JMIR Mental Health

Journal Impact Factor (JIF) (2023): 4.8
Volume 10 (2023) ISSN 2368-7959 Editor in Chief: John Torous, MD, MBI

Contents

Editorials

Original Papers

Reviews

Short Paper

Viewpoints

Corrigenda and Addendas

Editorial

Focusing on Digital Research Priorities for Advancing the Access and Quality of Mental Health

John Torous[1], MD; Nicole M Benson[2], MD; Keris Myrick[1,3], MBA; Gunther Eysenbach[4], MD

[1]Beth Israel Deaoness Medical Center, Boston, MA, United States
[2]McLean Hospital, Belmont, MA, United States
[3]Inseperable, Washington, DC, United States
[4]JMIR Publications, Toronto, ON, Canada

Corresponding Author:
John Torous, MD
Beth Israel Deaoness Medical Center
330 Brookline Ave
Boston, MA, 02446
United States
Phone: 1 6176676700
Email: jtorous@gmail.com

Abstract

Digital mental health solutions are now well recognized as critical to solving the global mental health crisis. As research accelerates, it is now clear that solutions ranging from computer-based therapy programs to virtual reality headsets and smartphone apps to large language model chatbots are of interest, feasible, and hold exciting potential to improve mental health. This research should now consider the next generation of scientific and clinical questions regarding if these new approaches are equitable, valid, effective, implementable, efficacious, and even cost-effective. This paper outlines several of the new frontiers for the next generation of research and introduces JMIR Publications' partnership with the Society of Digital Psychiatry to further advance these aims.

(JMIR Ment Health 2023;10:e47898) doi:10.2196/47898

KEYWORDS

digital phenotyping; mental health; depression; anxiety; smartphone

With the rapid acceleration of digital mental health, there is a growing need to ensure that the field delivers on its potential to increase access and quality of care. While the initial phase of research has already yielded important insights, it has also highlighted key challenges and areas that require further attention. Rather than providing a comprehensive overview of the field, the aim of this paper is to emphasize specific aspects of digital mental health that may be of particular interest as we move forward.

Equity in digital health has emerged as a key concern, as early assumptions that digital mental health tools would easily reach and serve all people have been replaced with the realization that several digital divides exist, especially around digital access, literacy, and skills. Therefore, addressing digital literacy and skills barriers may represent the single most important priority for ensuring that the full potential of digital health can be harnessed to benefit people with the greatest needs. It will also be important to explore the role of digital technology as a new social determinant of health as well as a tool to support addressing social determinants of mental health.

While studies that showcase the effectiveness of digital tools or interventions in single use cases are important, those demonstrating their multisite and global applications are particularly relevant. Adapting and tailoring digital tools to suit the needs of diverse cultural and regional contexts is still relatively uncommon, despite the potential for significant impact. Research focused on programs that can be scaled to serve diverse communities and provide care to underserved populations is therefore of great interest. Likewise, examining how ideas can flow in both directions and new solutions can emerge from all regions can yield valuable insights and promote more equitable access to digital health worldwide. Relatedly, equity in the voices of researchers and authors remains a priority, and *JMIR Mental Health* will continue its "Patient Perspective" pieces to offer a direct voice to those with lived experience.

Ensuring replicability is a cornerstone of all scientific research, and digital mental health research is no exception. Studies examining the impact of proprietary tools that are not accessible to others pose a significant threat to reproducible science. While such research is still valuable, authors must explore creative

solutions that allow others to engage with their approach and build on their findings. Similarly, to promote transparency and minimize potential bias, statistical and machine learning models should be presented in a clear and accessible manner that enables replication and assessment of bias. Relatedly, addressing challenges around missing data, especially from digital phenotyping research, and assumptions in preprocessing data are now critical for the success of this work. Preregistration of models, use of open-source tools, and data sharing are essential measures that will become increasingly critical to the field. Validation with biological markers will become more feasible and relevant, but validation through practical use and real-world positive outcomes will always be the most relevant.

Privacy is a top concern for the digital health space, particularly in the realm of digital mental health. While research that highlights this concern is valuable in measuring progress, papers that present novel solutions to these privacy issues are of even greater interest. Topics that are likely to have a high impact include efforts to increase public awareness and engagement around privacy as well as new regulatory proposals to enhance privacy protections. Relatedly, ethics in research remains equally salient, and questions about the informed nature of consent based solely on legalese buried in privacy policies will continue to raise concerns.

Efficacy is a key focus of the digital mental health space, and there is a wide range of claims about the clinical impact of various interventions. However, it has become clear that the level of human support offered can significantly impact the efficacy of these interventions. Therefore, research that assesses the optimal degree and mode of human support necessary to make interventions more effective is of great interest. Additionally, claims of efficacy should likely be compared against an active digital control intervention to ensure their rigor and relevance. While research without adequate control groups can help refine interventions and provide important feasibility data, it is less likely to contribute to a deeper understanding of efficacy. Likewise, retrospective studies of select cohorts of certain patient populations are interesting, but prospective studies are often the most informative.

As concerns about low engagement with digital health tools have become more apparent, there is an urgent need for innovative solutions. Studies that fail to provide clear insights into engagement present a significant challenge to the field, and transparent reporting has become more critical than ever. As noted earlier, providing human support can play a crucial role in enhancing engagement, and it is also important to focus on research that investigates design factors and adaptive interventions with context awareness. Furthermore, with growing recognition that clinician or coach support is vital for patient engagement, research on the adoption of clinical systems and reports on their implementation in care settings have become especially relevant and timely. Beyond engagement, research that explores the extent to which digital technologies impact client activation—clients taking an active role in treatment and maximizing shared decision-making between provider and client to move toward recovery outcomes—is also a needed area of focus.

As the field of digital health continues to grow, so too will the areas of research prioritization. Given the vast and dynamic nature of this field, there will always be exceptional topics that need to be covered despite any abovementioned concerns. For example, cross-sectional studies of interest in digital mental health have already been well reported on, but with new advances in large language models and related artificial intelligence language models, it will be important to understand interest, concerns, and potential even before more advanced research. New voices need to be heard and innovative pilot studies run with new advances in these language models. However, our collective goal remains the same: to highlight only the highest quality research. To achieve this goal, papers should continue to emphasize the clinical relevance of their findings and explain how their work can help advance access and quality of care.

Finally, as the field expands, there is a need to bring new voices into the field and transform research findings into practice. To accomplish these goals, JMIR Publications has partnered with the Society of Digital Psychiatry to make *JMIR Mental Health* the official society journal. As the society advances standards and advocates for high-quality practices/standards, the journal will serve to convene stakeholders and disseminate consensus papers. By bridging academics, industry, patient, clinician, regulatory, and family voices, the synergy of *JMIR Mental Health* and the Society of Digital Psychiatry will together help advance the mission of increased access and quality of mental health services for all.

Conflicts of Interest

JT is the editor in chief of *JMIR Mental Health* at the time of this publication and is a scientific advisor for Precision Mental Wellness. NB volunteers on the Epic Behavioral Health Specialty Steering Board. NB and KM serve on the editorial board for *JMIR Mental Health* at the time of this publication. GE is the founder, executive editor, and publisher at JMIR Publications, receives a salary, and owns equity.

Edited by T Leung; submitted 04.04.23; this is a non–peer-reviewed article;accepted 17.04.23; published 24.04.23.

Please cite as:
Torous J, Benson NM, Myrick K, Eysenbach G
Focusing on Digital Research Priorities for Advancing the Access and Quality of Mental Health
JMIR Ment Health 2023;10:e47898
URL: https://mental.jmir.org/2023/1/e47898
doi:10.2196/47898
PMID:37093624

Original Paper

Findings From the Step Up, Test Up Study of an Electronic Screening and Brief Intervention for Alcohol Misuse in Adolescents and Young Adults Presenting for HIV Testing: Randomized Controlled Efficacy Trial

Niranjan S Karnik[1], MD, PhD; Lisa M Kuhns[2], MPH, PhD; Anna L Hotton[3], PhD; Natascha Del Vecchio[3], MS; Moira McNulty[3], MD; John Schneider[3], MPH, MD; Geri Donenberg[4], PhD; Kristin Keglovitz Baker[5], PA-C; Rose Diskin[2], BA; Abigail Muldoon[2], MA; Juan Rivera[5], BA, CHES; Faith Summersett Williams[2], PhD; Robert Garofalo[2], MPH, MD

[1]Institute for Juvenile Research, Department of Psychiatry, University of Illinois Chicago, Chicago, IL, United States

[2]The Potocsnak Family Division of Adolescent and Young Adult Medicine, Ann & Robert H Lurie Children's Hospital of Chicago, Chicago, IL, United States

[3]The Chicago Center for HIV Elimination, The University of Chicago, Chicago, IL, United States

[4]Center for Dissemination and Implementation Science, University of Illinois Chicago, Chicago, IL, United States

[5]Center for Education, Research & Advocacy, Howard Brown Health, Chicago, IL, United States

Corresponding Author:
Niranjan S Karnik, MD, PhD
Institute for Juvenile Research
Department of Psychiatry
University of Illinois Chicago
1747 West Roosevelt Road
Chicago, IL, 60608
United States
Phone: 1 3122730185
Email: nkarnik@uic.edu

Abstract

Background: Substance use, particularly binge drinking of alcohol and noninjection substance use, is associated with increased risk for HIV infection among youth, but structured substance use screening and brief intervention are not often provided as part of HIV risk reduction.

Objective: The purpose of the study was to test the efficacy of a fully automated electronic screening and brief intervention, called Step Up, Test Up, to reduce alcohol misuse among adolescents and young adults presenting for HIV testing. Secondary objectives were reduction in sexual risk and uptake of pre-exposure prophylaxis (PrEP) for HIV prevention.

Methods: Youth aged 16 years to 25 years who presented for HIV testing at community-based locations were recruited for study participation. Those who screened at moderate to high risk on the Alcohol Use Disorders Identification Test were randomized (1:1) to either an electronic brief intervention or a time-attention control. The primary outcome was change in alcohol use at 1, 3, 6, and 12-month follow-ups. Negative binomial and log binomial regression analyses with generalized estimating equations were conducted to evaluate the intervention efficacy.

Results: Among a sample of 329 youth, there were no significant differences in alcohol use outcomes between conditions over time or at the 1, 3, 6, or 12-month time points. In terms of secondary outcomes, there was evidence of reduction in condomless insertive anal sex under the influence of alcohol and drugs at 12 months compared with 3 months in the intervention versus the attention control condition (incidence rate ratio=0.15, 95% CI 0.05-0.44); however, there were no other significant differences in sexual risk and no difference in PrEP engagement.

Conclusions: We found no effect of electronic brief intervention to reduce alcohol use and some effect on sexual risk among youth aged 16 years to 25 years who present for HIV testing.

Trial Registration: ClinicalTrials.gov number NCT02703116; https://clinicaltrials.gov/ct2/show/NCT02703116

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-020-8154-6

(JMIR Ment Health 2023;10:e43653) doi:10.2196/43653

KEYWORDS

HIV prevention; men who have sex with men; transgender women; alcohol intervention; HIV; gay; homosexual; MSM; alcohol; youth; screening; sexual behavior; sexual behavior; sexual risk; risky; pre-exposure prophylaxis; prophylaxis; prevention; efficacy; adolescent; young adult; testing; risk

Introduction

The syndemic intersection of HIV risk and substance misuse among young men who have sex with men (YMSM) and young transgender women (YTW) has been well established and well documented [1-5]. Recent studies have shown that substance use, particularly binge drinking of alcohol and noninjection substance use, increases sexual risks and the potential for HIV infection [6]. In 2019, over 20% of new HIV infections occurred in youth aged 13 years to 19 years [7]. Among young people, the vast majority (66%) of HIV infections continue to occur among those having male-to-male sexual contact with YMSM constituting the largest share. Among those with male-to-male sexual contact, 25% of infections occur among YMSM aged 13 years to 24 years, and 31% occur among those aged 25 years to 29 years [7]. In addition to YMSM, substantial evidence exists that YTW are also at risk for HIV, and in 2019, their rate of diagnosis increased by 5% [7]. Along with routine condom use and pre-exposure prophylaxis (PrEP), addressing alcohol and other drug (AOD) use are among the most modifiable individual-level factors for the prevention of HIV and sexually transmitted infections (STIs). Although there are many structural factors including racism, stigma, and poverty that contribute to the dynamics of HIV infection, prevention strategies have often focused on individual or personal-level factors. Recent years have seen shifts toward HIV interventions in community clinics that provide primary care or specialized HIV services. At the intersection of individual behavioral change and structural change are attempts to insert brief motivational strategies into larger systems of care. Motivational interviewing (MI)–based approaches promote screening, brief intervention, and referral to treatment (SBIRT). Electronic screening and brief intervention (eSBI) is a subset of SBIRT in an electronic medium suitable for use in primary care and other generalist settings. YMSM and YTW often seek HIV testing and other supportive services in community-based and outreach settings. These settings are underutilized as potential entry points for engagement in comprehensive care across the HIV prevention and care continuum, including PrEP for HIV-negative youth who are at risk of HIV acquisition. We theorized that these same community settings could serve as an access point for substance use interventions if these interventions were brief and scalable.

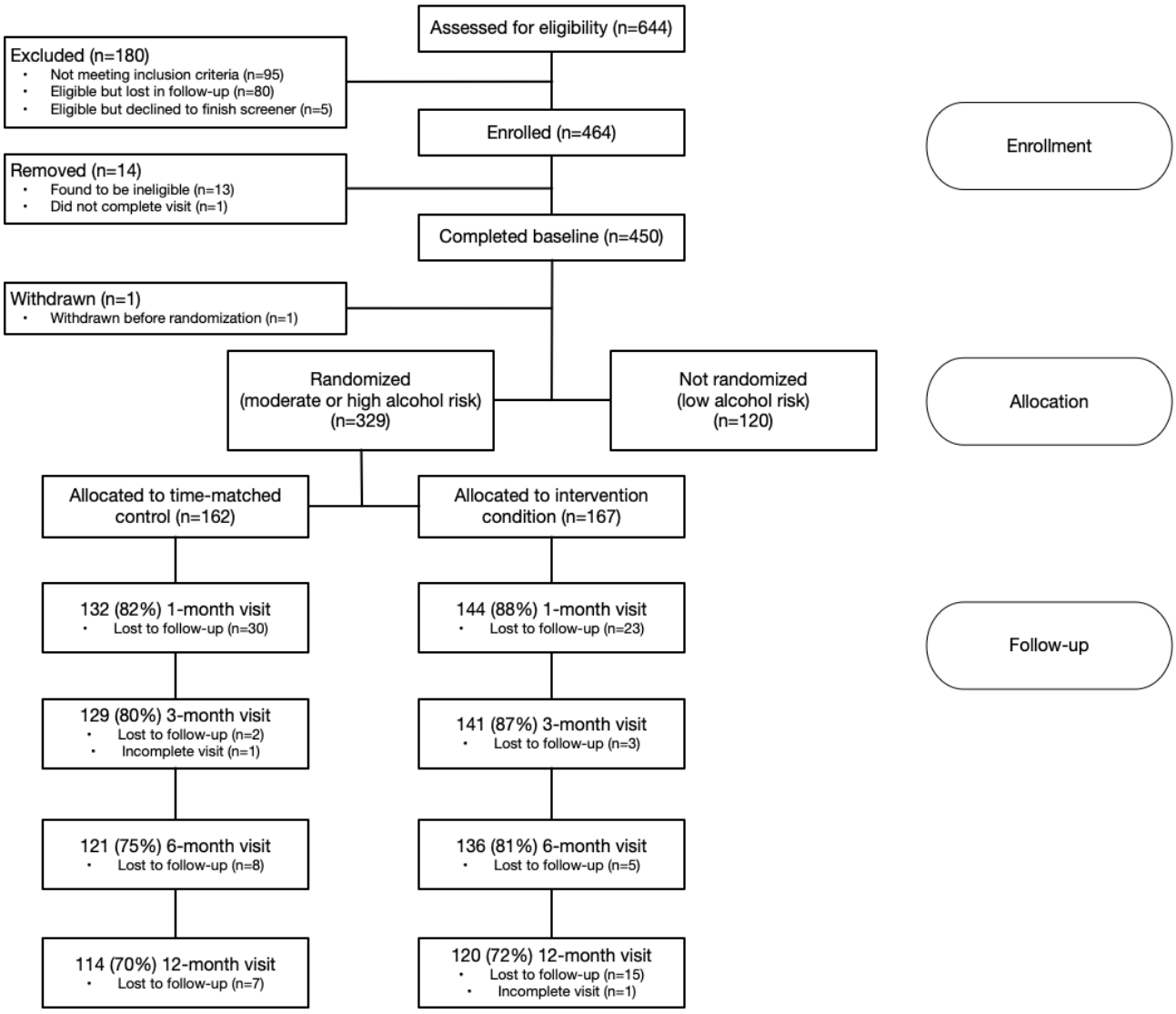
The purpose of this study was to assess the feasibility, acceptability, and initial efficacy of eSBI in comparison to an electronic attention control intervention (ie, promotion of good nutrition), coupled with standard HIV prevention, on alcohol use among YMSM and YTW in community-based HIV testing environments in Chicago. Secondary objectives were to assess intervention effects on sexual behavior, as well as engagement within the HIV prevention and care continuum, and to assess modification of the intervention effect by comorbid mental health problems (ie, symptoms of depression and anxiety).

Methods

Design

This study was a randomized controlled trial of eSBI among YMSM and YTW seeking HIV testing [8]. Using an electronic web-based portal, all participants were screened for alcohol misuse and received immediate feedback regarding their level of use (eg, how their use compares to others, whether it exceeds “safe use” guidelines); those who screened for moderate to high alcohol use, including binge drinking, were then randomized to either electronic intervention or control modules. All participants, regardless of randomization status, were followed for 12 months with in-person visits conducted at 1, 3, 6, and 12-month intervals (see CONSORT diagram, Figure 1). All participants were recruited from HIV testing clinics that utilized a seek, test, treat, and retain (STTR) model of care [9].

Figure 1. Step up, Test Up CONSORT diagram.



Ethical Approval

The protocol was approved by the Ann & Robert H. Lurie Children's Hospital Institutional Review Board (review #2015-703) with a waiver of parental permission for participation of minors (aged 16-17 years). All enrolled participants were consented prior to the start of any research activities using an electronic consent form. The consent form indicated that the purpose of the study was to learn more about whether screening and brief intervention for substance use is effective in reducing that use.

Identification and Recruitment of Participants

Youth were recruited from HIV testing centers at the 3 primary sites in Chicago: the Division of Adolescent Medicine at Lurie Children's Hospital, Howard Brown Health, and the Village at the University of Chicago. Individuals were eligible if they were (1) aged 16 years to 25 years, (2) interested in testing for HIV infection, (3) HIV-negative or HIV status unknown (per self-report; verified at the point of HIV testing), (4) identified as a man who has sex with men or a transgender woman who has sex with men (ie, born male, identify as female or transgender, and at any point in the gender transition process), and (5) English-speaking. Individuals who were eligible and interested were enrolled by study staff.

Randomization

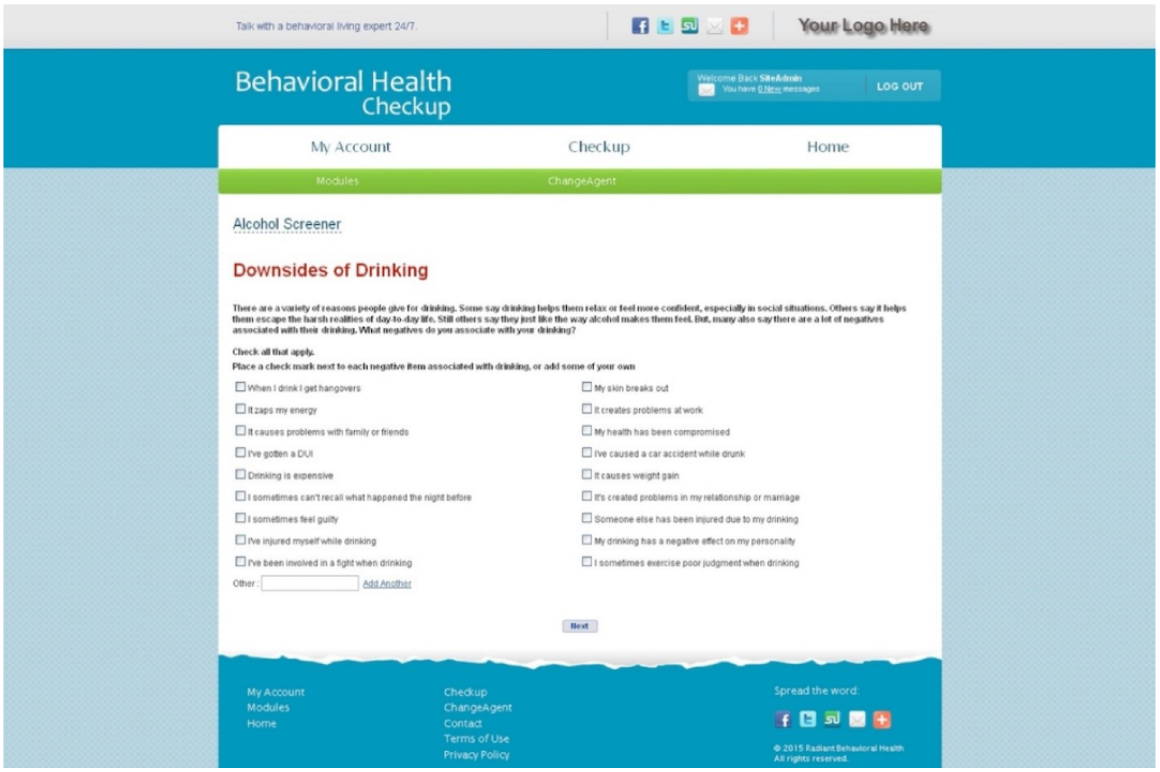
Upon enrollment, all participants completed HIV testing (using standard of preventive care testing and counseling at each site); individuals who screened reactive on the rapid HIV test were immediately referred for confirmatory HIV testing and were withdrawn from the study. All participants were screened for alcohol misuse using the Alcohol Use Disorders Identification Test (AUDIT) via an electronic portal. The AUDIT is a widely used, 10-item, validated screening tool for the assessment of alcohol use [10]. It contains items that reflect consumption, dependency, and alcohol-related harms. The AUDIT total score ranges from 0 to 40, divided into 4 Zones (I-IV), reflecting increasing levels of risk and recommended intervention (ie, Zone 1: alcohol education; Zone II: simple advice; Zone III: simple advice plus brief counseling and monitoring; Zone IV: referral to specialist for diagnostic evaluation and treatment). In addition, one item on the AUDIT reflects binge drinking ("How often do you have six or more drinks on one occasion?"). For our study, those scoring in Zones II-IV or endorsing the binge drinking item were included.

After completion of the AUDIT, participants received immediate feedback regarding their level of use. Participants who screened moderate to high risk for alcohol misuse or endorsed binge drinking on the AUDIT were randomized via computerized assignment (1:1) to either brief intervention modules to reduce alcohol misuse, eSBI, or a time and attention-matched control (ie, promotion of good nutrition). Study staff were blinded to random assignments. The randomization sequence was generated by NSK, uploaded to the Radiant platform (Radiant Interactive Group), and assigned automatically in the order of enrollment.

Description of the Intervention: eSBI

Those randomized to the intervention were asked to complete lessons or exercises on 11 topical areas focused on alcohol use (eg, importance of change, downsides of drinking, barriers to change, strategies for cutting back, change motivators) in an MI format, each with a single web page. The intervention is not a full MI intervention, as it was relatively simple in its presentation, but it did have components of the Elicit-Provide-Elicit approach. The intervention had been widely used in employee assistance programs. The intervention content was adapted for late adolescents and young adults in terms of voice and language (eg, open, nonjudgmental, inclusive), contexts (eg, school as well as work contexts), and reading level. The intervention was presented as a series of slides that contained an interactive element in which participants could click on checkboxes, drop-down menus, and short text fields as part of the brief MI exercises. An example slide is shown in Figure 2. By design, the intervention lacked any specific cultural tailoring for YMSM or YTW other than the age-appropriate language. This was intended to facilitate rapid scaling and broad adoption across the general population if the intervention proved effective. The electronic intervention and control modules were delivered using a customized eSBI platform developed by Radiant Interactive Group (Laguna Niguel, CA). Radiant uses industry-standard data encryption and security to manage protected health information recorded by their system.

Figure 2. Example slide from the electronic screening and brief intervention (eSBI).



Time and Attention-Matched Control: Diet and Nutrition

Those randomized to the control condition completed a brief time and attention-control intervention, also in an MI format and of equal length (ie, same number of intervention screens), which encouraged good nutrition but was not expected to impact alcohol use.

Standard of Care Prevention Services

The standard HIV preventive care at each site included rapid point-of-care HIV testing and counseling and linkage navigation to PrEP or HIV care, depending on the test result.

Study Assessments

Participants completed a baseline study visit that included a standardized assessment via computer-assisted self-interviewing (CASI). Follow-up CASI assessments were conducted at 1, 3, 6, and 12-month follow-up visits either in person or remotely via web-based assessment. Detailed contact information was collected to facilitate study retention.

Primary Outcome

All outcome measures were self-reported and constructed using the revised versions of the Daily Drinking Questionnaire [11] (DDQ-R [12]) including questions regarding the heaviest drinking week [13]. The questionnaire asks the participant to report the number of drinks consumed for each day of the week

for both a typical week and heavy drinking week in the past 30 days. The outcomes assessed for this analysis were (1) the total number of drinks consumed in a typical drinking week, (2) the total number of drinks consumed in a peak drinking week, (3) the number of days 5 or more drinks were consumed in a typical week, (4) the number of days 5 or more drinks were consumed in a heavy drinking week, and (5) any binge drinking (ie, 5 or more drinks consumed in 1 episode) in the prior month.

Secondary Outcomes

Sexual risk behavior was measured via a modified version of the AIDS-Risk Behavior Assessment [14] used in prior studies with YMSM and YTW [15,16] and included (1) condomless receptive anal sex (RAS) acts, (2) condomless RAS acts while under the influence of alcohol or drugs, (3) condomless insertive anal sex (IAS) acts, and (4) condomless IAS acts while under the influence of alcohol or drugs. Condomless sex was defined as sex in which "no condom was used or a condom was used but only for a part of the time." Attending at least one provider visit for PrEP care during the observation period was additionally included as a secondary outcome for HIV-negative participants who reported using PrEP (n=89).

We measured eSBI intervention satisfaction and acceptability using the Client Satisfaction Questionnaire (CSQ-8) [17]. Participants were asked to rate the quality of the intervention and their satisfaction, as well as whether they would recommend it to friends. We also include an open-ended item, "Do you have any other suggestions to improve this intervention?"

Moderators

Depression and anxiety were measured using the short version of the Center for Epidemiologic Studies Depression Scale (CES-D10) [18] and 7-item Generalized Anxiety Disorder (GAD-7) [19], respectively.

Statistical Analysis

The analysis used an intent-to-treat approach in which all randomized participants were included for analysis. Baseline characteristics of the intervention groups were inspected visually for imbalances; differences between the 2 intervention groups were assessed using chi-square or Fisher exact tests for categorical variables and *t* tests for numerical variables. Changes in drinking levels over time were examined descriptively using means with SDs and medians with IQRs. The targeted sample size of 450 was estimated to yield 80% power to detect effects for our primary (alcohol use) outcome under a range of retention levels (75%-85%).

To evaluate the efficacy of the intervention, generalized estimating equations (GEEs) with robust standard errors were conducted. Each outcome was regressed on intervention group, time, a group-by-time interaction, and the baseline value of the outcome, where the statistical significance of the interaction term indicates whether the groups differ in response to the intervention over time. The baseline outcomes were included in the model due to baseline imbalances in the outcomes between the groups. Unadjusted models are presented in Table S1 in Multimedia Appendix 1. Omnibus tests (Type III Score tests) were used to assess the overall effects of the intervention, time, and whether the differences between the groups varied across time points. Contrasts were performed to assess the effect of the intervention compared with the time and attention-matched control at each follow-up visit compared with first follow-up for exploratory purposes even when omnibus tests were not significant. Interaction contrasts represent the relative risk (RR) for binary outcomes or incidence rate ratio (IRR) for count outcomes in the intervention arm at that time point compared with the RR or IRR for the group comparison at first follow-up (ie, a ratio of ratios). All count outcomes were modeled as negative binomially distributed variables with log link functions to compute rate ratios; the binary outcome (having had more than 5 drinks in any single drinking session in the prior month) was modeled as a binomially distributed variable with log link function to estimate relative risk. An exchangeable working correlation was specified to account for repeated measurements on participants over time. All statistical analyses were conducted using SAS software version 9.4 (SAS Institute). Graphics were generated using STATA 17.

Results

Participants

A total of 644 individuals were screened for the study: 549 were eligible, and 95 were not eligible. Study enrollment began in December 2016, and all follow-up visits were completed by October 2020. Among those eligible, 85 (85/549, 15.5%) were lost to follow-up prior to enrollment, resulting in 464 enrolled (see CONSORT diagram, Figure 1). Of those enrolled, 14 were withdrawn or considered incomplete, resulting in a final sample of 450 (450/549, 82% of eligible).

All enrolled participants were screened with the AUDIT; 120 screened low risk and were not randomized. Therefore, a total of 329 participants were included in the analysis described herein. Table 1 presents descriptive information on the 329 participants randomized to the intervention or attention control conditions. The average age of the participants was 22.8 years. Most participants identified as non-White (221/329, 67.2%). The majority identified as cisgender male (294/329, 89.4%) and gay (205/329, 62.3%). Most of the participants were employed either full-time or part-time (222/329, 67.7%) and had at least some college education or attended trade school (224/329, 68.3%). There were 70 (70/329, 21.3%) with a history of incarceration. A little more than one-half of participants had a primary care provider (174/329, 52.9%). There were no significant differences between the intervention groups in these characteristics at baseline. For each measure of drinking, there were reductions in the mean amounts of drinking over time in both groups (Table 2). The average duration of the eSBI intervention was just under 10 minutes and of the control condition was just over 10 minutes.

Table 1. Demographics for the intervention (electronic screening and brief intervention [eSBI]; n=167) and control (n=162) groups.

Characteristics	eSBI[a]	Control[a]	*P* value
Age (years), mean (SD)	22.89 (1.99)	22.79 (1.97)	.65
Gender identity, n (%)			.88
Cisgender male	150 (89.8)	144 (88.9)	
Transgender female	15 (9.0)	15 (9.3)	
Transgender male	2 (1.2)	3 (1.9)	
Sexual orientation, n (%)			.98
Bisexual	42 (25.1)	41 (25.3)	
Gay	105 (62.9)	100 (61.7)	
Queer	16 (9.6)	16 (9.9)	
Straight/heterosexual	4 (2.4)	5 (3.1)	
Race/ethnicity, n (%)			.68
White NH[b]	60 (35.9)	48 (29.6)	
Black/African American NH	65 (38.9)	68 (42.0)	
Hispanic/Latino/a	24 (14.4)	26 (16.0)	
Other NH	18 (10.8)	20 (12.3)	
Employment, n (%)			.44
Employed full-time	54 (32.3)	63 (39.1)	
Employed part-time	56 (33.5)	49 (30.4)	
Unemployed	57 (34.1)	49 (30.4)	
Education, n (%)			.59
Less than high school	15 (9.0)	13 (8.1)	
High school or GED[c]	36 (21.6)	40 (24.8)	
Some college or trade school	50 (29.9)	55 (34.2)	
College degree or higher	66 (39.5)	53 (32.9)	
Arrest history, n (%)	44 (26.3)	53 (32.9)	.24
Incarceration history, n (%)	30 (18.0)	40 (24.8)	.17
Currently has a PCP[d], n (%)	82 (49.1)	92 (56.8)	.20
Baseline HIV test results, n (%)			.21
Nonreactive	167 (100.0)	159 (98.1)	
Reactive	0 (0)	2 (1.2)	
Confirmed positive	0 (0)	1 (0.6)	

[a]Not all sum to 329 due to missing data.

[b]NH: non-Hispanic.

[c]GED: General Education Development.

[d]PCP: primary care physician.

Table 2. Outcomes by treatment condition at all time points.

Outcomes	All participants (n=329)	Control (n=162)	eSBI[a] intervention (n=167)
Number of drinks in a typical drinking week, mean (SD)			
Baseline	11.89 (11.64)	10.51 (10.45)	13.24 (12.57)
1 month	8.92 (10.01)	9.33 (12.31)	8.55 (7.31)
3 months	8.51 (9.25)	7.94 (8.43)	9.04 (9.94)
6 months	7.90 (8.00)	7.49 (7.91)	8.27 (8.09)
12 months	8.68 (9.61)	8.35 (7.74)	8.98 (11.12)
Number of drinks in a typical drinking week, median (IQR)			
Baseline	8.00 (4.00-16.00)	8.00 (4.00-14.00)	10.00 (5.00-17.00)
1 month	6.00 (3.00-12.00)	6.00 (3.00-12.00)	7.00 (3.00-13.00)
3 months	6.00 (3.00-11.00)	6.00 (3.00-10.00)	6.00 (2.00-13.00)
6 months	6.00 (2.00-12.00)	5.00 (2.00-11.00)	6.00 (2.00-12.00)
12 months	7.00 (2.00-12.00)	7.00 (2.00-11.75)	6.00 (2.00-12.00)
Number of drinks in a peak drinking week, mean (SD)			
Baseline	6.36 (4.84)	5.83 (4.48)	6.87 (5.13)
1 month	5.09 (3.84)	5.15 (4.14)	5.04 (3.57)
3 months	4.96 (4.00)	4.71 (3.81)	5.18 (4.17)
6 months	4.86 (3.54)	4.62 (3.42)	5.08 (3.64)
12 months	4.90 (3.63)	4.75 (3.24)	5.04 (3.98)
Number of drinks in a peak drinking week, median (IQR)			
Baseline	5.00 (3.00-8.00)	5.00 (3.00-8.00)	6.00 (4.00-9.00)
1 month	5.00 (2.00-7.00)	4.00 (2.75-7.00)	5.00 (2.00-7.00)
3 months	4.00 (2.00-7.00)	4.00 (2.00-6.00)	4.00 (2.00-8.00)
6 months	4.00 (2.00-7.00)	4.00 (2.00-6.00)	4.50 (2.00-7.25)
12 months	5.00 (2.00-7.00)	5.00 (2.00-7.00)	5.00 (2.00-7.00)
Number of days with 5 or more drinks in a typical drinking week, mean (SD)			
Baseline	0.95 (1.41)	0.84 (1.30)	1.05 (1.51)
1 month	0.63 (1.14)	0.69 (1.24)	0.58 (1.03)
3 months	0.58 (1.02)	0.47 (0.85)	0.69 (1.14)
6 months	0.56 (1.01)	0.45 (0.88)	0.65 (1.10)
12 months	0.59 (1.07)	0.57 (0.99)	0.62 (1.14)
Number of days with 5 or more drinks in a typical drinking week, median (IQR)			
Baseline	0 (0-2.00)	0 (0-1.00)	0 (0-2.00)
1 month	0 (0-1.00)	0 (0-1.00)	0 (0-1.00)
3 months	0 (0-1.00)	0 (0-1.00)	0 (0-1.00)
6 months	0 (0-1.00)	0 (0-1.00)	0 (0-1.00)
12 months	0 (0-1.00)	0 (0-1.00)	0 (0-1.00)
Number of days with 5 or more drinks in a peak drinking week, mean (SD)			
Baseline	1.53 (1.78)	1.32 (1.62)	1.74 (1.90)
1 month	1.02 (1.33)	1.01 (1.44)	1.03 (1.23)
3 months	1.04 (1.38)	0.95 (1.28)	1.11 (1.47)
6 months	1.06 (1.46)	0.98 (1.29)	1.13 (1.59)
12 months	1.23 (1.54)	1.18 (1.44)	1.27 (1.64)

Outcomes	All participants (n=329)	Control (n=162)	eSBI[a] intervention (n=167)
Number of days with 5 or more drinks in a peak drinking week, median (IQR)			
Baseline	1.00 (0-2.00)	1.00 (0-2.00)	1.00 (0-3.00)
1 month	1.00 (0-2.00)	0.00 (0-1.00)	1.00 (0-2.00)
3 months	0.50 (0-2.00)	1.00 (0-2.00)	0 (0-2.00)
6 months	1.00 (0-2.00)	1.00 (0-1.00)	0.50 (0-2.00)
12 months	1.00 (0-2.00)	1.00 (0-2.00)	1.00 (0-2.00)
Number of any days with 5 or more drinks in 1 event in the past month, n (%)			
Baseline	193 (58.7)	85 (52.5)	108 (64.7)
1 month	141 (51.1)	65 (49.2)	76 (52.8)
3 months	131 (48.5)	62 (48.1)	69 (48.9)
6 months	128 (49.8)	60 (49.6)	68 (50.0)
12 months	125 (53.4)	61 (53.5)	64 (53.3)

[a]eSBI: electronic screening and brief intervention.

Primary Outcomes

The results of the adjusted GEE models for primary outcomes are presented in Table 3. The omnibus tests (Type III Score tests) showed that the overall effect of the intervention on substance use was not statistically significant (ie, no significant differences between groups across all time points) for any of the primary outcomes. The overall effect of time was statistically significant only for binge drinking in a peak drinking week at 12 months compared with 1 month (ie, there was an increase in drinking outcomes over time across both groups; IRR=1.32, 95% CI 1.05-1.66), and the intervention group-by-time interaction was not statistically significant (ie, differences between groups did not vary over time) for any of the primary outcomes. Figure 3 displays the results of the GEE models with the predicted values and lower and upper 95% CIs.

Table 3. Primary outcomes from the generalized estimating equation estimates of the treatment effect.

Model[a]	Incidence rate ratio (95% CI)	*P* value[b]
Number of drinks in a typical drinking week		
Treatment group		.49[c]
eSBI[d]	0.97 (0.80-1.16)	.71
Time		.56[c]
1 month	Reference	Reference
3 months	0.98 (0.83-1.14)	.74
6 months	0.98 (0.85-1.14)	.80
12 months	1.12 (0.97-1.30)	.14
Treatment effects (group × time)[e]		.55[c]
3 months to 1 month	1.04 (0.84-1.27)	.72
6 months to 1 month	1.00 (0.81-1.24)	.97
12 months to 1 month	0.89 (0.72-1.10)	.29
Number of drinks in a peak drinking week		
Treatment group		.88[c]
eSBI	0.98 (0.84-1.14)	.78
Time		.98[c]
1 month	Reference	Reference
3 months	0.95 (0.83-1.10)	.50
6 months	0.96 (0.86-1.06)	.43
12 months	1.00 (0.89-1.12)	.96
Treatment effects (group × time)[e]		.82[c]
3 months to 1 month	1.06 (0.88-1.28)	.52
6 months to 1 month	1.06 (0.91-1.24)	.43
12 months to 1 month	1.00 (0.84-1.19)	.99
Number of days with 5 or more drinks in a typical drinking week		
Treatment group		.76[c]
eSBI	0.88 (0.60-1.28)	.49
Time		.85[c]
1 month	Reference	Reference
3 months	0.75 (0.54-1.05)	.10
6 months	0.82 (0.58-1.15)	.24
12 months	1.02 (0.73-1.41)	.92
Treatment effects (group × time)[e]		.13[c]
3 months to 1 month	1.52 (0.99-2.31)	.056
6 months to 1 month	1.36 (0.88-2.13)	.17
12 months to 1 month	0.98 (0.63-1.53)	.94
Number of days with 5 or more drinks in a peak drinking week		
Treatment group		.62[c]
eSBI	0.98 (0.75-1.28)	.87
Time		.047[c]

Model[a]	Incidence rate ratio (95% CI)	*P* value[b]
1 month	Reference	Reference
3 months	1.02 (0.83-1.26)	.84
6 months	1.13 (0.89-1.43)	.33
12 months	1.32 (1.05-1.66)	.02
Treatment effects (group × time)[e]		.94[c]
3 months to 1 month	1.00 (0.76-1.33)	.97
6 months to 1 month	0.95 (0.69-1.31)	.77
12 months to 1 month	0.92 (0.66-1.27)	.60
Number of any days with 5 or more drinks in 1 event in the past month		
Treatment group		.33[c]
eSBI	0.99 (0.81-1.21)	.92
Time		.72[c]
1 month	Reference	Reference
3 months	0.97 (0.80-1.18)	.77
6 months	1.02 (0.86-1.22)	.80
12 months	1.08 (0.92-1.26)	.37
Treatment effects (group × time)[e]		.75[c]
3 months to 1 month	0.97 (0.76-1.23)	.79
6 months to 1 month	0.89 (0.69-1.15)	.37
12 months to 1 month	0.90 (0.71-1.15)	.40

[a]Each model also controlled for the baseline value of the outcome variable to control for baseline imbalances between the groups.

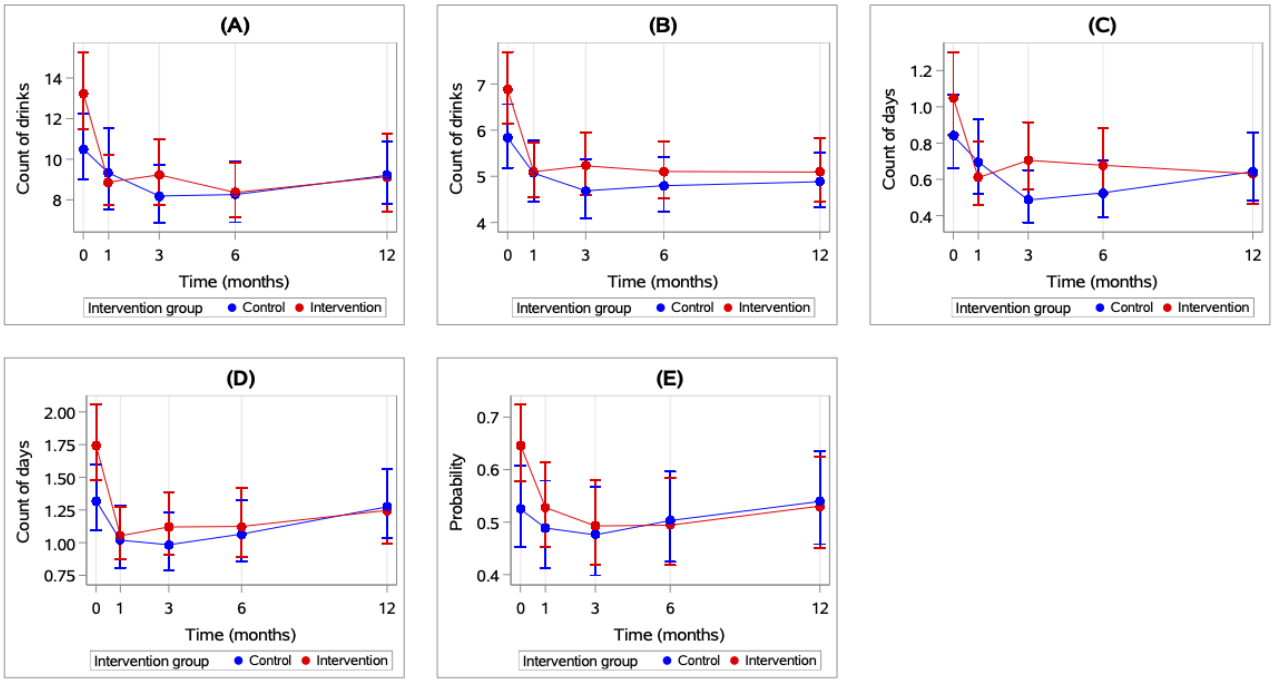
[b]Omnibus test *P* values were from Score tests; all other *P* values were from Wald tests.

[c]Omnibus *P* values.

[d]eSBI: electronic screening and brief intervention.

[e]The ratio of the intervention to control at the indicated time point versus baseline.

Figure 3. Primary outcome at each time point (0, 1, 3, 6, and 12 months) by intervention group based on generalized estimating equation (GEE) models: (A) number of drinks in a typical drinking week in the past 30 days, (B) number of drinks in a peak drinking week in the past 30 days, (C) number of days with 5 or more drinks in a typical drinking week in the past 30 days, (D) number of days with 5 or more drinks in a peak drinking week in the past 30 days, (E) any episodes with 5 or more drinks in the past 30 days.



Secondary Outcomes

The results of the adjusted GEE models for secondary outcomes are presented in Table 4. The intervention group showed a reduction in the number of condomless IAS acts while under the influence of alcohol or drugs, whereas the control group showed an increase in the number of acts from 3 months to 12 months (IRR=0.15, 95% CI 0.05-0.44) when examining the contrasts. There were no significant differences between the groups in the other sexual behavior outcomes or number of PrEP care visits between groups. The results of the GEE models are visualized in Figure 4.

In terms of acceptability and satisfaction, 89.7% (148/167) of intervention participants rated the quality of the intervention as "good" to "excellent," and 94.6% (156/167) indicated they were "mostly satisfied" to "very satisfied." A total of 89.7% (148/167) indicated that they would "probably" or "definitely" refer a friend to receive the intervention. The most common suggestions for improvement to the intervention were to include interaction with an educator or counselor (n=8); make the intervention less prescriptive for low levels of drinking (n=10); and include graphics, video, or music (n=3).

Table 4. Secondary outcomes from the generalized estimating equation estimates of the treatment effect.

Model[a]	Incidence rate ratio (95% CI)	*P* value[b]
Number of times of condomless receptive anal sex in the past 3 monthss		
Treatment group		.24[c]
eSBI	0.79 (0.46-1.36)	.40
Time		.16[c]
3 months	Reference	Reference
6 months	1.76 (1.26-2.45)	.001
12 months	1.17 (0.75-1.83)	.49
Treatment effects (group × time)[d]		.26[c]
6 months to 3 months	0.66 (0.37-1.17)	.16
12 months to 3 months	1.03 (0.51-2.07)	.94
Number of times of condomless receptive anal sex while under the influence of alcohol or drugs		
Treatment group		.33[c]
eSBI	0.80 (0.45-1.41)	.43
Time		.20[c]
3 months	Reference	Reference
6 months	1.87 (1.14-3.05)	.013
12 months	1.33 (0.81-2.17)	.26
Treatment effects (group × time)[d]		.49[c]
6 months to 3 months	0.74 (0.32-1.71)	.48
12 months to 3 months	1.10 (0.42-2.91)	.84
Number of times of condomless insertive anal sex in the past 3 months		
Treatment group		.29[c]
eSBI	1.08 (0.61-1.89)	.79
Time		.19[c]
3 months	Reference	Reference
6 months	1.48 (1.06-2.05)	.02
12 months	1.90 (1.11-3.26)	.02
Treatment effects (group × time)[d]		.25[c]
6 months to 3 months	0.73 (0.32-1.65)	.44
12 months to 3 months	0.53 (0.25-1.12)	.10
Number of times of condomless insertive anal sex while under influence of alcohol or drugs		
Treatment group		.94[c]
eSBI	2.48 (1.14-5.39)	.02
Time		.32[c]
3 months	Reference	Reference
6 months	2.39 (1.45-3.95)	.001
12 months	3.58 (1.69-7.56)	.001
Treatment effects (group × time)[d]		.06[c]
6 months to 3 months	0.41 (0.12-1.38)	.15
12 months to 3 months	0.15 (0.05-0.44)	.001

Model[a]	Incidence rate ratio (95% CI)	*P* value[b]
At least 1 PrEP[e] care visit in a 3-month period[f]		
Treatment group		.09[c]
eSBI	1.05 (0.80-1.37)	0.728
Time		.25[c]
3 months	Reference	Reference
6 months	0.83 (0.63-1.10)	.19
12 months	0.99 (0.74-1.32)	.94
Treatment effects (group × time)[d]		.41[c]
6 months to 3 months	1.28 (0.89-1.84)	.18
12 months to 3 months	1.17 (0.84-1.64)	.36

[a]Each model also controlled for the baseline value of the outcome variable to control for baseline imbalances between the groups.

[b]Omnibus test *P* values were from Score tests; all other *P* values were from Wald tests.

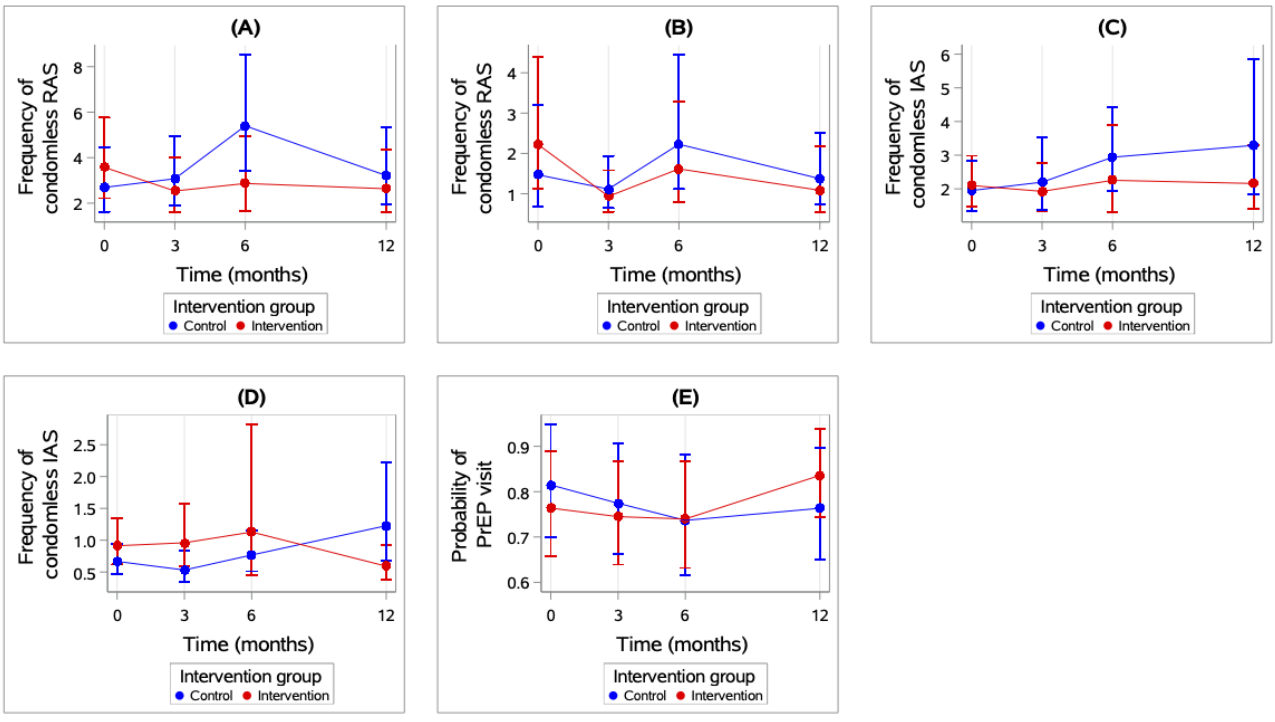
[c]Omnibus *P* values.

[d]The ratio of the intervention to control at the indicated time point versus baseline.

[e]PrEP: pre-exposure prophylaxis.

[f]Only including time points after the first indicated PrEP use.

Figure 4. Secondary outcome at each time point (0, 3, 6, and 12 months) by intervention group based on generalized estimating equation (GEE) models: (A) number of times of condomless receptive anal sex (RAS), (B) number of times of condomless RAS while under the influence of alcohol or drugs, (C) number of times of condomless insertive anal sex (IAS), (D) number of times of condomless IAS while under the influence of alcohol or drugs, (E) any visits to a provider for pre-exposure prophylaxis (PrEP).



Moderation Effects

As an exploratory aim, we conducted tests of moderation to determine whether the efficacy of the intervention was modified by the presence of depression or anxiety. The observed effects for anxiety were unstable and varied in direction, magnitude, and statistical significance depending on the analytic approach. When we controlled for baseline imbalances in the outcomes, the intervention appeared to be more efficacious among participants with fewer symptoms of anxiety; however, when we did not control for baseline imbalances, the intervention appeared more efficacious among participants with more anxiety symptoms. The contradictory findings may be due to associations between anxiety and the baseline outcomes, but the result is difficult to interpret and should be considered with

caution. No moderation effects were observed for depression in any of the analyses.

Discussion

Principal Findings

In this study, we found no effect of the eSBI intervention on alcohol misuse, but there was evidence of an effect on sexual risk. This study was born from a desire to create a low-cost, scalable, and brief single time-point intervention to address substance misuse in community settings. Prior literature has shown that computerized interventions can successfully reduce substance misuse using an MI approach. For example, Ondersma and colleagues [20] demonstrated that a single 20-minute eSBI intervention, with a more comprehensive MI approach, successfully improved abstinence from drugs at 3 months and 6 months among postpartum drug users. In contrast, our eSBI approach failed to show any major effect of eSBI on drinking behavior in moderate to high-risk drinkers. Although there was an effect at 12 months with significant increase in the number of binge drinking episodes in a peak week for both the eSBI group and our control condition, these differences were not seen at 3 months or 6 months, and the time-by-group effect was not significant. Our intervention's lack of effect, in contrast to the results by Ondersma et al [20], may be the result of differing populations (HIV at-risk YMSM and YTW vs postpartum women) and the need for cultural tailoring of the intervention for the former. In addition, the trial by Ondersma et al [20] used animation to deliver the intervention, which may have improved engagement with the content.

Since this study was conceived, there has been a number of trials of eSBI that have found no effect [21,22] or some impact in particular populations including college age drinkers [23], nonpregnant women of childbearing age [24], individuals on antiretroviral therapy [25], and graduate students using cannabis [26]. The literature on eSBI can best be described as mixed, and there is evidence that the more efficacious eSBI interventions have strong cultural tailoring, which our intervention did not have due to a desire to scale this intervention broadly. In addition, our MI approach was very brief and included relatively short MI components. It was by design simple, and we were seeking to test a system that might be broadly implemented. However, this brevity may have come with a price, as the intervention may not have had the robustness of a full MI intervention.

Despite the lack of findings for substance misuse, there was evidence of reduction in sexual risk in the intervention group in comparison to the control group for condomless IAS under the influence at 12 months. These analyses should be interpreted with caution, as there were multiple secondary analyses. Nevertheless, this finding is intriguing and should be the subject of additional study.

At the time that this study was conceived in 2014, trials had mixed findings on the impact of brief interventions on AOD. Since that time, observers have raised criticism of the value of screening and brief interventions [27-29]. The Substance Abuse and Mental Health Services Administration (SAMHSA)–promoted SBIRT model has generally shown impact on individuals with high levels of substance use and in primary care settings [27]. It is possible that our intervention was too brief and lacking in boosters or follow-up. Further trials may be merited with these elements in a dynamic trial design that might allow simultaneous comparisons.

Another promising line of inquiry that has shown impact is the use of relational agents (ie, using avatars, conversational bots, or other interfaces) that simulate face-to-face dynamics and allow for a more holistic experience during screening and a brief intervention [30-32]. Relative to this type of relational agent, our intervention was static; designed with simple, scalable elements; and presented in a slide format. The simplistic design was by intention, as we hypothesized that many clinics would not have the infrastructure, computing capabilities, and high bandwidth access that are often required for more computationally complex frameworks. We suspect that our eSBI system lacked enough appealing components to produce a robust attentional response; however, as we did not measure attention to the eSBI system, this is purely speculative.

Although, overall, the intervention received high ratings for acceptability and satisfaction, some participants suggested the addition of more dynamic and engaging components (eg, more graphics, video, and music), commented that the fully automated version was impersonal and would have preferred more interaction with an educator or counselor, or felt the intervention was too prescriptive for perceived low levels of drinking. These are important considerations to inform future iterations in this population.

Finally, since we conceptualized this study in 2014, eSBI approaches have been tested for feasibility among people living with HIV in sub-Saharan Africa, where alcohol misuse is associated with both HIV transmission risk as well as poor HIV treatment outcomes but is often underreported [25,33]. In Uganda and Namibia, an electronic screening approach was designed specifically to increase reporting of alcohol misuse and, in Namibia, was then coupled with brief intervention. In both studies, evidence of feasibility was apparent, although with concerns about confidentiality (ie, providers being aware of drinking levels), which hampered acceptability [25,33]. These types of initiatives will provide important additional tests of effectiveness of eSBI to reduce alcohol misuse in HIV care environments and, given our findings, may also hold promise for reducing the risk for ongoing HIV transmission (ie, via sexual risk reduction).

Conclusions

This study found a lack of longer-term impact of a single time point intervention, eSBI, on alcohol misuse among YMSM and YTW presenting to community-based HIV testing clinics. The intervention showed an effect in reducing one of the sexual risk outcomes (number of condomless IAS acts while under the influence). Given that this intervention was embedded within larger systems of care (ie, placed within HIV testing clinics without the need for staffing), there would appear to be some value in considering further trials in this space. Further adaptation of the eSBI approach might result in greater impact. This eSBI intervention was embedded within an STTR model

of care that has been shown to be highly effective in addressing HIV risk and engaging individuals in care. The study showed excellent recruitment and retention of the sample over a 1-year follow-up period. Our eSBI system can be considered a branch of the broader SBIRT model of care. This model has come under some criticism in recent years due to a lack of compelling evidence to show its impact as a primary prevention and intervention approach. Our eSBI system may have failed to achieve the intended goals of alcohol misuse reduction because of the brevity of the intervention, the lack of boosters or follow-up, and the relatively simple design of the intervention. Regardless of the reason, this study falls into alignment with other negative trials of SBIRT. Future research may need to consider relational agents, digital interventions hybridized with human counselors or therapists, and interventions with boosters of some type. In addition, interventions that can address contextual factors (interpersonal, social, and environmental) may help boost the effect of digital interventions like eSBI. Given the instability of results for moderation of the intervention effect by symptoms of anxiety, future studies may need larger samples or be limited to the subpopulation of individuals with higher levels of anxiety to explore these effects more fully.

Acknowledgments

This study is funded by the National Institute on Drug Abuse of the National Institutes of Health (NIH) under Award Number R01-DA041071. Research reported in this publication was also supported by the NIH's National Center for Advancing Translational Sciences, Grant Number UL1-TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We would like to thank member of the Step Up, Test Up team at all study sites for their contribution to study protocol and procedures. Special thanks to Marbella Cervantes for assistance with data cleaning and management.

Authors' Contributions

NSK, RG, and LMK contributed to the funding and design of all aspects of the study. GD, JS, KKB, and MM contributed to the study design and selection of measures. ALH designed the proposed data analysis, and ALH and NDV completed the analysis and created all tables. NSK and LMK drafted the manuscript. The manuscript has been read and approved by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1: generalized estimating equation (GEE) estimates of treatment effect unadjusted for baseline difference in alcohol use.
[DOCX File , 28 KB - mental_v10i1e43653_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 95 KB - mental_v10i1e43653_app2.pdf]

References

1. Brennan J, Kuhns LM, Johnson AK, Belzer M, Wilson EC, Garofalo R, Adolescent Medicine Trials Network for HIV/AIDS Interventions. Syndemic theory and HIV-related risk among young transgender women: the role of multiple, co-occurring health problems and social marginalization. Am J Public Health 2012 Sep;102(9):1751-1757 [FREE Full text] [doi: 10.2105/AJPH.2011.300433] [Medline: 22873480]
2. Lee JS, Safren SA, Bainter SA, Rodríguez-Díaz CE, Horvath KJ, Blashill AJ. Examining a syndemics network among young Latino men who have sex with men. Int J Behav Med 2020 Feb;27(1):39-51 [FREE Full text] [doi: 10.1007/s12529-019-09831-1] [Medline: 31820288]
3. Mimiaga MJ, Hughto JMW, Biello KB, Santostefano CM, Kuhns LM, Reisner SL, et al. Longitudinal analysis of syndemic psychosocial problems predicting HIV risk behavior among a multicity prospective cohort of sexually active young transgender women in the United States. J Acquir Immune Defic Syndr 2019 Jun 01;81(2):184-192 [FREE Full text] [doi: 10.1097/QAI.0000000000002009] [Medline: 30839380]
4. Mustanski B, Phillips G, Ryan DT, Swann G, Kuhns L, Garofalo R. Prospective effects of a syndemic on HIV and STI incidence and risk behaviors in a cohort of young men who have sex with men. AIDS Behav 2017 Mar;21(3):845-857 [FREE Full text] [doi: 10.1007/s10461-016-1607-3] [Medline: 27844298]
5. Teixeira da Silva D, Bouris A, Voisin D, Hotton A, Brewer R, Schneider J. Social networks moderate the syndemic effect of psychosocial and structural factors on HIV risk among young Black transgender women and men who have sex with men. AIDS Behav 2020 Jan 9;24(1):192-205. [doi: 10.1007/s10461-019-02575-9] [Medline: 31289985]
6. Robbins T, Wejnert C, Balaji AB, Hoots B, Paz-Bailey G, Bradley H, NHBS-YMSM Study Group. Binge drinking, non-injection drug use, and sexual risk behaviors among adolescent sexual minority males, 3 US cities, 2015. J Urban Health 2020 Oct;97(5):739-748 [FREE Full text] [doi: 10.1007/s11524-020-00479-x] [Medline: 32918154]

7. Centers FDC. Diagnoses of HIV Infection in the United States and Dependent Areas 2019. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance/vol-32/index.html [accessed 2023-03-18]
8. Kuhns LM, Karnik N, Hotton A, Muldoon A, Donenberg G, Keglovitz K, et al. A randomized controlled efficacy trial of an electronic screening and brief intervention for alcohol misuse in adolescents and young adults vulnerable to HIV infection: step up, test up study protocol. BMC Public Health 2020 Jan 08;20(1):30 [FREE Full text] [doi: 10.1186/s12889-020-8154-6] [Medline: 31914971]
9. Chandler RK, Kahana SY, Fletcher B, Jones D, Finger MS, Aklin WM, et al. Data collection and harmonization in HIV research: the seek, test, treat, and retain initiative at the National Institute on Drug Abuse. Am J Public Health 2015 Dec;105(12):2416-2422. [doi: 10.2105/AJPH.2015.302788] [Medline: 26469642]
10. Babor T, Higgins-Biddle J, Saunders J, Monteiro M. AUDIT : the Alcohol Use Disorders Identification Test : guidelines for use in primary health care. World Health Organization. 2001. URL: https://www.who.int/publications/i/item/WHO-MSD-MSB-01.6a [accessed 2023-03-18]
11. Collins RL, Parks GA, Marlatt GA. Social determinants of alcohol consumption: the effects of social interaction and model status on the self-administration of alcohol. J Consult Clin Psychol 1985 Apr;53(2):189-200. [doi: 10.1037//0022-006x.53.2.189] [Medline: 3998247]
12. Morean ME, Corbin WR. Subjective alcohol effects and drinking behavior: the relative influence of early response and acquired tolerance. Addict Behav 2008 Oct;33(10):1306-1313. [doi: 10.1016/j.addbeh.2008.06.007] [Medline: 18619740]
13. Murphy JG, McDevitt-Murphy ME, Barnett NP. Drink and be merry? Gender, life satisfaction, and alcohol consumption among college students. Psychol Addict Behav 2005 Jun;19(2):184-191. [doi: 10.1037/0893-164X.19.2.184] [Medline: 16011389]
14. Donenberg GR, Emerson E, Bryant FB, Wilson H, Weber-Shifrin E. Understanding AIDS-risk behavior among adolescents in psychiatric care: links to psychopathology and peer relationships. J Am Acad Child Adolesc Psychiatry 2001 Jun;40(6):642-653 [FREE Full text] [doi: 10.1097/00004583-200106000-00008] [Medline: 11392341]
15. Garofalo R, Hotton AL, Kuhns LM, Gratzer B, Mustanski B. Incidence of HIV infection and sexually transmitted infections and related risk factors among very young men who have sex with men. J Acquir Immune Defic Syndr 2016 May 01;72(1):79-86 [FREE Full text] [doi: 10.1097/QAI.0000000000000933] [Medline: 26745827]
16. Garofalo R, Kuhns LM, Reisner SL, Biello K, Mimiaga MJ. Efficacy of an empowerment-based, group-delivered HIV prevention intervention for young transgender women: the Project LifeSkills randomized clinical trial. JAMA Pediatr 2018 Oct 01;172(10):916-923 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.1799] [Medline: 30105381]
17. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
18. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10(2):77-84. [Medline: 8037935]
19. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
20. Ondersma SJ, Svikis DS, Thacker LR, Beatty JR, Lockhart N. Computer-delivered screening and brief intervention (e-SBI) for postpartum drug use: a randomized trial. J Subst Abuse Treat 2014 Jan;46(1):52-59 [FREE Full text] [doi: 10.1016/j.jsat.2013.07.013] [Medline: 24051077]
21. Bendtsen M. An electronic screening and brief intervention for hazardous and harmful drinking among Swedish university students: reanalysis of findings from a randomized controlled trial using a Bayesian framework. J Med Internet Res 2019 Dec 17;21(12):e14420 [FREE Full text] [doi: 10.2196/14420] [Medline: 31845903]
22. Johnson NA, Kypri K, Saunders JB, Saitz R, Attia J, Latter J, et al. Effect of electronic screening and brief intervention on hazardous or harmful drinking among adults in the hospital outpatient setting: A randomized, double-blind, controlled trial. Drug Alcohol Depend 2018 Oct 01;191:78-85. [doi: 10.1016/j.drugalcdep.2018.06.030] [Medline: 30096637]
23. Ganz T, Braun M, Laging M, Schermelleh-Engel K, Michalak J, Heidenreich T. Effects of a stand-alone web-based electronic screening and brief intervention targeting alcohol use in university students of legal drinking age: A randomized controlled trial. Addict Behav 2018 Feb;77:81-88. [doi: 10.1016/j.addbeh.2017.09.017] [Medline: 28985586]
24. Nayak MB, Kaskutas LA, Mericle AA. Randomized trial of an innovative electronic screening and brief intervention for reducing drinking among women of childbearing age. J Addict Med 2019;13(6):450-459 [FREE Full text] [doi: 10.1097/ADM.0000000000000518] [Medline: 30882553]
25. Tang AM, Hamunime N, Adams RA, Kanyinga G, Fischer-Walker C, Agolory S, et al. Introduction of an alcohol-related electronic screening and brief intervention (eSBI) program to reduce hazardous alcohol consumption in Namibia's antiretroviral treatment (ART) program. AIDS Behav 2019 Nov;23(11):3078-3092 [FREE Full text] [doi: 10.1007/s10461-019-02648-9] [Medline: 31444711]
26. Goodness TM, Palfai TP. Electronic screening and brief intervention to reduce cannabis use and consequences among graduate students presenting to a student health center: A pilot study. Addict Behav 2020 Jul;106:106362. [doi: 10.1016/j.addbeh.2020.106362] [Medline: 32120198]

27. Jonas DE, Garbutt JC, Amick HR, Brown JM, Brownley KA, Council CL, et al. Behavioral counseling after screening for alcohol misuse in primary care: a systematic review and meta-analysis for the U.S. Preventive Services Task Force. Ann Intern Med 2012 Nov 06;157(9):645-654 [FREE Full text] [doi: 10.7326/0003-4819-157-9-201211060-00544] [Medline: 23007881]
28. Saitz R. Screening and brief intervention enter their 5th decade. Subst Abus 2007;28(3):3-6. [doi: 10.1300/J465v28n03_02] [Medline: 18077299]
29. Saitz R. 'SBIRT' is the answer? Probably not. Addiction 2015 Sep;110(9):1416-1417. [doi: 10.1111/add.12986] [Medline: 26223169]
30. Bickmore T, Gruber A. Relational agents in clinical psychiatry. Harv Rev Psychiatry 2010;18(2):119-130. [doi: 10.3109/10673221003707538] [Medline: 20235777]
31. Grekin ER, Beatty JR, McGoron L, Kugler KC, McClure JB, Pop DE, et al. Testing the efficacy of motivational strategies, empathic reflections, and lifelike features in a computerized intervention for alcohol use: A factorial trial. Psychol Addict Behav 2019 Sep;33(6):511-519 [FREE Full text] [doi: 10.1037/adb0000502] [Medline: 31436446]
32. Prochaska JJ, Vogel EA, Chieng A, Baiocchi M, Maglalang DD, Pajarito S, et al. A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. Drug Alcohol Depend 2021 Oct 01;227:108986 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108986] [Medline: 34507061]
33. Emenyonu N, Kekibiina A, Woolf-King S, Kyampire C, Fatch R, Dawson-Rose C, et al. Digital health screening in people with HIV in Uganda to increase alcohol use reporting: qualitative study on the development and testing of the self-administered digital screener for health. JMIR Form Res 2022 Sep 01;6(9):e35015 [FREE Full text] [doi: 10.2196/35015] [Medline: 36048519]

Abbreviations

AOD: alcohol and other drugs
AUDIT: Alcohol Use Disorders Identification Test
CASI: computer-assisted self-interviewing
CES-D10: Center for Epidemiologic Studies Depression Scale 10
CSQ-8: Client Satisfaction Questionnaire
DDQ-R: Daily Drinking Questionnaire Revised
eSBI: electronic screening and brief intervention
GAD-7: Generalized Anxiety Disorder 7
GEE: generalized estimating equation
IAS: insertive anal sex
IRR: incidence rate ratio
MI: motivational interviewing
PrEP: pre-exposure prophylaxis
RAS: receptive anal sex
RR: relative risk
SAMHSA: Substance Abuse and Mental Health Services Administration
SBIRT: screening, brief intervention, and referral to treatment
STI: sexually transmitted infection
STTR: seek, test, treat, and retain
YMSM: young men who have sex with men
YTW: young transgender women

Edited by J Torous, G Eysenbach; submitted 19.10.22; peer-reviewed by S Naar, D Levine; comments to author 18.01.23; revised version received 16.02.23; accepted 02.03.23; published 29.03.23.

Please cite as:
Karnik NS, Kuhns LM, Hotton AL, Del Vecchio N, McNulty M, Schneider J, Donenberg G, Keglovitz Baker K, Diskin R, Muldoon A, Rivera J, Summersett Williams F, Garofalo R
Findings From the Step Up, Test Up Study of an Electronic Screening and Brief Intervention for Alcohol Misuse in Adolescents and Young Adults Presenting for HIV Testing: Randomized Controlled Efficacy Trial
JMIR Ment Health 2023;10:e43653
URL: https://mental.jmir.org/2023/1/e43653
doi:10.2196/43653
PMID:36989027

Original Paper

Developing a Theoretically Informed Implementation Model for Telemedicine-Delivered Medication for Opioid Use Disorder: Qualitative Study With Key Informants

Joseph Tay Wee Teck[1], MD, MSc; Rosalind Gittins[2], MPharm; Giedre Zlatkute[1], PhD; Alberto Oteo Pérez[1], PhD; Susanna Galea-Singer[3], MD; Alexander Baldacchino[1], MD, PhD

[1]Digital Health Interventions in Addiction Services Project, Population and Behavioural Science, School of Medicine, University of St Andrews, St Andrews, United Kingdom

[2]Via, London, United Kingdom

[3]NHS Fife Addiction Services, Leven, United Kingdom

Corresponding Author:
Joseph Tay Wee Teck, MD, MSc
Digital Health Interventions in Addiction Services Project, Population and Behavioural Science
School of Medicine
University of St Andrews
North Haugh
St Andrews, KY16 9TF
United Kingdom
Phone: 44 07983344674
Email: jbtwt1@st-andrews.ac.uk

Abstract

Background: Telemedicine-delivered medication for opioid use disorder (TMOUD) has become more prevalent during the COVID-19 pandemic, particularly in North America. This is considered a positive development as TMOUD has the potential to increase access to evidence-based treatment for a population heavily affected by the opioid crisis and consequent rising mortality and morbidity rates in relation to opioid use disorder. Despite the increase in the use of TMOUD, there are no established service- and process-focused models to guide the implementation of this intervention.

Objective: This study aims to develop a process- and service-focused implementation model in collaboration with key stakeholders and bring together peer-reviewed literature, practice-based knowledge, and expert opinions.

Methods: The simple rules for evidence translation in complex systems framework was applied to guide the development of a 6-step qualitative study. The steps were definition of the scope and objectives of the model, identification of evidence, stakeholder engagement, draft model development, key informant consultation, and final model specification.

Results: The final specification for the TMOUD implementation model incorporated key strategic priorities, service delivery prerequisites, service design elements, stakeholder identification and engagement, key process domains, and iterative cycles of evaluation and improvement.

Conclusions: Through stakeholder engagement and key informant consultation, we produced a process- and service-focused TMOUD implementation model. The model is modifiable to different contexts and settings while also in keeping with the current evidence base and national and international standards of high-quality opioid use disorder care.

(JMIR Ment Health 2023;10:e47186) doi:10.2196/47186

KEYWORDS

telemedicine; medication for opioid use disorder; implementation model; telebuprenorphine; opioid use disorder; mobile phone

Introduction

Background

Scotland has been experiencing a rising trend in drug-related deaths (DRDs) for 2 decades [1]. In 2020, the per capita DRD rate in Scotland was 245 deaths per million population, the highest in the European Union and approaching US rates of 277 deaths per million population [2]. In June 2019, the Scottish government established the Drug Deaths Task Force to identify policies, measures, and interventions that could reduce DRDs and harms [3]. A key task force recommendation was the introduction of evidence-based standards to enable the delivery of high-quality, safe, timely, and accessible drug treatments across Scotland, incorporating medication for opioid use disorder (MOUD) [4].

The Scottish government's decision to improve the provision and quality of MOUD was in line with the international evidence base and priorities for reducing DRDs [5]. MOUD, such as methadone and buprenorphine, can notably reduce overdose risk and associated deaths, thus improving health and social outcomes, particularly when patients remain on treatment [6,7]. Nevertheless, internationally, only 1 in every 6 people seeking treatment for opioid use disorder (OUD) receives it [8]. Several factors contribute to this so-called treatment gap [9], including stigma [10], limited clinical resources, insufficient investment, ambivalence toward and difficulties accessing MOUD [11-14], and stringent regulatory barriers [15]. Furthermore, vulnerable groups such as younger people [16], pregnant women, mothers [17,18], rural residents [19], racial and ethnic minority groups [20], and people experiencing displacement [21,22] or homelessness [23] disproportionately struggle to access MOUD. Consequently, the Scottish government's first 5 standards for medication-assisted treatment (MAT), which focuses on MOUD, include provision for same-day access, choice of medication, assertive outreach and anticipatory care, harm reduction, and retention in treatment [24] (Textbox 1).

Textbox 1. Scottish medication-assisted treatment (MAT) standards 1 to 5, which focus on the provision of medication for opioid use disorder.

Standard 1

- All the people accessing services have the option to start MAT from the same day of presentation.

Standard 2

- All the people are supported to make an informed choice on what medication to use for MAT and the appropriate dose.

Standard 3

- All the people at high risk of drug-related harm are proactively identified and offered support to commence or continue MAT.

Standard 4

- This standard comprises the number of services offering evidence-based harm reduction at the point of MAT delivery.

Standard 5

- All the people will receive support to remain in treatment for as long as requested.

To embed these standards across Scotland by April 2022, the Scottish government funded several demonstration projects, a research program, and an innovation fund to address gaps in knowledge and service delivery [25]. The COVID-19 pandemic created significant disruptions to this work program [25] yet created the opportunity for an exploration of the use of telemedicine in delivering MOUD. Telemedicine-delivered MOUD (TMOUD) has shown promise in addressing the treatment gap by increasing immediacy and responsiveness; reducing geographical and logistical barriers, stigma, and costs; enabling more consistent and regular patient contact; and increasing service efficiency [26-31]. In response to service closures, reduced staffing capacity, and the introduction of measures to reduce COVID-19 transmission, some countries introduced regulatory, funding, and clinical governance changes allowing for TMOUD to be used more widely [32-37]. This was particularly the case in the United States and Canada, where a rapid acceleration of TMOUD adoption was observed [32].

In October 2020, the Digital Health Interventions in Addiction Services project at St Andrews University received funding through the Scottish Drug Deaths Taskforce and the Corra Foundation to explore ways of developing and implementing TMOUD as part of high-quality drug treatment in Scotland. This project was timely as we were able to observe and learn from international experiences with TMOUD implementation in response to the COVID-19 pandemic. Indeed, an early finding from the accelerated and necessary transitions to telemedicine to provide MOUD was that it created uncertainty and anxiety among clinicians and service providers in a variety of settings and countries [36,38]. This largely regarded issues related to governance; regulatory compliance; safety; confidentiality; and best practice approaches to history taking, physical examination, and investigations such as drug and blood-borne virus testing [39].

Objectives

From an implementation science perspective, such contextual, operational, and clinical concerns are key pivot points determining the success or failure of new TMOUD services [40]. However, to our knowledge, there are no specific implementation frameworks or models to facilitate the embedding of TMOUD into standard drug treatment practice.

The literature on TMOUD before the COVID-19 pandemic tended to comprise retrospective data analyses comparing the efficacy of in-person versus telemedicine delivery of MOUD [41], with a steep rise in publications from 2020 onward describing pandemic-related rapid service transitions to TMOUD [32]. Consequently, our aim with this work was to describe the development of a theoretically informed TMOUD implementation model.

Methods

Setting

Addiction treatment services are complex systems of care with multiple interdependent components and autonomous actors whose agency in performing their everyday roles can hamper or disrupt attempts to introduce sustainable change [42-44]. Furthermore, the delivery of MOUD both in Scotland and internationally occurs within a highly regulated clinical guidance, governance, and legislative framework associated with a high level of scrutiny on compliance from both people with OUD seeking treatment and treatment providers [45]. In the context of this TMOUD project, at least 3 additional layers of intersecting complexity were identified.

First, initial actions in 2020 to introduce and implement quality standards for MOUD delivery in Scotland were enabling in nature, involving stakeholder engagement, quality improvement support, funding for innovative pilot projects, development of processes and tools, and their wide dissemination [24]. Unfortunately, a progress report in March 2021 [24], echoed by a national benchmarking process [46], identified significant barriers to services achieving these standards, including challenges with leadership and financial planning, conflicting local priorities, and a lack of transparency and evaluation data [24,46]. This culminated in the Scottish government using legislative powers to compel health boards, integration authorities, and local authorities to drive the changes necessary to deliver on these quality standards, particularly same-day access to MOUD [47]. Improving access to and retention of MOUD will inevitably result in growing numbers in treatment, and services will need to adapt, evolve, and innovate to provide ongoing, safe, and effective care. Both the United States and Canada, with similar drug death crises, have developed models of TMOUD to improve service efficiency and the capacity to expand and meet the demand for treatment [48].

Second, Scotland has the necessary telemedicine-specific infrastructure, policy, and governance frameworks for TMOUD implementation. Specifically, there has been a strategic development plan for telemedicine and telecare in Scotland since 2012 [49] and an established platform to deliver telemedicine through the Attend Anywhere/Near Me video consulting service since 2016 supported by an implementation framework provided by Technology Enabled Care Scotland [50]. Indeed, this background work meant that Scotland was able to successfully introduce a rapid rollout initiative of video consulting in the context of the COVID-19 pandemic [51]. Unfortunately, addiction services consistently lagged behind other health specialties in using the Near Me video consulting service both before and after the rapid rollout initiative. When compared with psychiatry, psychology, and community mental health services, for example, addiction services made up 0.77% (44/5745) versus 9.5% (546/5745) of telemedicine consultations from January 2019 to December 2019 [50], and this number reduced further after the rapid rollout in response to the pandemic with 0.25% (651/260,547) in contrast to 36.41% (94,876/260,547) [52] of telemedicine consultations. Nevertheless, although the Scottish government mandated quality standards for MOUD delivery, there were no targeted financial or policy drivers to incentivize the uptake of TMOUD.

Third, the peer-reviewed evidence base for TMOUD has expanded greatly and continued to grow since the COVID-19 pandemic [48]. There were nearly 4 times more peer-reviewed publications on TMOUD in January 2023 compared with before 2020 [48]. This evidence base is dominated by US-based research, which differs in critical ways from the Scottish and UK contexts [48]. Key differences include a highly restricted approach to methadone provision and the preponderance of buprenorphine as a first-line MOUD in the United States as opposed to the United Kingdom, where both medications are available at community pharmacies with a roughly 60:40 split in favor of methadone [53]. Consequently, it was necessary to find ways to adapt existing knowledge to local processes and iteratively review the growing evidence base as knowledge gaps emerged throughout this project.

Theoretical Underpinning: Applying the Successful Healthcare Improvements From Translation of Evidence Into Practice Framework

The setting of this TMOUD project necessitated a theoretical framework suited to managing complexity, and the Successful Healthcare Improvements From Translation of Evidence Into Practice (SHIFT-Evidence) framework was identified as particularly comprehensive in this regard [54]. This framework describes 3 strategic principles and 12 simple rules on how to make sense of and intervene in complex systems with a focus on service delivery and processes [54]. The first strategic principle, to act scientifically and pragmatically, is an exhortation to tailor and iteratively adapt implementation approaches to match local contexts, problems, and opportunities [54]. The second principle, to embrace complexity, essentializes the need to fully understand usual care practices and processes and identify and address existing and emergent issues through evidence translation efforts [54]. The third principle, to engage and empower, highlights the importance of aligning evidence translation with the concerns and motivations of the people who underpin, perform, and engage with real-world processes and practices [54]. Associated with these 3 principles are 12 simple rules that have been applied to design this study, described in detail in Multimedia Appendix 1.

Research Design

We used a logic model [55] to describe the overall approach taken to develop this TMOUD implementation model (Multimedia Appendix 2). The logic model is based on the application of the 12 SHIFT-Evidence framework rules detailed in Multimedia Appendix 1.

Step 1: Define the Scope and Objectives of the Implementation Model

This model was intended to focus on process development and improvement (eg, workflows, role allocation, infrastructure, and resources) to facilitate TMOUD implementation rather than define ideal clinical standards or predict or analyze barriers and facilitators. A key tenet of the project was to promote a participatory approach such that the implementation model produced would be informed by clinical expertise, professional values, and obligations; customizable to specific contexts; and implementable through collective effort rather than a top-down imposition. In keeping with theories that explain the "kinds of work" necessary to successfully implement an intervention, such as normalization process theory [56], the model needed to characterize TMOUD as distinct from standard MOUD practices, identify critical stakeholder relationships and perspectives, and describe necessary processes and actions and appraisal work for iterative improvements [56].

Step 2: Identification of Evidence to Inform the Implementation Model

As discussed previously, the evidence base for TMOUD has been rapidly evolving and growing, and implementation- or practice-specific knowledge has been lacking. We addressed this issue by carrying out a series of scoping reviews [48,57] that are flexible, inclusive, and iterative in nature [58] along with the sourcing of expert case studies through addiction professional networks such as the International Society of Addiction Medicine and the Canadian Society of Addiction Medicine.

Step 3: Stakeholder Identification and Engagement

In Scotland, local-level drug and alcohol strategy and policy implementation is devolved to 31 Alcohol and Drug Partnerships (ADPs), which work across statutory and third-sector health and social care providers. We identified stakeholders through ADPs but also through national advocacy organizations; drug and alcohol research networks; and professional organizations such as the Royal College of Psychiatrists, Royal College of General Practitioners, College of Mental Health Pharmacy, and the Royal Pharmaceutical Society. As the use of TMOUD was not prominent on the national agenda and local initiatives were nonexistent, we launched the participatory component of this research through a web-based engagement event linking stakeholders with an international panel of TMOUD experts. Each presenter was a service provider from the United States, Canada, England, Ireland, or Scotland and provided a 10-minute presentation on their experiences in implementing aspects of TMOUD. A summary of the topics and content covered is provided in Multimedia Appendix 3, and recordings of the presentations are available [59]. Stakeholders were able to engage in discussions with expert panelists in a synchronous chat. Participants at the engagement event were invited to indicate whether they wished to take part in the future TMOUD implementation model development.

Step 4: Produce a Draft Implementation Model for Consultation

The first iteration of the draft implementation model was based on a review of policy documents [4,60,61], information collected from our expert panel, synchronous discussion transcripts from the engagement seminar, and scoping reviews conducted to identify examples of program implementation and models and processes underpinning TMOUD delivery [48,57].

Step 5: Key Informant (Stakeholder) Consultation

The draft implementation model was made available in an electronic PDF file (Multimedia Appendix 4) as well as an interactive web-based collaborative document (Multimedia Appendix 5). Participants from the stakeholder engagement event who had indicated an interest in commenting on the draft model were invited to sign up to do this via a unique link to the web-based collaboration platform. This web-based platform allowed participants to provide asynchronous comments on the same interactive document, adding a layer of transparency to the overall process. This was not convenient for all our participants as some could not access the web-based platform owing to organizational firewalls, whereas others preferred to retain a degree of anonymity when commenting. A PDF copy of the draft document was emailed to these participants if so requested. Comments, conflicts, and proposed amendments were incorporated into a revised document, which was then discussed at a final roundtable meeting with key implementation decision makers and leaders.

Step 6: Produce a Final Specification of the Implementation Model

Following the key informant and stakeholder consultation process, the updated document was combined with additional findings of a literature review carried out throughout the research period. This confluence of research and practice-based evidence and expert opinions formed the final specification of the TMOUD implementation model.

Ethics Approval

Ethics approval was granted by the St Andrews University Teaching and Research Ethics Committee (MD15635).

Results

Key Informant (Stakeholder) Consultation

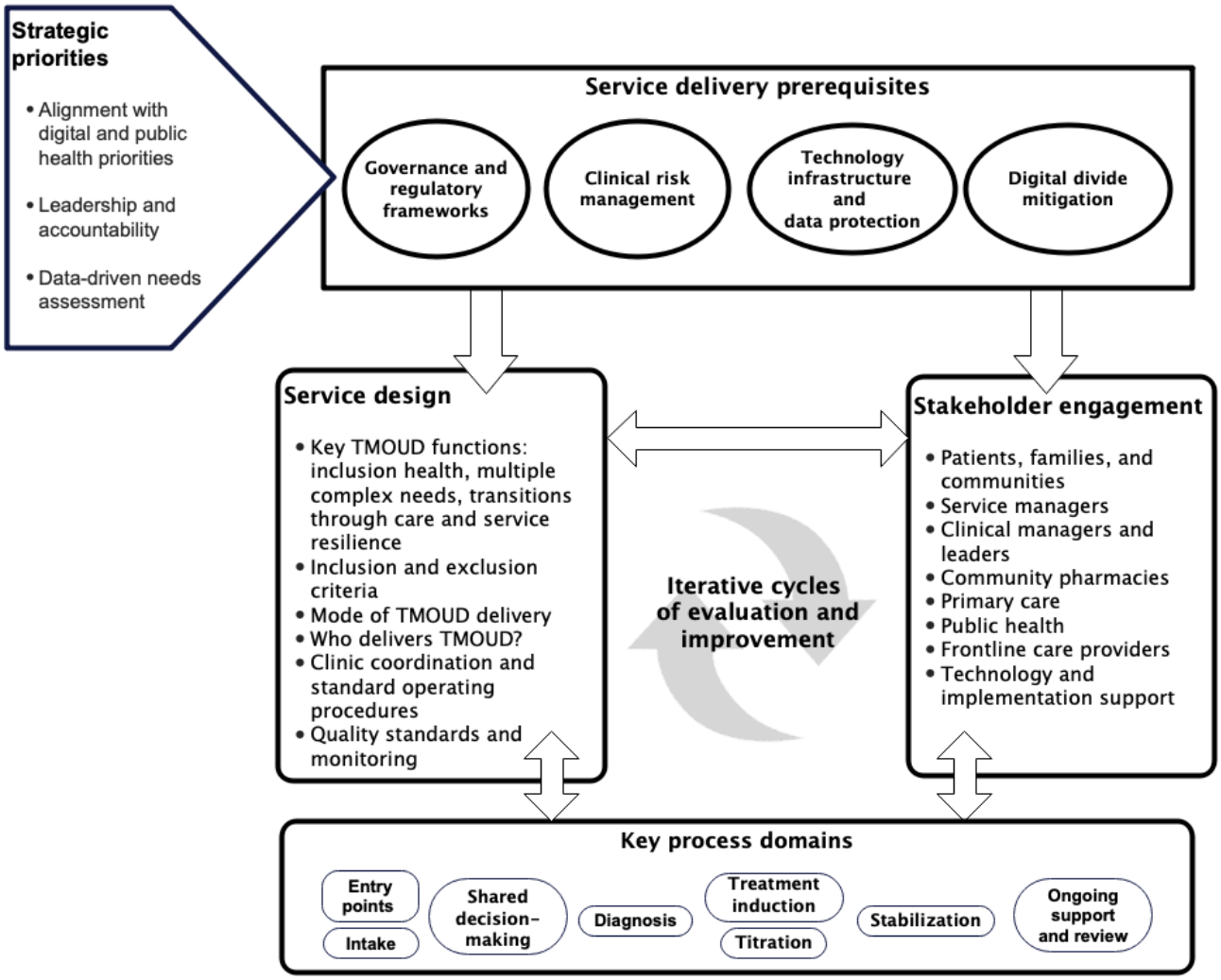
There was diverse representation across key stakeholder groups in our initial engagement event and subsequent participatory activities. A total of 264 participants logged on to the engagement event, including National Health Service representatives (n=50, 18.9%), third-sector care providers (n=42, 15.9%), participants from policy and patient advocacy organizations and international nongovernmental organizations, pharmaceutical representatives, non-UK addiction care providers, participants from the Scottish government, and ADP representatives. A total of 10.6% (28/264) of the people attending the engagement event actively interacted with the panelists through chat room discussions. In total, 71 participants were invited to review the first draft of the implementation model, with 18 (25%) representing diverse stakeholders actively

contributing. A total of 5.7% (15/264) of the participants took part in the final roundtable discussion. This roundtable meeting was focused on obtaining input from decision makers and clinical leaders on gaps and vulnerabilities within the draft implementation model. Multimedia Appendix 6 details the stakeholder roles represented in each of the participatory activities we conducted.

Final Specification of the TMOUD Implementation Model

Figure 1 provides a graphical representation of the TMOUD implementation model. The key components of the model were strategic priorities, service delivery prerequisites, service design factors, stakeholder engagement, key process domains, and iterative cycles of evaluation and improvement.

Figure 1. Telemedicine-delivered medication for opioid use disorder (TMOUD) implementation model.



Strategic Priorities

Strategic Alignment With Regional Digital Health Policy and Public Health Priorities

There was broad recognition among key informants that prospective TMOUD projects brought together the public health priority of reducing DRD rates and national digital health strategies. Critically, local innovation ecosystems are frequently linked to national digital health policy, and an understanding of these by addiction service providers can facilitate the internal buy-in of TMOUD implementation processes. More specifically, significant investments in digital infrastructure, training, and other resources were readily available as a consequence of national digital health policies [61], ready to support local TMOUD adoption. Some key informants believed that supporting the engagement of addiction services and clinicians with other specialties such as urgent care, which was already making use of telemedicine, and introducing telemedicine champions and technology training programs may help build confidence in TMOUD implementation.

Leadership and Accountability

Leadership and strategic direction were identified as critically important to the implementation of TMOUD [48,62-64]. For example, early guidance from the Irish College of General Practitioners in conjunction with the National Health Service Executive Office for Social Inclusion [65] reassured clinicians who were concerned about divergence from standard in-person practice when delivering TMOUD [65]. One of the key informants with telemedicine implementation experience reinforced the value of organizational-level clinical leadership to gain an understanding of strengths, infrastructure, and risks; adapt a generic telemedicine model to the needs of the local

service providers and population; and engage frontline clinicians in TMOUD processes, management, and delegation, a perspective supported by the literature [62,64,66].

Data-Driven Needs Assessment

In total, 2 areas requiring a needs assessment were highlighted. The first was to understand the geographic distribution of people who use drugs, including travel time to service, availability of a dispensing pharmacy, transportation availability and costs, broadband coverage, service and clinician capacity, and the person's preference for contact type. Some of these data may be readily available from existing geo-mapping initiatives, whereas other data may require active engagement with local communities and auditing the service. The second aspect involved understanding the needs of local people and what may be missed when a service is transferred to telemedicine. Several examples of this were provided by key informants, for example, blood-borne virus testing and treatment, sexual health, housing support, overdose awareness, and take-home naloxone provision. Specifically, TMOUD services need to consider how more holistic aspects of drug treatment services can continue within a telemedicine model of care.

Service Delivery Prerequisites

Governance and Regulatory Frameworks

Variability in jurisdictional legal and professional regulatory frameworks and local guidelines and practice regarding the prescription of controlled drugs such as methadone and buprenorphine may act as barriers to TMOUD implementation [63,66]. These frameworks are not only often legally mandated but are also often perceived as synonymous with good clinical practice [67]. Examples of regulatory restrictions that require particular attention in the development of TMOUD include regional prescription drug monitoring programs to limit the risk of inappropriate provision of a medication that could cause dependency and mandatory in-person consultations before the provision of MOUD [27,68,69]. However, growing evidence has shown that overregulation results in barriers to MOUD provision, pushing some patients into the illicit market to manage opioid withdrawal [70].

To mitigate the impact of the COVID-19 pandemic, many jurisdictions introduced changes to regulatory, funding, and clinical governance policies regarding MOUD delivery [71]. Areas that allowed more extensive regulatory easing and policy flexibility also had more developed and successful TMOUD implementation [30,32,36,37,64]. For example, because of nonnegotiable prescribing arrangements for several controlled drugs, including methadone, India was unable to implement TMOUD in any meaningful way [38] compared with its North American counterparts [27,64,69,72]. Other important regulatory considerations include reimbursement arrangements and payment parity between telemedicine and in-person consultations [27,68,69,73,74], information governance safeguards [38,68], the availability of sanctioned telemedicine platforms [66,75], and medical indemnity coverage for physicians [62,68].

It follows, then, that an overarching priority must be for services to identify the policy and legislative limits to TMOUD implementation and whether temporary or enduring exemptions or memoranda of understanding are in place or negotiable. For example, the US TMOUD expansion throughout the COVID-19 pandemic was almost certainly facilitated by payment parity approvals for telemedicine by health insurance companies [32], permission for a broad range of videoconferencing tools to be used for telemedicine, and waiving the requirement for an in-person examination before controlled substances were prescribed [75]. Addiction care providers in the United States have been keen to document the value of pandemic-related regulatory easing in narrowing inequalities to MOUD access and the OUD treatment gap [66] and work to sustain the environment that made TMOUD implementation so successful [76]. A similar advocacy approach may be valuable to other jurisdictions keen to develop locally appropriate TMOUD services.

One key informant highlighted that addiction care providers may not always fit within the same governance framework; thus, the limitations, barriers, and facilitators may not be congruent. For example, in Scotland, taxpayer-funded and statutorily governed National Health Service addiction care providers use a single validated secure communications platform and have specific reimbursement arrangements for MOUD [77]. In England, it is often third-sector organizations that are commissioned by local authorities to provide addiction services, each having its own negotiated service-level contracts and funding arrangements [78]. Similarly, in Canada, the decentralization of methadone regulation to provinces in 1995 allowed each jurisdiction to adapt according to local needs and geographic and staffing practices, resulting in some areas, such as Ontario, being able to deliver TMOUD since 2013 [26]. However, in the United States, the complexities of local variations in legislation and reimbursement policies, together with overarching federal laws such as the Ryan Haight Online Pharmacy Consumer Protection Act of 2008, have resulted in provider organizations having to expend considerable effort to ensure they do not inadvertently break the law when providing TMOUD [68].

Clinical Risk Management

Clinical risk management (CRM) is the process by which health care organizations ensure the quality and safety of the services they provide, identify circumstances that contribute to clinical risk, and act to prevent or control those risks [79]. CRM is typically described as a 4-step process starting with risk identification, analysis of severity and frequency, development and application of a reduction or elimination plan, and iterative evaluation of the impact of this plan on risk reoccurrence [79]. Although CRM is a critical component of TMOUD service design and delivery, 2 main concerns were highlighted. The first was the way in which bureaucratic risk responses by prescribers derived from organizational CRM may widen the digital divide and structural inequalities [48]. Specifically, a scoping review of TMOUD highlighted the practice of only allowing telemedicine to be used for the delivery of MOUD to people who were already established in treatment, excluding those who were new to the service as the perception was that the latter group posed a high risk of indemnity if not seen in person at the outset [48]. This was despite the regulatory easing

that sanctioned buprenorphine induction without an initial in-person visit [80]. The outcome of this in the United States was the exclusion of marginalized and ethnic minority groups from accessing MOUD as they were not already within the treatment system at the pandemic onset. Critically, these groups did not have a mechanism to challenge this decision or highlight the increased risks they experienced as a result [48].

The second issue was related to prescriber attitudes toward risk, which are typically influenced by social relationships, attitudes, trust, values, power dynamics and hierarchies, experience, and knowledge [48]. Consequently, providing evidence of good outcomes or mandating the delivery of TMOUD without addressing the local risk culture would be unlikely to enable successful implementation. A theme among the key informants was the importance of ongoing engagement with stakeholders, including people with OUD, in the design of TMOUD services, risk assessments, and CRM activities. Specifically, stakeholders should be involved in the iterative evaluation of the impact of CRM in reducing risks to individuals in the context of digitally delivered care rather than assuming that the situation remains the same as with in-person care. Consequently, the cornerstone of CRM is a full understanding of existing MOUD processes and the associated or potential risks when transitioning to TMOUD, an area described in detail in the *Key Process Domains* section.

Technology Infrastructure and Data Protection

Issues such as dropped calls, poor-quality images or sound, time lag disrupting the flow of communication, and fluctuating access to bandwidth that interferes with clinical consultations are common in telemedicine consults [36,38,80,81]. Service providers have needed to transfer the consultation to a telephone call in some cases, potentially limiting the amount of information available to support diagnosis and treatment [36,38,80,81]. Furthermore, although the Scottish National Health Service has had access to a nationally approved telemedicine platform with appropriate governance and security measures in place [60], this is not the case for many nonstatutory service providers, who have relied on commercial services such as Skype for business or encrypted Zoom calls. Each of these platforms may have different privacy law compliance and may differ in their integration with electronic health records. Several key informants recommended dedicated resources to ensure that the technology underpinning telemedicine was robustly tested and that troubleshooting support was readily available. Clear guidance and reassurance should be provided to clinicians that consultation platforms are compliant with data protection legislation.

Digital Divide Mitigation

There was an awareness of and efforts to mitigate the digital divide in Scotland [82-84], which encompassed addressing technical skills, confidence, broadband and data poverty, and equipment access and affordability. Some patients with OUD experiencing homelessness may struggle to safeguard or charge devices, find the privacy to join a telemedicine consult, or have the "headspace" to engage with technology because of other more immediate life-critical priorities [85]. Several initiatives to address these more complex barriers to TMOUD were identified, including fixed or mobile telemedicine pods, using bridge or link workers as telemedicine support, and diversifying digital options to include telephone contact.

Mitigation strategies should include an awareness of how clinical practice or policies may directly widen the digital divide. For example, concerns have been raised over audio-only versus video-based telemedicine being less efficient and effective [86]. Regulators in the United States are considering mandating against audio-based TMOUD despite better engagement of people at the extreme margins of society [87,88]. However, the implication for developing video-only models of TMOUD may include the marginalization of those who feel able to use the telephone to receive care even if data-heavy mobile apps or video calls are unaffordable or unappealing.

Another example is the inclusion criterion for accepting patients into a TMOUD program. Clinical inclusion criteria may typically incorporate measures of stability such as housing status, criminal activity, polysubstance use, and the existence of concurrent mental health issues [57]. A TMOUD service that excludes people who are unstably housed or have comorbid health conditions may systematically exclude the most vulnerable among its catchment population [57]. Furthermore, in the United States, people categorized as "unstable" based on these criteria tend to belong to people of color populations, leading to exclusion along racial lines [57].

Stakeholder Engagement

It is important to engage with stakeholders at all levels of service delivery. Nevertheless, depending on the standard model of care in a particular jurisdiction, not all stakeholder groups may play a critical role in TMOUD implementation. Most services will need to engage effectively with leadership and management stakeholders, such as clinical, organizational, operational, and nursing managers, and prescribers, such as medical directors, physicians, physician assistants, and nurse practitioners. However, not all services will have the same community pharmacy arrangements depending on the availability of pharmacy outlets in the region, the existence of locally negotiated service-level agreements, and whether electronic prescriptions for controlled drugs are permissible within that jurisdiction.

Other differences across jurisdictions may be the level of access to behavioral health personnel such as social workers, clinical psychologists, mental health therapists, substance use counselors, and case managers. In addition, some aspects of care may be delivered by organizations without the relevant data-sharing arrangements and access to shared records, for example, community outreach workers or employees of third-sector housing, harm reduction, or advocacy services. Different agencies may have different levels of clinical and IT administrator and program coordinator support. Finally, patients, particularly those not already known to the service, may be at a significant disadvantage as their needs may be entirely invisible to TMOUD service implementors.

What this effectively means is that, after identifying key stakeholders, a protracted and challenging period of engagement and negotiation may be necessary to enable them to play their

roles in TMOUD delivery. Among the most challenging arrangements are those involving community pharmacies and electronic prescriptions. Unlike in the United States, community pharmacies in the United Kingdom may dispense methadone and buprenorphine following the instructions of the prescriber, including the provision of observed dosing [89]. Although this is a significant advantage to MOUD provision in the United Kingdom, the regulatory aspects and reimbursement structure for electronic prescribing and dispensing are currently ambiguous and pose a barrier to TMOUD.

Service Design

Inclusion and Exclusion Criteria

A nuanced approach to identifying suitability for TMOUD is to use criteria stratified according to the risk level and stage of treatment. This should be in addition to regulatory requirements and existing service-level policies. Considerations may include (1) clinical circumstances such as comorbidities, particularly severe mental health diagnoses, pregnancy, polypharmacy, polysubstance use, previous treatment experiences, and frailty, and (2) social and personal circumstances such as the remoteness of the person's location, poor transport access, local COVID-19 restrictions, access to technology, risk of travel versus the risk of overdose, and whether the person lives alone or in supported accommodation. A standard template for triage or suitability for TMOUD should be developed to direct referrals appropriately, and alternative methods of assessment should be identified for people excluded from TMOUD. Multimedia Appendix 7 provides examples of exclusion and inclusion criteria for fully web-based and hybrid TMOUD.

Models of TMOUD Delivery

As the practice of TMOUD has grown throughout the pandemic, several models of care surrounding this intervention have emerged. One of the reviews carried out within this project [57] categorized 4 overlapping models of TMOUD delivery elaborated on in Textbox 2. These were TMOUD to provide an inclusion health focus, facilitate transitions through care, meet complex care needs, and increase service resiliency and quality. Services intending to adopt TMOUD may find that their existing or intended care model matches some elements of each of these categories or is in alignment with one specific category. Defining the key principles of care delivery underpinning TMOUD will have several implications, including the selection of quality and outcome measures of importance. This will be elaborated on in a later section.

Textbox 2. Models of telemedicine-delivered medication for opioid use disorder (TMOUD).

TMOUD model (TM) 1.0: inclusion health focus

- Definition: the model is focused on reaching marginalized, easily ignored groups at the intersections of multiple forms of exclusion, including race, homelessness, criminal justice involvement, and forced displacement.
- Key aspects of care include outreach through street medicine, mobile clinics, health care facilitators, harm reduction workers, telemedicine facilitators, and multicomponent care.
- TM 1.1: places of safety. Care is brought to the patient where they are and feel safe.
- TM 1.2: trauma informed. The service is guided by trauma-informed principles and culturally appropriate, peer-supported care.
- TM 1.3: low threshold. Provides low-threshold, no–wrong door, low-barrier entry into treatment.
- TM 1.4: social determinants of health. The service is sensitive to social determinants of health, including digital exclusion, and mitigates them.

TM 2.0: transitions through care

- Definition: the model is focused on promoting the patient's social stability and continuity of care across jurisdictional boundaries and points of transition where overdose risk may be magnified.
- Key aspects of care include bridge clinics and shared care arrangements among specialist, primary care, and multidisciplinary services such as joint opioid use disorder and perinatal services.
- TM 2.1: patient journey mapping. Comprises the identification of points of high risk of patient attrition, for example, leaving prison; discharge from hospital, emergency department, or treatment service; and becoming homeless or relocating.
- TM 2.2: process mapping and risk mitigation. Removes barriers to accessing a medication for opioid use disorder (MOUD) prescription, clinical review, or MOUD supply through a dispensing pharmacy.
- TM 2.3: chronic condition management. Incorporates principles of managing a chronic relapsing or remitting condition where longitudinal continuity of care reduces the risk of adverse health outcomes.
- TM 2.4: multiple episodes of care. Comprises recognizing, accepting, and allowing for multiple episodes of care across the condition.

TM 3.0: complexity of care needs

- Definition: the model is designed around multiple unmet care needs, including housing; legal and income support; treatment of comorbid physical and mental health conditions; and harm reduction interventions to prevent blood-borne virus transmission, soft tissue infections, and overdose deaths.
- Key aspects of care include care navigators or medical case managers to coordinate interactions with health services, navigate complex bureaucratic systems, create individualized care plans, and support patients in meeting their goals.
- TM 3.1 comprises multidisciplinary and multisectoral care incorporating social workers, counselors, nurses, primary care physicians, case managers, legal advisors, and outreach workers.
- TM 3.2 comprises the flexibility to tailor interventions or programs to different needs, for example, services for people with HIV or Hepatitis C, inclusion health primary care, or street medicine services.
- TM 3.3 comprises education and empowerment of patients through peers or culturally appropriate workers to facilitate self-advocacy.

TM 4.0: increasing service resilience and quality

- Definition: the model is designed around an existing service's needs to fill resource or geographic gaps; remove administrative or clinical burden; introduce cost efficiencies; or increase flexibility to innovate, change, adapt, or survive.
- Key aspects of care are determined at the organizational level with defined goals, criteria, boundaries, and processes, and there may be limited opportunities for service user involvement in its design and delivery.
- TM 4.1: internally or externally sourced. TMOUD pathways and expertise are developed internally or brought in from an external vendor.
- TM 4.2: predefined intake criteria. The TMOUD service has specific intake criteria. Where a TMOUD vendor is used, intake may rely on shared care protocols with the host organization to ensure that an appropriate screening process has taken place before referral.
- TM 4.3: coproduced clinical pathways. Where a vendor is used, service delivery pathways will need to be coproduced with the host organization to fit with the organization's strategic and pragmatic aims in introducing TMOUD.

Note: Models from Teck et al [57].

Agreed Upon Modes of TMOUD Delivery: Overview

A total of 4 modes of TMOUD delivery have been identified in the literature. One of the 4 modes exclusively involved web-based contact, whereas the other 3 were hybrid forms of TMOUD in which some in-person contact was involved. The hub-home model is frequently described in the literature and involves the clinician performing the TMOUD visit with the person in their home [36]. An initial screening assessment and triage may have taken place via telephone before prescriber

contact [65], and specific arrangements may have had to be put in place for identity confirmation and urine drug testing before the consultation [36,65,90]. The triadic hub-home model is a derivative of the more common hub-home model, in which a health care assistant or bridge worker attends the person's home to facilitate the TMOUD consultation. This is particularly helpful when the person may be disabled, struggles to use technology, or has additional needs such as the administration of a depot medication. Examples were not readily found in the literature; however, some key informants had experience using this model.

The hub-spoke model involves the person accessing technology to enable TMOUD in a setting outside their home. The triadic version of this model involved the presence of a health care assistant or bridge worker to facilitate the TMOUD consultation. The flexibility offered by the hub-spoke model meant that it was offered in a broad range of settings, including prisons, needle and syringe provision services, and rural primary care [62,63,91].

One of our stakeholder organizations—an advocacy group for families affected by drugs and alcohol—highlighted the lack of examples of family-inclusive practices in the literature on TMOUD modes of delivery. For example, within the context of the hub-spoke model, family members who often perform informal caring roles may be well suited to being TMOUD facilitators, ensuring that the person can go on the web, providing corroborative history, advocating for their family member's needs, or collecting medications from a dispensing pharmacy.

Agreed Upon Modes of TMOUD Delivery: Agreement on Who Should Carry Out the TMOUD Consultation

The choice of prescriber to perform TMOUD consultations depends on a variety of factors. By choice of prescriber, we mean the profession to which they belong (eg, pharmacy, nursing, or medicine); their level of seniority, experience, and training; and the degree of autonomy permitted to them by their respective organizations and governing bodies. For example, the General Medical Council in the United Kingdom highlighted specific areas of concern and risk when delivering telemedicine [92]. Issues must have been identified and addressed before there can be an expectation of a UK-registered physician to start providing TMOUD consultations, and different professions may have different requirements.

An example of a medical student conducting the initial assessment of a person accessing TMOUD, followed by a discussion and consultation with a fully qualified physician, was identified [93], highlighting the value for the next generation of clinicians preparing for what the future of addiction care may look like. Indeed, the incorporation of digital health competencies into the medical school curriculum is becoming increasingly important worldwide [94]. Another consideration is that clinical seniority and experience in providing MOUD may not translate easily to high-quality care via TMOUD. Indeed, there is evidence that some clinicians experienced difficulties in the rapid transition from in-person consultations to telemedicine at the pandemic onset [36]. Understanding TMOUD as a digital health intervention with specific skills and competencies is currently an underresearched area.

Clinic Coordination and Standard Operating Procedures

Key issues in this area included the ways in which people contact the service to access TMOUD and how the service would accommodate people with transient lifestyles, limited access to technology, and difficulties keeping their handsets safe or charged. Examples of street outreach services actively engaging homeless populations [35] and needle and syringe provision services acting as locations for TMOUD [91] may help overcome some of these issues. Key informants who provided frontline support services highlighted the increased demand for alternate nonverbal means of communication, such as web chat, email, or SMS text messaging, particularly in cases in which people may be tentatively exploring the idea of engaging with services. The need to involve support workers, families, or telemedicine facilitators outside the main provider's governance structure may require additional guidance or memoranda of understanding, allowing staff to work in more flexible yet defensible ways when making appointments or coordinating other aspects of care. Furthermore, standard lone workers or physical consultation risk assessments and safety policies will need to be modified for web-based environments. Multimedia Appendix 8 details other important factors to consider before TMOUD consultations may begin. We summarized in visual form the steps involved in an actual TMOUD consultation, adapting the work by the Welsh National Video Consultation Service Toolkit [95], in Multimedia Appendix 9.

Quality Standards and Monitoring

Quality standards for TMOUD provision will relate closely to the type of care that should be offered in response to the gaps and needs of the current system and, where relevant, statutory benchmarking, for example, the MAT standards outlined in Textbox 1. In Table 1, we outline specific service-level measures that relate to the Scottish MAT standards and show how they map to models of TMOUD derived from the international evidence base. Using the measures in Table 1 will not only support ongoing evaluation and service improvement but also enable benchmarking against international examples of TMOUD models of care.

Table 1. Medication-assisted treatment (MAT) standards 1 to 5 mapped against telemedicine-delivered medication for opioid use disorder (TMOUD) models of care.

Scottish MAT standards and what to measure	Mapping against TMOUD models of care	Implications for TMOUD quality
Standard 1: same-day access		
Time from first contact with any partner in the multiagency partnership within an episode of care to the commencement of MOUD[a]	• TM[b] 1.0: inclusion health focus • TM 1.1: places of safety • TM 1.3: low threshold • TM 3.1: multidisciplinary and multisectoral care	Same-day access is made possible by low thresholds for treatment initiation (TM 1.3). Multidisciplinary and multisectoral care (TM 3.1) increases opportunity for people to be seen where they feel safe (TM 1.1).
Number of people started on MOUD	• TM 1.3: low threshold • TM 4.0: service resilience and quality	Higher numbers in treatment are a measure of treatment threshold (TM 1.3) and may be an outcome measure required by a service adopting TMOUD (TM 4.0).
Standard 2: choice		
The number of people in treatment taking methadone or sublingual buprenorphine at the appropriate dose and with opportunities to change their choice	• TM 1.2: trauma informed • TM 3.3: education and empowerment • TM 4.0: service resilience and quality • TM 4.3: coproduced clinical pathways	Choice in treatment is a principle of trauma-informed care (TM 1.2), and people with opioid use disorder may need to be empowered to demand these choices (TM 3.3). Services need to have time and resources (TM 4.0) and appropriate clinical pathways (TM 4.3) to respond to demand.
The number of people in treatment receiving injectable buprenorphine	• TM 1.0: inclusion health focus • TM 3.3: education and empowerment • TM 4.0: service resilience and quality	Some groups may be excluded from being offered injectable buprenorphine (TM 1.0) or may not have knowledge of it (TM 4.0). Services must adapt to introduce injectable buprenorphine, and adopting TMOUD may increase the capacity to do so (TM 4.0).
The number of people receiving HAT[c]	• TM 1.0: inclusion health focus • TM 3.0: complexity of care needs • TM 4.0: service resilience and quality	HAT is a relatively novel intervention, potentially benefiting marginalized populations (TM 1.0) with complex care needs (TM 3.0). It is a resource-heavy intervention, and TMOUD may be used to free up capacity to deliver it (TM 4.0).
Standard 3: assertive outreach and anticipatory care		
Duration from when first identified as at risk to initial contact and assessment; number of people followed up with and for whom an initial assessment is performed	• TM 2.0: transition through care • TM 2.1: patient journey mapping • TM 2.2: process mapping and risk mitigation	This measure may provide empirical data to test service understandings of the patient journey (TM 2.1) and processes that may increase or decrease risk (TM 2.2) during transitions in care (TM 2.0).
Proportion of people by age, gender, and race identified as at risk by source of risk event (service that identified and actioned the risk); identification and outcome of the intervention	• TM 1.0: inclusion health focus • TM 1.1: places of safety • TM 2.0: transition through care • TM 2.2: process mapping and risk mitigation	Disaggregation of data by age, gender, race, and other characteristics measures the extent to which inclusivity is improved (TM 1.0), what services are favored by specific groups (TM 1.1), and whether risk mitigation (TM 2.2) during care transitions (TM 2.0) reaches all affected groups equally.
Standard 4: harm reduction		
Proportion of MOUD services offering BBV[d] testing and vaccination, naloxone and overdose awareness, wound care, assessment of injecting risk and injecting equipment provision, and virtual supervised injecting services	• TM 3.0: complexity of care needs • TM 3.2: flexibility to tailor interventions • TM 4.0: service resilience and quality • TM 4.3: coproduced clinical pathways	This measure quantifies the extent to which TMOUD services the complexity of the needs of people with opioid use disorder (TM 3.0) and the extent to which it can innovate or adapt to address gaps (TM 3.2). Furthermore, services may introduce telemedicine (TM 4.0) and coproduce pathways to integrate harm reduction and MOUD provision.
Standard 5: retention		

Scottish MAT standards and what to measure	Mapping against TMOUD models of care	Implications for TMOUD quality
Attrition rate: number of people currently on MOUD treatment and number of people discharged within a set time by setting, age, gender, and race	• TM 2.0: transition through care • TM 2.1: patient journey mapping • TM 2.2: process mapping and risk mitigation • TM 4.0: service resilience and quality • TM 4.3: coproduced clinical pathways	This directly measures the service's ability to support continuity through transitions in care (TM 2.0) and provides empirical data to test mapping and risk mitigation (TM 2.1 and 2.2). TMOUD may be adopted to improve retention in treatment (TM 4.0) through specific clinical pathways (TM 4.3).
Reason for discharge (eg, planned or unplanned)	• TM 3.0: complexity of care needs • TM 3.2: flexibility to tailor interventions • TM 4.0: service resilience and quality • TM 4.2: predefined intake criteria	Services designed to meet complex needs (TM 3.0) develop interventions to match these needs (TM 3.2) and, theoretically, will have low unplanned discharges. Services introducing TMOUD to increase efficiency (TM 4.0) may set high-threshold intake criteria (TM 4.2), which excludes some groups.

[a]MOUD: medication for opioid use disorder.

[b]TM: TMOUD model.

[c]HAT: heroin-assisted treatment.

[d]BBV: blood-borne virus.

Key Process Domains

Understanding how MOUD is defined and offered within a service is essential to understanding not only where the introduction of telemedicine would be most beneficial but also what the potential risks and CRM strategies should be. Several national and international guidance documents have set out how the core aspects and quality standards of MOUD services should be provided and demonstrated, including those from the United Kingdom [67], Scotland [4], the United States [96], and the World Health Organization [97]. With reference to these documents, we divided the processes involved in MOUD into those performed before the commencement of medication (accessing the service, enrollment, diagnosis of OUD, and treatment planning) and processes associated with medication commencement (induction, medication titration, substance use stabilization, and regular medication and clinical reviews). Textboxes 3 and 4 describe these processes and the associated risks involved in designing a TMOUD service. When introducing a pilot TMOUD project, clinicians and planners may decide to offer telemedicine for specific domains while also ensuring that arrangements are in place for other nontelemedicine domains. It is also critical for providers to evaluate the extent to which they can accommodate people who do not wish to have telemedicine consultations, particularly outside of pandemic or public health–mandated situations. These processes are often illustrated in the form of a process map [98], and an example of this is provided in Multimedia Appendix 10.

Textbox 3. Processes before the commencement of medication for opioid use disorder (MOUD) and considerations when transitioning to telemedicine.

Entry points

- Discovery: How will people find out about the telemedicine-delivered MOUD (TMOUD) service? Examples: word of mouth, internet, social media, referrals from other agencies, advertisements or flyers, or outreach or in-reach services.
- Accessibility: What will the service opening times be? By being on the web or through the telephone, are drop-in options no longer possible? What about out-of-hours access? Will there be a free phone number? Will this free phone number be integrated alongside other service numbers, for example, for unscheduled or emergency care? Will the service be by referral only?
- Safety and inclusivity: How are the principles of a psychologically informed, welcoming, trustworthy, private, transparent, and respectful service maintained through telemedicine? How are peer support; family involvement; collaborative working; and sex-, language-, and culturally appropriate services provided through telemedicine?
- Associated risks in planning for access:
  - Easily ignored or marginalized groups may not access the relevant information on TMOUD services in a timely way or have the skills to act on them.
  - Fear of stigmatizing or disrespectful attitudes may discourage people from attending a TMOUD consult.
  - Rigid appointment times may not suit this group, making the availability of a drop-in option important.
  - People with no fixed address will not receive postal appointments and may struggle to retain the same contact number throughout their treatment.
  - Although telemedicine has been described as having the potential to improve access to MOUD, the reality needs to be closely monitored as delays are associated with increased mortality and morbidity among people with opioid use disorder (OUD).

Enrollment

- Identity: confirming the identity of people experiencing homelessness and other itinerant or marginalized groups can be problematic yet essential when providing controlled drugs as medication. In addition, registration with a primary care service and the ability to check that the person is not already accessing these medications through another service are critical and may be more difficult via a fully web-based service. In some settings, eligibility to access the service and consent forms may need to be verified in person.
- Past medical history: timely access to medical records; the person's MOUD treatment history; recent investigations including liver function tests, drug tests, and electrocardiograms; information on comorbidities; recent hospital admissions; other prescribed medication; allergies; and contraindications are essential for safe MOUD induction.
- Additional needs such as blood-borne virus (BBV) testing; sterile injecting equipment access; take-home naloxone supplies; primary care; mental health, financial, housing, and nutritional support; advocacy; and a dispensing pharmacy are necessary components of care to be arranged.
- Risks:
  - Delays in the enrollment processes may delay safe prescribing, increasing the risk of withdrawals and contributing to increased mortality and morbidity risks.
  - Opioid withdrawals are independently associated with receptive syringe sharing and nonfatal overdose.
  - Failure to conduct the necessary checks may result in double prescribing or harms such as overdose from unsafe prescribing. The safety standards of the service may be brought into question, resulting in sanctions.
  - MOUD is most effective in preventing HIV and hepatitis C transmission if provided in the context of a robust sterile injecting equipment supply.

Diagnosis

- Assessment: information that contributes to a robust assessment includes the substances used, preferred route of administration (eg, sniffing, smoking, or injection), daily use (amount and cost), desired effect (eg, up, down, or feeling normal), tolerance, withdrawal symptoms, age at first use, strengths or skills, social support, criminal justice involvement, drug mixing, infection risk, nonfatal overdose frequency, psychiatric symptoms, previous treatment experiences, and whether the person meets validated criteria for OUD diagnosis.
- Examination: includes assessment of pulse, evidence of tremor or agitation, flushing, sweating, pupil size, yawning, irritability, mental state, injecting sites, and jaundice.
- Investigations or standardized tools: drug testing, pregnancy testing, liver function tests, electrocardiograms, and BBV near-person testing form part of standard good practice in-person MOUD provision. Many services are also required to collect a minimum data set to meet statutory or insurance provider obligations.
- Risks:
  - A diagnosis of OUD is an absolute prerequisite to initiate MOUD such as methadone and buprenorphine.
  - An overestimation of a person's level of tolerance to opioids may result in an overdose.

- Where the necessary information is collated, it is possible to make a diagnosis of dependence and make a prescribing decision at the first appointment.
- Individuals already well known to the service can often be safely reassessed and restarted on treatment rapidly, particularly if their information is relatively recent.
- Arbitrary attendance to multiple appointments before initiating treatment increases risks of harm to the person.

Treatment planning

- Reviewing potential harms and risks both of entering treatment and of the person's current situation: this includes the risk of overdose and death, infection, social isolation, acquisitive crime and incarceration, drug debts, violence, exploitation, homelessness, family breakdown, and increased suicidality.
- Exploring the person's priorities and perspectives: there is an ethical and legal obligation upon care providers to ensure that the person's values and priorities are understood and considered in clinical decision-making. For example, the person may wish to find a way to continue to use drugs safely, reduce the harms of their ongoing drug use, reduce their use to controllable levels, or stop entirely. They may wish to work toward abstinence and move on to rehabilitation services or seek sufficient stability to pursue other forms of personal development.
- Supporting an informed treatment choice: the mechanism of action of each MOUD option (typically methadone or buprenorphine) needs to be understood by the person, including the risks and benefits. There may be pharmacological prerequisites to be understood, such as needing a period without using before initiating buprenorphine. The clinician is typically able to make recommendations only after a discussion on the potential risk-benefit and the person's preference.
- Risks:
  - Credibility, trust, and rapport are required for treatment recommendations to be accepted.
  - A mismatch between the clinician's and the individual's outcome goals can result in disengagement from services.
  - Telemedicine may hamper the degree of communication required for a person to fully understand the information provided. Both clinicians and individuals may feel pressured into concluding the consultation quickly owing to discomfort with the web-based platform.

Textbox 4. Processes during and following the commencement of medication for opioid use disorder (MOUD) and considerations when transitioning to telemedicine.

Induction

- Safety: the identity of the person taking the medication needs to be confirmed, and a clinician able to determine their state at the time of induction (intoxication, withdrawals, or neutral) is required. A breath alcohol level may be necessary. Immediate action or follow-up arrangements may be needed if a risk of an overdose or harm is noted.
- Medication-specific issues: buprenorphine has a lower risk of respiratory depression and overdose in someone who has an opioid tolerance, and unobserved home induction may be possible. Conventionally, this requires definite withdrawal symptoms to avoid the possibility of induced withdrawals. As a full opioid agonist, methadone carries additional risk, and there are greater risks of harm with unobserved dosing at this stage of treatment.
- The first dose of treatment: Where will dispensing occur? Who will supervise the dose? Is the person required to be seen again after receiving their first dose? What is the feasibility of offering injectable formulations of MOUD via telemedicine?
- Risks:
  - Both the clinician and the dispensing pharmacy will need to confirm the person's identity.
  - With polysubstance use, for example, alcohol or benzodiazepine codependence, the status of the person at the point of induction may be difficult to determine via voice or video-based telemedicine.
  - Clear plans need to be in place for when a person attending a telemedicine-delivered MOUD (TMOUD) consultation is clearly unwell or intoxicated. These plans need to be feasible and timely, for example, ensuring emergency services or that take-home naloxone administration can occur to prevent an overdose.

Titration

- Attendance: titrating the dose of MOUD to an adequate level may require several encounters with a prescriber (daily, twice a week, or weekly). This is a critical phase as the person may feel the need to continue to use until the treatment takes the withdrawals away. Nonattendance to a review appointment is difficult to manage in both a telemedicine or in-person context. A plan needs to be in place to manage nonattendance, including how safeguarding or check-ins will be performed. Other considerations include a management plan for losses to follow-up and strategies to increase retention in treatment.
- Support: outreach support, TMOUD facilitators, appointment reminders, tracing service if appointments are missed, peer or psychosocial support, and ongoing harm reduction interventions such as take-home naloxone provision can significantly improve engagement and safety of the titration phase and compensate for the lack of in-person clinical contact.
- Dosing-related issues: frequency of dose increases and consideration of how these are managed with the dispensing pharmacy are critical. Will the person need to come to the service to pick up a prescription? Will the pharmacy accept electronic prescriptions? How will missed doses or intoxication observed by the pharmacist be managed?
- Risks:
  - The risk of death increases at the beginning and end of MOUD treatment.
  - Methadone initiation requires prescribing at subtherapeutic doses and subsequent titration to a therapeutic dose over a few weeks, resulting in risks with ongoing drug use and death, making the first 4 weeks of treatment risky.
  - Supervised consumption is used to reduce the risk of overdose and diversion and is associated with a reduction in illicit heroin and alcohol use but may be associated with decreased retention in treatment.
  - Mortality risk at treatment onset is lower among those initiated on buprenorphine than methadone; however, retention may be better with methadone.
  - Currently, experience with TMOUD has been predominantly with buprenorphine rather than methadone.

Stabilization

- Comorbidities: a period of stabilization may provide opportunities for people engaging with MOUD to begin to manage coexisting conditions such as HIV, hepatitis C, and tuberculosis treatment.
- Managing other substances: polysubstance use with alcohol and illicit substances or prescription opioids limits the benefits gained from MOUD. The period of stability through MOUD commencement may support the person to seek help with these other substances.
- Psychosocial interventions: contingency management, cognitive behavioral therapy, relapse prevention, dialectical behavior therapy, group drug counseling, mutual aid including 12-step and Specific, Measurable, Attainable, Relevant, and Time-Bound goals, employment support, education and training, upskilling, benefit maximization, legal support, advocacy, compassion-focused therapy, and mindfulness have all been successfully delivered via a telemedicine platform and should be considered in the design of TMOUD services.
- Risks:
  - People who have stabilized in their substance use may become more aware of previously hidden psychiatric symptoms and will need timely access to mental health services. Anticipating and offering this within a telemedicine service is important though not always feasible.

- A service transitioning to telemedicine needs to provide avenues to access other services critical to the person's ongoing welfare to maximize the benefits of MOUD.

Medication reviews

- Adapting to changing circumstances: the TMOUD service will need to respond to the changing health circumstances of individuals, for example, the onset of cardiovascular or respiratory disease as the person ages. An individual's needs may change through their treatment journey, and flexibility in dispensing arrangements may be required, which should be incorporated into the TMOUD service policies. The TMOUD service will need to have provision for prescribing adjustments, for example, where individuals have palliative or chronic pain needs.
- Managing relapses: a clear plan for how relapses are managed must be in place, including dosing adjustments and harm reduction interventions. Is the person at greater risk of relapsing without the appropriate support or would it be undetected if engaging with a TMOUD rather than an in-person service?
- Moving forward: Does the service provide or link with other services? Are there opportunities to access long-acting MOUD formulations, which reduce the person's need to attend a dispensing pharmacy regularly?
- Risks:
  - A telemedicine service may not provide the usual touch points available in an in-person service, which may trigger professional curiosity or opportunities for exploring topics unrelated to MOUD provision alone.
  - Consequently, the TMOUD service may become less relevant to the individuals' circumstances; however, the need for arrangements for change, transfer, or discharge may not be anticipated.

Iterative Cycles of Evaluation and Improvement

A Quality Improvement Focus

Existing guidance on telemedicine service transition during and beyond the COVID-19 pandemic has recommended a quality improvement approach to TMOUD implementation [62,69] echoed by several key informants. This may include the development of a theory of change, a phased implementation approach, and iterative testing to develop and refine the processes informed by feedback from stakeholders [99]. Another approach identified in the literature is the use of formal implementation frameworks to evaluate TMOUD projects. For example, the Consolidated Framework for Implementation Research was used in one study as a conceptual guide for the systematic assessment of barriers and facilitators in different multilevel implementation contexts to influence TMOUD effectiveness [62]. Using the Consolidated Framework for Implementation Research, Brunet et al [62] extended the transferability of their evaluation findings to other services in diverse settings.

Evaluation

In addition to service-level quality measures identified in Table 1, validated tools to evaluate TMOUD more generally may be useful. Several evaluation tools were identified in the literature, including the Telemedicine Satisfaction Questionnaire [100] evaluating the usability of the telemedicine service [101] and the Telemedicine Service Maturity Model [102,103] to evaluate the implementation of telemedicine services. One key study reviewed several existing tools to develop a survey-based measurement tool for TMOUD satisfaction among people who use drugs receiving this treatment in rural areas [104].

Discussion

Principal Findings

We have reported on a multistakeholder, process-focused implementation model for TMOUD. This model is intended to provide clear step-by-step guidance on TMOUD implementation informed by evidence from the peer-reviewed literature as well as internationally derived case study expertise. Although this work is admittedly focused on the Scottish context, the use of the SHIFT-Evidence framework to underpin model development, along with the international evidence base, may increase its relevance to a wider, international audience. Through stakeholder engagement and consultation, we tried to mitigate the perception of a top-down imposition on how TMOUD should be delivered or implemented. Our consultation efforts resulted in the capture and inclusion of a wide range of knowledge, perspectives, experiences, and opinions. A limitation is the lack of a specific mechanism to form a consensus on what is critical to the model, for example, through a survey-based Delphi process. We have secured funding to field-test this TMOUD implementation model in Scotland, after which we will seek formal consensus on its validity in other settings.

Another limitation of our study was a reliance on a US-dominated evidence base, which >70% of TMOUD research currently comes from [48]. Experience in the United States tends to involve the provision of buprenorphine via telemedicine, whereas services in the United Kingdom tend to prescribe methadone more often than buprenorphine. This may mean that UK services need to develop a robust homegrown evidence base to provide methadone via telemedicine to fully benefit from TMOUD implementation. Furthermore, the evidence base grew substantially from the time this work began to the point of submission for peer review. Built into the design of this project were iterative and flexible scoping reviews and ongoing consultation with the community of practice derived from the case study identification stage.

There is currently a gap in evidence-based risk mitigation strategies when delivering TMOUD, identified in the process-mapping activity described previously. Furthermore, there is a need to develop technological approaches to improve the role of telemedicine platforms to act as a diagnostic platform as well as a medium for consultations when delivering TMOUD. Proposals for this include adapting smartphone-based pupillometry techniques to assess the degree of withdrawal from opioids [105] and a process for web-based point-of-care drug testing [106]. Although it is recognized that frontline practitioners and people who use drugs should be partners in the design, implementation, and evaluation of TMOUD [48], there is little evidence on how to engage these groups in a meaningful and impactful way.

Conclusions

In conclusion, we have produced a process-driven TMOUD implementation model intended to bridge gaps in knowledge at the local level while also providing flexibility to respond to local needs and contexts. A critical next step will be to field-test the model and seek international consensus on its validity across different settings.

Acknowledgments

This project was funded by the Scottish government's Drug Deaths Task Force and the Corra Foundation. The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript. The views expressed are those of the authors and not those of the Scottish government's Drug Deaths Task Force or the Corra Foundation. The authors would also like to thank all the key informants and Heidi Dritschel for her work in administering the project.

Authors' Contributions

JTWT contributed to conceptualization, methodology, funding acquisition, formal analysis, and writing. RG contributed to writing, review, and editing. GZ contributed to normal analysis and project administration. AO contributed to formal analysis, project administration, and review. SG-S contributed to reviewing and editing. AB contributed to conceptualization, funding acquisition, and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Applying the 12 Successful Healthcare Improvements From Translation of Evidence Into Practice rules to the study design.
[DOCX File , 17 KB - mental_v10i1e47186_app1.docx]

Multimedia Appendix 2
Logic model.
[DOCX File , 362 KB - mental_v10i1e47186_app2.docx]

Multimedia Appendix 3
Web-based stakeholder engagement event topic list.
[DOCX File , 20 KB - mental_v10i1e47186_app3.docx]

Multimedia Appendix 4
Draft implementation model (PDF form).
[PDF File (Adobe PDF File), 4145 KB - mental_v10i1e47186_app4.pdf]

Multimedia Appendix 5
Screenshots of the web-based platform used to consult on the implementation model.
[PDF File (Adobe PDF File), 583 KB - mental_v10i1e47186_app5.pdf]

Multimedia Appendix 6
Diversity of stakeholders and key informants in the consultation phase.
[DOCX File , 15 KB - mental_v10i1e47186_app6.docx]

Multimedia Appendix 7
Examples of inclusion and exclusion criteria for telemedicine-delivered medication for opioid use disorder.
[DOCX File , 15 KB - mental_v10i1e47186_app7.docx]

Multimedia Appendix 8
Clinic coordination and administration.
[DOCX File , 16 KB - mental_v10i1e47186_app8.docx]

Multimedia Appendix 9
Visual guide to conducting a telemedicine-delivered medication for opioid use disorder consultation.
[DOCX File , 1490 KB - mental_v10i1e47186_app9.docx]

Multimedia Appendix 10
Telemedicine-delivered medication for opioid use disorder process map.
[PDF File (Adobe PDF File), 30 KB - mental_v10i1e47186_app10.pdf]

References

1. McAuley A, Fraser R, Glancy M, Yeung A, Jones HE, Vickerman P, et al. Mortality among individuals prescribed opioid-agonist therapy in Scotland, UK, 2011-20: a national retrospective cohort study. Lancet Public Health 2023 Jul;8(7):e484-e493 [FREE Full text] [doi: 10.1016/S2468-2667(23)00082-8] [Medline: 37295452]
2. Baumgartner R. Precision medicine and digital phenotyping: digital medicine's way from more data to better health. Big Data Soc 2021 Jul;8(2):205395172110664 [FREE Full text] [doi: 10.1177/20539517211066452]
3. Changing lives: our final report. Scottish Drug Deaths Taskforce. 2022. URL: https://drugstaskforce.knowthescore.info/wp-content/uploads/sites/2/2022/08/Changing-Lives-updated-1.pdf [accessed 2023-06-19]
4. Medication Assisted Treatment (MAT) standards for Scotland access, choice, support. Scottish Government. 2021. URL: https://tinyurl.com/cfputvef [accessed 2021-09-12]
5. Santo TJ, Clark B, Hickman M, Grebely J, Campbell G, Sordo L, et al. Association of opioid agonist treatment with all-cause mortality and specific causes of death among people with opioid dependence: a systematic review and meta-analysis. JAMA Psychiatry 2021 Sep 01;78(9):979-993 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.0976] [Medline: 34076676]
6. Volkow ND, Frieden TR, Hyde PS, Cha SS. Medication-assisted therapies--tackling the opioid-overdose epidemic. N Engl J Med 2014 May 29;370(22):2063-2066. [doi: 10.1056/NEJMp1402780] [Medline: 24758595]
7. Winograd RP, Presnall N, Stringfellow E, Wood C, Horn P, Duello A, et al. The case for a medication first approach to the treatment of opioid use disorder. Am J Drug Alcohol Abuse 2019 May 14;45(4):333-340. [doi: 10.1080/00952990.2019.1605372] [Medline: 31084515]
8. Marchand K, Beaumont S, Westfall J, MacDonald S, Harrison S, Marsh DC, et al. Conceptualizing patient-centered care for substance use disorder treatment: findings from a systematic scoping review. Subst Abuse Treat Prev Policy 2019 Sep 11;14(1):37 [FREE Full text] [doi: 10.1186/s13011-019-0227-0] [Medline: 31511016]
9. Donroe JH, Tetrault JM. Narrowing the treatment gap in managing opioid use disorder. CMAJ 2018 Mar 05;190(9):E236-E237 [FREE Full text] [doi: 10.1503/cmaj.180209] [Medline: 29507154]
10. Barry CL, McGinty EE, Pescosolido BA, Goldman HH. Stigma, discrimination, treatment effectiveness, and policy: public views about drug addiction and mental illness. Psychiatr Serv 2014 Oct;65(10):1269-1272 [FREE Full text] [doi: 10.1176/appi.ps.201400140] [Medline: 25270497]
11. McGinty EE, Stone EM, Kennedy-Hendricks A, Bachhuber MA, Barry CL. Medication for opioid use disorder: a national survey of primary care physicians. Ann Intern Med 2020 Jul 21;173(2):160-162 [FREE Full text] [doi: 10.7326/M19-3975] [Medline: 32311740]
12. Eisenberg MD, McCourt A, Stuart EA, Rutkow L, Tormohlen KN, Fingerhood MI, et al. Studying how state health services delivery policies can mitigate the effects of disasters on drug addiction treatment and overdose: protocol for a mixed-methods study. PLoS One 2021 Dec 16;16(12):e0261115 [FREE Full text] [doi: 10.1371/journal.pone.0261115] [Medline: 34914779]
13. Miclette MA, Leff JA, Cuan I, Samet JH, Saloner B, Mendell G, et al. Closing the gaps in opioid use disorder research, policy and practice: conference proceedings. Addict Sci Clin Pract 2018 Nov 13;13(1):22 [FREE Full text] [doi: 10.1186/s13722-018-0123-3] [Medline: 30424803]
14. Mojtabai R, Mauro C, Wall MM, Barry CL, Olfson M. Medication treatment for opioid use disorders in substance use treatment facilities. Health Aff (Millwood) 2019 Jan;38(1):14-23 [FREE Full text] [doi: 10.1377/hlthaff.2018.05162] [Medline: 30615514]
15. Stringer KL, Langdon KJ, McKenzie M, Brockmann B, Marotta P. Leveraging COVID-19 to sustain regulatory flexibility in the treatment of opioid use disorder. J Subst Abuse Treat 2021 Apr;123:108263 [FREE Full text] [doi: 10.1016/j.jsat.2020.108263] [Medline: 33612196]
16. Feder KA, Krawczyk N, Saloner B. Medication-assisted treatment for adolescents in specialty treatment for opioid use disorder. J Adolesc Health 2017 Jun;60(6):747-750 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.12.023] [Medline: 28258807]
17. Frazer Z, McConnell K, Jansson LM. Treatment for substance use disorders in pregnant women: motivators and barriers. Drug Alcohol Depend 2019 Dec 01;205:107652. [doi: 10.1016/j.drugalcdep.2019.107652] [Medline: 31704383]

18. Krans EE, Kim JY, James AE3, Kelley D, Jarlenski MP. Medication-assisted treatment use among pregnant women with opioid use disorder. Obstet Gynecol 2019 May;133(5):943-951 [FREE Full text] [doi: 10.1097/AOG.0000000000003231] [Medline: 30969219]
19. Salvador J, Bhatt S, Fowler R, Ritz J, James R, Jacobsohn V, et al. Engagement with project ECHO to increase medication-assisted treatment in rural primary care. Psychiatr Serv 2019 Dec 01;70(12):1157-1160 [FREE Full text] [doi: 10.1176/appi.ps.201900142] [Medline: 31434561]
20. Tyndall M, Dodd Z. How structural violence, prohibition, and stigma have paralyzed north American responses to opioid overdose. AMA J Ethics 2020 Aug 01;22(1):E723-E728 [FREE Full text] [doi: 10.1001/amajethics.2020.723] [Medline: 32880362]
21. Allden K, Murakami N, Maung C. Trauma and Recovery on War's Border: A Guide for Global Health Workers. Darby, MT: Dartmouth College Press; 2015.
22. Kane JC, Ventevogel P, Spiegel P, Bass JK, van Ommeren M, Tol WA. Mental, neurological, and substance use problems among refugees in primary health care: analysis of the Health Information System in 90 refugee camps. BMC Med 2014 Nov 24;12(1):228 [FREE Full text] [doi: 10.1186/s12916-014-0228-9] [Medline: 25420518]
23. Miler JA, Carver H, Masterton W, Parkes T, Maden M, Jones L, et al. What treatment and services are effective for people who are homeless and use drugs? A systematic 'review of reviews'. PLoS One 2021 Jul 14;16(7):e0254729 [FREE Full text] [doi: 10.1371/journal.pone.0254729] [Medline: 34260656]
24. Dickie E, Clusker T, McCormick D. Medication assisted treatment MAT standards for Scotland access, choice, support interim report. Scottish Drug Deaths Taskforce. 2021. URL: https://www.drugsandalcohol.ie/33995/1/MAT-Subgroup-Interim-Report-on-programme-to-date_Mar21.pdf [accessed 2023-06-28]
25. The Scottish Drug Deaths Task Force evidence paper. The Scottish Drug Deaths Task Force. URL: https://drugstaskforce.knowthescore.info/wp-content/uploads/sites/2/2022/08/scottish-drug-deaths-taskforce-evidence-paper-final-version.pdf [accessed 2023-06-19]
26. Eibl JK, Morin K, Leinonen E, Marsh DC. The state of opioid agonist therapy in Canada 20 years after federal oversight. Can J Psychiatry 2017 Jul 19;62(7):444-450 [FREE Full text] [doi: 10.1177/0706743717711167] [Medline: 28525291]
27. Eibl JK, Gauthier G, Pellegrini D, Daiter J, Varenbut M, Hogenbirk JC, et al. The effectiveness of telemedicine-delivered opioid agonist therapy in a supervised clinical setting. Drug Alcohol Depend 2017 Jul 01;176:133-138 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.01.048] [Medline: 28535455]
28. Huskamp HA, Busch AB, Souza J, Uscher-Pines L, Rose S, Wilcock A, et al. How is telemedicine being used in opioid and other substance use disorder treatment? Health Aff (Millwood) 2018 Dec;37(12):1940-1947 [FREE Full text] [doi: 10.1377/hlthaff.2018.05134] [Medline: 30633671]
29. Molfenter T, Boyle M, Holloway D, Zwick J. Trends in telemedicine use in addiction treatment. Addict Sci Clin Pract 2015 May 28;10:14 [FREE Full text] [doi: 10.1186/s13722-015-0035-4] [Medline: 26016484]
30. Wang L, Weiss J, Ryan EB, Waldman J, Rubin S, Griffin JL. Telemedicine increases access to buprenorphine initiation during the COVID-19 pandemic. J Subst Abuse Treat 2021 May;124:108272 [FREE Full text] [doi: 10.1016/j.jsat.2020.108272] [Medline: 33771276]
31. Weintraub E, Greenblatt AD, Chang J, Himelhoch S, Welsh C. Expanding access to buprenorphine treatment in rural areas with the use of telemedicine. Am J Addict 2018 Dec;27(8):612-617. [doi: 10.1111/ajad.12805] [Medline: 30265425]
32. Krawczyk N, Fawole A, Yang J, Tofighi B. Early innovations in opioid use disorder treatment and harm reduction during the COVID-19 pandemic: a scoping review. Addict Sci Clin Pract 2021 Nov 13;16(1):68 [FREE Full text] [doi: 10.1186/s13722-021-00275-1] [Medline: 34774106]
33. Cantor J, McBain RK, Kofner A, Hanson R, Stein BD, Yu H. Telehealth adoption by mental health and substance use disorder treatment facilities in the COVID-19 pandemic. Psychiatr Serv 2022 Apr 01;73(4):411-417. [doi: 10.1176/appi.ps.202100191] [Medline: 34407631]
34. Tofighi B, McNeely J, Walzer D, Fansiwala K, Demner A, Chaudhury CS, et al. A telemedicine buprenorphine clinic to serve New York City: initial evaluation of the NYC public hospital system's initiative to expand treatment access during the COVID-19 pandemic. J Addict Med 2022;16(1):e40-e43 [FREE Full text] [doi: 10.1097/ADM.0000000000000809] [Medline: 33560696]
35. Tringale R, Subica AM. COVID-19 innovations in medication for addiction treatment at a Skid Row syringe exchange. J Subst Abuse Treat 2021 Feb;121:108181 [FREE Full text] [doi: 10.1016/j.jsat.2020.108181] [Medline: 33129635]
36. Uscher-Pines L, Sousa J, Raja P, Mehrotra A, Barnett M, Huskamp HA. Treatment of opioid use disorder during COVID-19: experiences of clinicians transitioning to telemedicine. J Subst Abuse Treat 2020 Nov;118:108124 [FREE Full text] [doi: 10.1016/j.jsat.2020.108124] [Medline: 32893047]
37. Watson DP, Swartz JA, Robison-Taylor L, Mackesy-Amiti ME, Erwin K, Gastala N, et al. Syringe service program-based telemedicine linkage to opioid use disorder treatment: protocol for the STAMINA randomized control trial. BMC Public Health 2021 Mar 31;21(1):630 [FREE Full text] [doi: 10.1186/s12889-021-10669-0] [Medline: 33789642]
38. Ghosh A, Mahintamani T, Pillai RR, Subodh BN, Mattoo SK, Basu D. Telemedicine-assisted stepwise approach of service delivery for substance use disorders in India. Asian J Psychiatr 2021 Apr;58:102582 [FREE Full text] [doi: 10.1016/j.ajp.2021.102582] [Medline: 33607350]

39. Why trust digital health? Understanding the perspectives of communities affected by BBVs/STIs and social stigma. Research Gate. 2023 Feb 06. URL: https://www.researchgate.net/publication/348097955_Why_trust_digital_health_Understanding_the_perspectives_of_communities_affected_by_BBVsSTIs_and_social_stigma [accessed 2022-03-29]
40. Kho J, Gillespie N, Martin-Khan M. A systematic scoping review of change management practices used for telemedicine service implementations. BMC Health Serv Res 2020 Sep 01;20(1):815 [FREE Full text] [doi: 10.1186/s12913-020-05657-w] [Medline: 32873295]
41. Lin LA, Casteel D, Shigekawa E, Weyrich MS, Roby DH, McMenamin SB. Telemedicine-delivered treatment interventions for substance use disorders: a systematic review. J Subst Abuse Treat 2019 Jun;101:38-49. [doi: 10.1016/j.jsat.2019.03.007] [Medline: 31006553]
42. Nabitz U, van Den Brink W, Jansen P. Using concept mapping to design an indicator framework for addiction treatment centres. Int J Qual Health Care 2005 Jun;17(3):193-201. [doi: 10.1093/intqhc/mzi037] [Medline: 15831547]
43. Teck JT, Baldacchino AM. COVID-19 and substance use disorders: syndemic responses to a global pandemic. In: el-Guebaly N, Carrà G, Galanter M, Baldacchino AM, editors. Textbook of Addiction Treatment: International Perspectives. Cham, Switzerland: Springer; 2021:1269-1281.
44. Trujols J, Garijo I, Siñol N, del Pozo J, Portella MJ, Pérez de los Cobos J. Patient satisfaction with methadone maintenance treatment: the relevance of participation in treatment and social functioning. Drug Alcohol Depend 2012 Jun 01;123(1-3):41-47. [doi: 10.1016/j.drugalcdep.2011.10.014] [Medline: 22071121]
45. Walters SM, Perlman DC, Guarino H, Mateu-Gelabert P, Frank D. Lessons from the first wave of COVID-19 for improved medications for opioid use disorder (MOUD) treatment: benefits of easier access, extended take homes, and new delivery modalities. Subst Use Misuse 2022;57(7):1144-1153 [FREE Full text] [doi: 10.1080/10826084.2022.2064509] [Medline: 35443862]
46. National benchmarking report on implementation of the medication assisted treatment (MAT) standards. Public Health Scotland. 2022. URL: https://www.publichealthscotland.scot/publications/national-benchmarking-report-on-implementation-of-the-medication-assisted-treatment-mat-standards/national-benchmarking-report-on-implementation-of-the-medication-assisted-treatment-mat-standards-20222023/ [accessed 2023-07-13]
47. Update on medication assisted treatment standards. Health and social care news, Scottish Government. URL: http://www.gov.scot/news/update-on-medication-assisted-treatment-standards/ [accessed 2023-07-13]
48. Teck JT, Zlatkute G, Perez A, Dritschel H, Ghosh A, Potenza MN, et al. Key implementation factors in telemedicine-delivered medications for opioid use disorder: a scoping review informed by normalisation process theory. The Lancet Psychiatry 2023 Jan;10(1):50-64. [doi: 10.1016/s2215-0366(22)00374-1]
49. Scottish G. A national telehealth and telecare delivery plan for Scotland to 2015. The Scottish Government. 2012. URL: https://www.gov.scot/binaries/content/documents/govscot/publications/advice-and-guidance/2012/12/national-telehealth-telecare-delivery-plan-scotland-2016-driving-improvement-integration-innovation/documents/00411586-pdf/00411586-pdf/govscot%3Adocument/00411586.pdf [accessed 2023-07-13]
50. Wherton J, Greenhalgh T. Evaluation of the attend anywhere/near me video consulting service in Scotland, 2019-20. Scottish Government Edinburgh. URL: https://tinyurl.com/y7a76stt [accessed 2023-07-13]
51. Wherton J, Greenhalgh T, Shaw SE. Expanding video consultation services at pace and scale in Scotland during the COVID-19 pandemic: national mixed methods case study. J Med Internet Res 2021 Oct 07;23(10):e31374 [FREE Full text] [doi: 10.2196/31374] [Medline: 34516389]
52. Wherton J, Greenhalgh T. Evaluation of the near me video consulting service in Scotland during COVID-19. Scottish Government. 2020. URL: https://tec.scot/programme-areas/near-me/equality-and-accessibility/evaluation-near-me-consulting-service-scotland-during-covid-19 [accessed 2023-09-29]
53. Joudrey PJ, Adams ZM, Bach P, Van Buren S, Chaiton JA, Ehrenfeld L, et al. Methadone access for opioid use disorder during the COVID-19 pandemic within the United States and Canada. JAMA Netw Open 2021 Jul 01;4(7):e2118223 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.18223] [Medline: 34297070]
54. Reed JE, Howe C, Doyle C, Bell D. Simple rules for evidence translation in complex systems: a qualitative study. BMC Med 2018 Jun 20;16(1):92 [FREE Full text] [doi: 10.1186/s12916-018-1076-9] [Medline: 29921274]
55. McLaughlin JA, Jordan GB. Logic models: a tool for telling your programs performance story. Eval Program Plann 1999 Mar;22(1):65-72 [FREE Full text] [doi: 10.1016/S0149-7189(98)00042-1]
56. Mishuris RG, Palmisano J, McCullagh L, Hess R, Feldstein DA, Smith PD, et al. Using normalisation process theory to understand workflow implications of decision support implementation across diverse primary care settings. BMJ Health Care Inform 2019 Oct 19;26(1):e100088 [FREE Full text] [doi: 10.1136/bmjhci-2019-100088] [Medline: 31630113]
57. Tay Wee Teck J, Butner JL, Baldacchino A. Understanding the use of telemedicine across different opioid use disorder treatment models: a scoping review. J Telemed Telecare 2023 Sep 04:1357633X231195607 [FREE Full text] [doi: 10.1177/1357633X231195607] [Medline: 37661829]
58. Gottlieb M, Haas MR, Daniel M, Chan TM. The scoping review: a flexible, inclusive, and iterative approach to knowledge synthesis. AEM Educ Train 2021 Jul;5(3):e10609 [FREE Full text] [doi: 10.1002/aet2.10609] [Medline: 34222749]
59. Gordon I. Seminar delivering medication assited treatment for addictions via telehealth. School of Medicine, University of St Andrews. 2021 Apr 19. URL: https://medicine.st-andrews.ac.uk/digitas/2021/04/19/seminar-title/ [accessed 2023-10-03]

60. Coronavirus resilience planning: use of near me videoconsulting- organisation. Technology Enabled Care Scotland. 2020. URL: https://tec.scot/sites/default/files/2021-06/Near-Me-Covid19-Organisational-Guidance.pdf [accessed 2022-07-29]
61. Digital citizen delivery plan 2021/2022. Technology Enabled Care Scotland. 2021. URL: https://tec.scot/sites/default/files/2021-06/Digital-Citizen-Delivery-Plan-final-21-22.pdf [accessed 2022-07-29]
62. Brunet N, Moore DT, Lendvai Wischik D, Mattocks KM, Rosen MI. Increasing buprenorphine access for veterans with opioid use disorder in rural clinics using telemedicine. Subst Abus 2022 Feb 20;43(1):39-46. [doi: 10.1080/08897077.2020.1728466] [Medline: 32078492]
63. Duncan A, Sanders N, Schiff M, Winkelman TN. Adaptations to jail-based buprenorphine treatment during the COVID-19 pandemic. J Subst Abuse Treat 2021 Feb;121:108161 [FREE Full text] [doi: 10.1016/j.jsat.2020.108161] [Medline: 33371945]
64. Tofighi B, Isaacs N, Byrnes-Enoch H, Lakew R, Lee JD, Berry C, et al. Expanding treatment for opioid use disorder in publicly funded primary care clinics: exploratory evaluation of the NYC health + hospitals buprenorphine ECHO program. J Subst Abuse Treat 2019 Nov;106:1-3 [FREE Full text] [doi: 10.1016/j.jsat.2019.08.003] [Medline: 31540604]
65. Crowley D, Delargy I. A national model of remote care for assessing and providing opioid agonist treatment during the COVID-19 pandemic: a report. Harm Reduct J 2020 Jul 17;17(1):49 [FREE Full text] [doi: 10.1186/s12954-020-00394-z] [Medline: 32680520]
66. Fiacco L, Pearson BL, Jordan R. Telemedicine works for treating substance use disorder: the STAR clinic experience during COVID-19. J Subst Abuse Treat 2021 Jun;125:108312 [FREE Full text] [doi: 10.1016/j.jsat.2021.108312] [Medline: 34016299]
67. Clinical guidelines on drug misuse and dependence update. Independent Expert Working Group. 2017. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673978/clinical_guidelines_2017.pdf [accessed 2023-09-29]
68. Lingam HA, Caudill RL. Teleprescribing controlled substances: flowchart analysis of the Ryan Haight act and state telemedicine laws. J Technol Behav Sci 2019 May 29;4(4):346-359. [doi: 10.1007/s41347-019-00100-2]
69. Weintraub E, Seneviratne C, Anane J, Coble K, Magidson J, Kattakuzhy S, et al. Mobile telemedicine for buprenorphine treatment in rural populations with opioid use disorder. JAMA Netw Open 2021 Aug 02;4(8):e2118487 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.18487] [Medline: 34448869]
70. Lofwall MR, Walsh SL. A review of buprenorphine diversion and misuse: the current evidence base and experiences from around the world. J Addict Med 2014;8(5):315-326 [FREE Full text] [doi: 10.1097/ADM.0000000000000045] [Medline: 25221984]
71. Nosyk B, Slaunwhite A, Urbanoski K, Hongdilokkul N, Palis H, Lock K, et al. Evaluation of risk mitigation measures for people with substance use disorders to address the dual public health crises of COVID-19 and overdose in British Columbia: a mixed-method study protocol. BMJ Open 2021 Jun 09;11(6):e048353 [FREE Full text] [doi: 10.1136/bmjopen-2020-048353] [Medline: 34108170]
72. Day N, Wass M, Smith K. Virtual opioid agonist treatment: Alberta's virtual opioid dependency program and outcomes. Addict Sci Clin Pract 2022 Jul 28;17(1):40 [FREE Full text] [doi: 10.1186/s13722-022-00323-4] [Medline: 35902924]
73. Durand L, Keenan E, Boland F, Harnedy N, Delargy Í, Scully M, et al. Consensus recommendations for opioid agonist treatment following the introduction of emergency clinical guidelines in Ireland during the COVID-19 pandemic: a national Delphi study. Int J Drug Policy 2022 Aug;106:103768 [FREE Full text] [doi: 10.1016/j.drugpo.2022.103768] [Medline: 35738029]
74. Hew A, Arunogiri S, Lubman DI. Challenges in delivering telemedicine to vulnerable populations: experiences of an addiction medical service during COVID-19. Med J Aust 2021 Sep 06;215(5):237-2e1 [FREE Full text] [doi: 10.5694/mja2.51213] [Medline: 34365643]
75. FAQs: provision of methadone and buprenorphine for the treatment of opioid use disorder in the COVID-19 emergency. Mental Health Services Administration. 2020. URL: https://integrationacademy.ahrq.gov/resources/13966 [accessed 2023-09-29]
76. Hser YI, Ober AJ, Dopp AR, Lin C, Osterhage KP, Clingan SE, et al. Is telemedicine the answer to rural expansion of medication treatment for opioid use disorder? Early experiences in the feasibility study phase of a national drug abuse treatment clinical trials network trial. Addict Sci Clin Pract 2021 Apr 20;16(1):24 [FREE Full text] [doi: 10.1186/s13722-021-00233-x] [Medline: 33879260]
77. NHS in Scotland 2019. Audit Scotland. 2019. URL: https://www.audit-scotland.gov.uk/uploads/docs/report/2019/nr_191024_nhs_overview.pdf [accessed 2021-08-15]
78. Black C. Review of drugs part two: prevention, treatment, and recovery. Government of UK. 2021. URL: https://www.gov.uk/government/publications/review-of-drugs-phase-two-report/review-of-drugs-part-two-prevention-treatment-and-recovery [accessed 2022-08-06]
79. Briner M, Kessler O, Pfeiffer Y, Wehner T, Manser T. Assessing hospitals' clinical risk management: development of a monitoring instrument. BMC Health Serv Res 2010 Dec 13;10(1):337 [FREE Full text] [doi: 10.1186/1472-6963-10-337] [Medline: 21144039]

80. Oesterle TS, Kolla B, Risma CJ, Breitinger SA, Rakocevic DB, Loukianova LL, et al. Substance use disorders and telehealth in the COVID-19 pandemic era: a new outlook. Mayo Clin Proc 2020 Dec;95(12):2709-2718 [FREE Full text] [doi: 10.1016/j.mayocp.2020.10.011] [Medline: 33276843]
81. Chan B, Bougatsos C, Priest K, McCarty D, Grusing S, Chou R. Opioid treatment programs, telemedicine and COVID-19: a scoping review. Subst Abus 2022;43(1):539-546. [doi: 10.1080/08897077.2021.1967836] [Medline: 34520702]
82. Holmes H, Burgess G. Digital exclusion and poverty in the UK: how structural inequality shapes experiences of getting online. Digit Geogr Soc 2022;3:100041 [FREE Full text] [doi: 10.1016/j.diggeo.2022.100041]
83. Connecting Scotland: phase 1 evaluation. Scottish Government. 2022. URL: https://www.gov.scot/publications/connecting-scotland-phase-1-evaluation/#:~:text=Evaluation%20of%20the%20first%20phase,restrictions%20resulting%20from%20COVID%20%2D19 [accessed 2023-09-29]
84. Hadi D, Hannah C, Catriona M, Tessa P, Joe S, Graeme S. Digital inclusion to prevent drug related deaths: literature review. Drugs Research Network Scotland. 2022. URL: https://dspace.stir.ac.uk/retrieve/0b14ab27-b05c-4876-a068-1956bc749985/Literature%20review%20-%20interim%20report%20final%20140222.pdf [accessed 2023-08-10]
85. Dorney-Smith S, Burridge S, Bell J, Ellis J, Snowball L. Digital health inclusion for people who have experienced homelessness- is this a realistic aspiration? Pathway. URL: https://www.pathway.org.uk/wp-content/uploads/2013/05/Digital-Inclusion-Report-Final-for-web.pdf [accessed 2023-09-29]
86. Wunsch C, Wightman R, Pratty C, Jacka B, Hallowell BD, Clark S, et al. Thirty-day treatment continuation after audio-only buprenorphine telehealth initiation. J Addict Med 2023 Mar;17(2):206-209. [doi: 10.1097/ADM.0000000000001077] [Medline: 36102540]
87. Yeo EJ, Kralles H, Sternberg D, McCullough D, Nadanasabesan A, Mayo R, et al. Implementing a low-threshold audio-only telehealth model for medication-assisted treatment of opioid use disorder at a community-based non-profit organization in Washington, D.C. Harm Reduct J 2021 Dec 09;18(1):127 [FREE Full text] [doi: 10.1186/s12954-021-00578-1] [Medline: 34886850]
88. Kennedy AJ, George JS, Rossetti G, Brown CO, Ragins K, Dadiomov D, et al. Providing low-barrier addiction treatment via a telemedicine consultation service during the COVID-19 pandemic in Los Angeles, county: an assessment 1 year later. J Addict Med 2023 Jan;17(1):e64-e66 [FREE Full text] [doi: 10.1097/ADM.0000000000001034] [Medline: 35839323]
89. Brandee I, Frances M, Sheri D. UK pharmacies offer sense of normalcy for methadone patients. PEW Charitable Trust. URL: https://pew.org/41G01fi [accessed 2023-08-11]
90. Guille C, Simpson A, Douglas E, Boyars L, Cristaldi K, McElligott J, et al. Treatment of opioid use disorder in pregnant women via telemedicine: a nonrandomized controlled trial. JAMA Netw Open 2020 Jan 03;3(1):e1920177 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20177] [Medline: 32003816]
91. Lambdin BH, Kan D, Kral AH. Improving equity and access to buprenorphine treatment through telemedicine at syringe services programs. Subst Abuse Treat Prev Policy 2022 Jul 15;17(1):51 [FREE Full text] [doi: 10.1186/s13011-022-00483-1] [Medline: 35841036]
92. Regulatory approaches to telemedicine. Europe Economics. General Medical Council. 2022. URL: https://www.gmc-uk.org/about/what-we-do-and-why/data-and-research/research-and-insight-archive/regulatory-approaches-to-telemedicine [accessed 2023-09-29]
93. Castillo M, Conte B, Hinkes S, Mathew M, Na CJ, Norindr A, et al. Implementation of a medical student-run telemedicine program for medications for opioid use disorder during the COVID-19 pandemic. Harm Reduct J 2020 Nov 17;17(1):88 [FREE Full text] [doi: 10.1186/s12954-020-00438-4] [Medline: 33203460]
94. Khurana MP, Raaschou-Pedersen DE, Kurtzhals J, Bardram JE, Ostrowski SR, Bundgaard JS. Digital health competencies in medical school education: a scoping review and Delphi method study. BMC Med Educ 2022 Feb 26;22(1):129 [FREE Full text] [doi: 10.1186/s12909-022-03163-7] [Medline: 35216611]
95. Using video consultations in secondary and hospital care: a toolkit for clinicians. DigitalHealth. URL: https://digitalhealth.wales/sites/default/files/2020-10/Using%20Video%20Consultations%20in%20Secondary%20Care%20V1.2%20141020.pdf [accessed 2022-11-22]
96. TIP 63: medications for opioid use disorder for healthcare and addiction professionals, policymakers, patients, and families. Substance Abuse and Mental Health Services Administration. 2021. URL: https://store.samhsa.gov/product/TIP-63-Medications-for-Opioid-Use-Disorder-Full-Document/PEP21-02-01-002 [accessed 2022-11-22]
97. International standards for the treatment of drug use disorders: revised edition incorporating results of field-testing. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/international-standards-for-the-treatment-of-drug-use-disorders [accessed 2023-09-29]
98. Antonacci G, Lennox L, Barlow J, Evans L, Reed J. Process mapping in healthcare: a systematic review. BMC Health Serv Res 2021 Apr 14;21(1):342 [FREE Full text] [doi: 10.1186/s12913-021-06254-1] [Medline: 33853610]
99. An evaluation of our rapid implementation support for NHS Near Me 23 March-30 June 2020. Healthcare Improvement Scotland. 2020. URL: https://ihub.scot/media/7355/20200901-near-me-interactive-report-v12-final.pdf [accessed 2022-07-29]
100. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the telemedicine satisfaction questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003 Jun 24;9(1):46-50. [doi: 10.1258/135763303321159693] [Medline: 12641893]

101. Parmanto B, Lewis ANJ, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016 Jul 01;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
102. Otto L, Whitehouse D, Schlieter H. On the road to telemedicine maturity: a systematic review and classification of telemedicine maturity models. BLED 2019 Proceedings. 2019. URL: https://aisel.aisnet.org/bled2019/35/ [accessed 2023-09-29]
103. van Dyk L, Fortuin J, Schutte C. A maturity model for telemedicine implementation. In: Proceedings of the 4th International Conference on eHealth, Telemedicine, and Social Medicine. 2012 Presented at: eTELEMED '12; January 30-February 4, 2012; Valencia, Spain p. 78-84 URL: https://www.researchgate.net/publication/267426846_A_Maturity_Model_for_Telemedicine_Implementation
104. Cole TO, Robinson D, Kelley-Freeman A, Gandhi D, Greenblatt AD, Weintraub E, et al. Patient satisfaction with medications for opioid use disorder treatment via telemedicine: brief literature review and development of a new assessment. Front Public Health 2020;8:557275 [FREE Full text] [doi: 10.3389/fpubh.2020.557275] [Medline: 33553083]
105. McGrath LB, Eaton J, Abecassis IJ, Maxin A, Kelly C, Chesnut RM, et al. Mobile smartphone-based digital pupillometry curves in the diagnosis of traumatic brain injury. Front Neurosci 2022 Jul 1;16:893711 [FREE Full text] [doi: 10.3389/fnins.2022.893711] [Medline: 35844221]
106. Virtual MOUD treatment: virtual point-of-care toxicology testing to accompany virtual medication assisted treatment for opioid use disorder. US National Library of Medicine. 2022. URL: https://clinicaltrials.gov/ct2/show/NCT05448118 [accessed 2022-11-21]

Abbreviations

ADP: Alcohol and Drug Partnership
CRM: clinical risk management
DRD: drug-related death
MAT: medication-assisted treatment
MOUD: medication for opioid use disorder
OUD: opioid use disorder
SHIFT-Evidence: Successful Healthcare Improvements From Translation of Evidence Into Practice
TMOUD: telemedicine-delivered medication for opioid use disorder

Edited by J Torous; submitted 10.03.23; peer-reviewed by J Reed, K Johnson; comments to author 26.05.23; revised version received 16.08.23; accepted 02.09.23; published 18.10.23.

Please cite as:
Tay Wee Teck J, Gittins R, Zlatkute G, Oteo Pérez A, Galea-Singer S, Baldacchino A
Developing a Theoretically Informed Implementation Model for Telemedicine-Delivered Medication for Opioid Use Disorder: Qualitative Study With Key Informants
JMIR Ment Health 2023;10:e47186
URL: https://mental.jmir.org/2023/1/e47186
doi:10.2196/47186
PMID:37851506

Review

When Heart Beats Differently in Depression: Review of Nonlinear Heart Rate Variability Measures

Milena Čukić[1], PhD; Danka Savić[2], PhD; Julia Sidorova[3], PhD

[1]Empa Materials Science and Technology, Empa Swiss Federal Institute, St Gallen, Switzerland

[2]Vinča Institute for Nuclear Physics, Laboratory of Theoretical and Condensed Matter Physics 020/2, Vinca Institute, University of Belgrade, Belgrade, Serbia

[3]Bioinformatics Platform, Hospital Clínic, Barcelona, Spain

Corresponding Author:
Milena Čukić, PhD
Empa Materials Science and Technology
Empa Swiss Federal Institute
Lerchenfeldstrasse 5
St Gallen, 9014
Switzerland
Phone: 41 +41587657070
Fax: 41 +41587657070
Email: milena.cukic@gmail.com

Abstract

Background: Disturbed heart dynamics in depression seriously increases mortality risk. Heart rate variability (HRV) is a rich source of information for studying this dynamics. This paper is a meta-analytic review with methodological commentary of the application of nonlinear analysis of HRV and its possibility to address cardiovascular diseases in depression.

Objective: This paper aimed to appeal for the introduction of cardiological screening to patients with depression, because it is still far from established practice. The other (main) objective of the paper was to show that nonlinear methods in HRV analysis give better results than standard ones.

Methods: We systematically searched on the web for papers on nonlinear analyses of HRV in depression, in line with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 framework recommendations. We scrutinized the chosen publications and performed random-effects meta-analysis, using the *esci* module in jamovi software where standardized effect sizes (ESs) are corrected to yield the proof of the practical utility of their results.

Results: In all, 26 publications on the connection of nonlinear HRV measures and depression meeting our inclusion criteria were selected, examining a total of 1537 patients diagnosed with depression and 1041 healthy controls (N=2578). The overall ES (unbiased) was 1.03 (95% CI 0.703-1.35; diamond ratio 3.60). We performed 3 more meta-analytic comparisons, demonstrating the overall effectiveness of 3 groups of nonlinear analysis: detrended fluctuation analysis (overall ES 0.364, 95% CI 0.237-0.491), entropy-based measures (overall ES 1.05, 95% CI 0.572-1.52), and all other nonlinear measures (overall ES 0.702, 95% CI 0.422-0.982). The effectiveness of the applied methods of electrocardiogram analysis was compared and discussed in the light of detection and prevention of depression-related cardiovascular risk.

Conclusions: We compared the ESs of nonlinear and conventional time and spectral methods (found in the literature) and demonstrated that those of the former are larger, which recommends their use for the early screening of cardiovascular abnormalities in patients with depression to prevent possible deleterious events.

(JMIR Ment Health 2023;10:e40342) doi:10.2196/40342

KEYWORDS

heart rate variability; HRV; electrocardiogram; ECG; depression; autonomous nervous system; ANS; nonlinear measures; cardiac risk; cardiovascular; mortality; heart dynamics; ECG analysis; analysis; online

Introduction

Cardiovascular diseases (CVDs) are the number one cause of death globally according to the World Health Organization [1,2]. Depression is the number one mental health–related contributors to the global burden of disease [3]. When combined, these 2 diseases can lead to increased mortality risk [4-6]. Recently, the European Society of Cardiology published a position paper about the mechanisms linking depression and CVD, based on abundant evidence from literature [7]. Although this connection was discovered a long time ago [8-10], the CVD screening of patients with depression is still far from routine.

In nearly 70% of patients with depression, somatic symptoms, such as lack of energy, sleep disturbance, lack of appetite, decreased sex drive, general pains, etc, dominate the clinical picture [11]. These symptoms are due to autonomous nervous system (ANS) dysfunction. Heart rate variability (HRV) is regulated by the ANS, and its disturbance is a marker of CVD. The relation between HRV and depression has been well understood [7,12-15]. We registered at least 14 reviews that meta-analytically compared conventional methods of analysis of this relation [16-28].

Medical professionals interested in the detection of depression may be uninformed of the knowledge and methods offering additional insights into a patient's condition, with the knowledge coming from theoretical research—mathematical analysis, complex systems dynamic theory, and information theory. These methods can be used to extract information embedded in electrophysiological signals, represented as time series—electrocardiogram (ECG), electroencephalogram, electromyogram, etc. Current view of what electrophysiological signals can yield is quite obsolete and limited by a reductionist approach established in clinical practice, because most devices for recording physiological signals have built-in algorithms based on Fourier analysis [29]. These standard (time and frequency) methods of electrical signal analysis are designed for (predictable) electro-mechanical systems and are not well suited for (complex) physiological systems. A number of review studies [16-27] offer very detailed comparative analyses of time and frequency measures of HRV related to depression. They rely on the assumption that the dynamics of the system may be linearized, where valuable information is lost in the case of electrophysiological signals.

Physiological systems are complex. Complex systems are composed of multiple subunits that interact in a nonlinear fashion producing unpredictable behaviors [29,30]. Although homeostasis is usually perceived as a still condition, "healthy heartbeat displays highly complex, apparently unpredictable fluctuations even under steady-state conditions" [29], whereas heart failure, for example, shows "slow periodic oscillations that correlate with Cheyne-Stokes breathing" [31]. The theory of complex dynamic systems applies to such a system. Its behavior can be predicted at best for short intervals, and it is characterized by long-range correlations and organized variability.

In information theory, the rate at which a system is producing information is described by Shannon entropy (ShanEn)—a quantity reflecting the number of possible states a system can occur in, that is, the level of uncertainty (unpredictability). Pincus et al [32,33] adapted the ShanEn for cardiology research and devised the approximate entropy (ApEn) algorithm, a statistic quantifying serial irregularity. Further, Richman and Moorman [34] refined this measure into sample entropy (SampEn), which was later improved by Costa et al [35], proposing the multiscale entropy (MSE) algorithm that calculates irregularity changes on multiple scales [35]. Costa et al [36] performed a series of studies focusing on methods of analysis of ECG, and their work was a significant step in the acceptance of nonlinear methodology. Translated to signals, the higher the entropy, the higher the irregularity of a signal, which is most often interpreted as higher complexity. This "awkward fact," as Vargas et al [30] noted, is paradoxical as complexity assumes a structure that is highly ordered. Glass and Mackey [37] stated that "Random outputs result from degraded control mechanisms and/or breakdown of the coupling among them," that is, from the loss of complexity. Nevertheless, as much as this confusion makes the insights into control mechanisms more difficult, the measures of complexity or irregularity differ between health and disease rendering them suitable for nonspecific markers of ill-health.

Neural control mechanisms, which demonstrate fractal properties, generate "organized variability" (previously thought to be the "noise" in the signal) characteristic of a healthy physiological system [38]. Physiological systems are scale-free; self-similar fluctuations are observed on different time scales. From one moment to the next, the fluctuations detected in the same signal are quite variable [31,32]. A system that is fractal can demonstrate irregularity across a wide range of scales, but the type of "disorder" or "roughness" on different scales is statistically similar [39-41]. Goldberger et al [31] stated that "organized variability is an inevitable consequence of fractal self-organization." According to the number of publications (in cardiology), the most popular fractal-based methods in use for analyzing HRV is detrended fluctuation analysis (DFA), which is based on correlation properties and uses random walk [42].

In interpreting the results and choosing the measures to be used, the physical meaning of the applied nonlinear analysis and the physiological context of a particular disease have to be kept in mind. Beside entropy- and fractal-based measures, there are other families of nonlinear measures that are methodologically very different. Poincaré plots are among the most accurate measures applied in cardiology [43], and being a graphical representation, they are very convenient for clinical application. Largest Lyapunov exponents (LLE) [44], which were used often in the beginning of the field, detect the level of chaos in a signal. Lempel-Ziv complexity (coming from information theory) quantifies the uncertainty contained in time-series data and is still among the frequently used measures in physiology [45]. Several correlation-based measures also showed promising results, but to describe even the basic methodology for all of them is out of the scope of this manuscript.

The fundamental difference between irregularity statistics and conventional variability measures is that the conventional approach is focusing on tasks of quantifying the degree of spread around the central value while the order of the input data is

irrelevant; whereas in irregularity statistics, nonlinear measures track changes from random to very regular and the order of samples is essential to the algorithm [40]. Nonlinear measures have been shown to be very effective in detecting the slightest differences between healthy and ill heart dynamics—even when time series of the compared states are varying around almost the same mean values [29,41]. An impressive example of the advantage of nonlinear methods is the case of detecting sudden infant death syndrome based on entropy measures calculated from ECG; the standard method was not able to detect any difference between healthy and babies under a serious risk [41].

This study is a random-effects meta-analysis of the most important studies that used nonlinear methods to confirm the connection between HRV (as a marker of CVD) and depression. We calculated effect sizes (ESs) from these studies and compared them with the ESs of standard (conventional) methods found in the literature. The aim of this work was to help convince clinicians to (1) introduce cardiological screening to patients with depression, since depression is confirmed to bear a risk for CVD [10,16,22,46-48] and (2) apply nonlinear methods to HRV analysis for more accurate and reliable screening results.

Methods

Overview

Since there are a considerable number of recently published meta-analytic studies [16-27] regarding the classical (spectral or conventional) approach to analyzing HRV, we decided to include only those studies that performed any nonlinear method of analysis or had mixed analytic approaches (applying both standard and nonlinear methods of analysis of heart rhythm to compare the effectiveness of analytics). This meta-analysis was performed in agreement with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines [49] with the main aim to present an integrated, realistic ES of the nonlinear measures in distinguishing between major depressive disorder (MDD) and controls, in comparison to conventionally used measures reported in previously published meta-analytic studies. Our work does not compare effects of interventions; thus, it was not preregistered (ie, the review protocol registration does not exist).

Only case-control studies or longitudinal studies that used ECG recordings or measurements of heart rate or any automated ECG diagnosis (for example, including early arrhythmia detection) and successive nonlinear analyses were included in our pool. Our query was kept as broad as possible to retrieve as many relevant papers as possible. We searched web resources, such as Springer, Wiley, Elsevier, IEEE, National Institute of Mental Health, Frontiers, PlosOne, Hindawi, Web of Science, PubMed, Cochrane Library, Scopus, and National Center for Biotechnology Information, with sets of keywords: (“depression” OR “Major depressive disorder” OR “bipolar depression”) AND (“ECG” OR “electrocardiography” OR “HRV” OR “heart rate variability”) AND (“Nonlinear analysis” OR “Fractal analysis” OR “entropy”).

We then scrutinized the abstracts and full texts (in English) and discarded those that were purely theoretical considerations of the connection of HRV and depression without the quantification of HRV measures (even when they included keywords from all 3 sets); papers on mood disorders without separate data on depression; studies where samples comprised solely of healthy subjects or subjects under the age of 18 years; studies without age-matched controls; animal studies; and papers without peer review.

Depression was diagnosed using the Diagnostic and Statistical Manual of Mental Disorders (DSM; DSM-III-R, DSM-IV-TR, and DSM-5), International Classification of Diseases 10th Revision, Mini-International Neuropsychiatric Interview, Beck Depression Inventory, or Montgomery-Åsberg Depression Rating Scale. Some studies, found as references through hand searching of citation lists in review papers, were downloaded from the ResearchGate platform or were included even if nonretrievable as full text. Papers with standard analyses of ECG, added for comparison, were found in part via web-based search, from reference lists of review papers, and from the authors who were kind to send us full texts. The web search was performed in February and March 2021. Our screening was finalized in October 2021.

During the process (after removing the duplicates), we kept the working sheet (updated by all authors) with the basic data extracted from every paper included in the study (the first author’s name and the year of publication as an identifier, sample size, the mean and SD calculated for groups, the measures used in the research, the effect detected, and specific observations about the accuracies of applied analyses). Where we were unable to extract the data, we used WebPlotDigitizer software [50]. After coding all the data from the included papers, ES estimates were transformed into the same metric to be compared (Cohen’s d_s was corrected, hence having the same value as Hedges’ g). We used the *esci* module in jamovi exploratory software (open-source statistical software written in R) [50] to calculate overall ESs (correction of Cohen’s d) and 95% CIs and to generate forest plots [51]. The forest plots were used to visually display the individual and overall ESs.

ES is a quantitative description of the strength of evidence about a phenomenon. Cohen’s d describes the standardized mean difference of an effect [52-55]. In between 2 groups of independent observations, Cohen’s d_s is:

where M_1 and M_2 are the variable means of the 2 groups (patients and controls); in the denominator, the pooled SD is the Bessel correction for bias in the estimation of population variance (based on the least squares estimator [54,55]); SD_1 and SD_2 are the respective SDs; and n_1 and n_2 are the sample sizes of the groups. Cohen’s d_s is also directly related to the t test:

where t is the t statistic, and n_1 and n_2 are as above. This is a direct relation between the ES and statistical significance. Here,

statistical significance is expressed regardless of whether the 95% CI around Cohen's d_s includes 0 or not. Hedges and Olkin [54] showed that the formula for Cohen's *d* based on sample averages gives a biased estimate of the population ES (especially for smaller sample sizes, n<20). The Cohen's *d* that we calculated is actually Cohen's d_s, described in Lakens [55] (where SDs are pooled as in the formula above, not a single average of both SDs from samples 1 and 2). Thus, it is the Cohen's *d* of a sample, d_s. Further to be corrected for biases, according to Hedges and Olkin [54], it must be multiplied by another Bessel correction (1 – 3 / 4(N1 + N2) – 9) [55]. After calculating the corrected Cohen's d_s for all the included studies, by applying pooled SD in the process, we confirmed that what is calculated as a correction (for biases) in the *esci* module in jamovi software is actually Hedges' *g*. The authors of the software also describe that the product of their calculation (included in forest plot that the program is generating) is equal to Hedges' *g* [51,56]. In several studies that reported *t* values, we calculated Cohen's d_s according to the second formula. This interpretation was done based on previous literature [51-56].

Ethical Considerations

Since all the studies included in our review have already received prior approvals from their local ethics committees, and we did not use nor collect any additional data from the patients and only reanalyzed already published data, we do not report any ethic approval for this particular study.

Results

Our initial search (based on the logical formula given in *Methods*) in the abovementioned web services yielded 867 papers. The elimination was performed through phases shown in the flow chart (Figure 1) showing the identification, screening, eligibility check, and inclusion of studies in accordance with PRISMA 2020 [49]. The chosen 26 papers originated from the following databases: Elsevier (n=10), PubMed (n=8), Frontiers (n=2), Web of Science (n=4), Springer (n=1), and IEEE (n=1). They encompassed a total of 1537 patients diagnosed with depression and 1041 healthy controls. The studies included those that used nonlinear analyses or both nonlinear and standard analyses.

Direct quantitative comparisons could not have been made, as studies varied in methodologies, as well as in research questions—detecting biomarkers or predictors of depression, CVD mortality risk estimation or risk analysis, effects of different therapies, etc. Therapies included those exploring medication effects or spillover on HRV and psychological or psychiatric interventions; some examined inflammation or other important physiological markers, but some also used historical medical data (for example, from Medicare archives in the United States, see [57-59]). As the effects of therapies are not the topic of this paper, we compared the studies grouped by family of measures used (in nonlinear analysis) and summarized their results and conclusions concerning only the detection of the relation between depression and CVD mortality risk.

Figure 2 shows the information about the studies' methodologies (the majority [19/26, 73%] used more than one; only 7 [27%] studies were based on one nonlinear measure) and their conclusions. After initial random-effects meta-analysis of overall ESs of all included studies, we identified 3 distinct (methodological) groups of research and performed 3 additional meta-analyses. The first group used DFA (8 studies) with reported Cohen's d_s (corrected for biases) and 95% CIs. The second group of studies used methods from the large family of entropy measures: ApEn in 5 studies, SampEn in 5 studies, MSE in 3 studies, and ShanEn, Renyi entropy and refined composite multiscale entropy, and multilag tone-entropy, each in one study. The third group comprised various nonlinear analyses: Poincaré plots (n=4), LLE (n=2), symbolic measures (n=2), Lempel-Ziv complexity (n=1), complex variability, mutual information, autonomic information flow, beat decay NN, logarithmic respiratory sinus arrhythmia, recurrence plot analysis, Complex Correlation Measure, correlation dimension, and Katz fractal dimension. This group also demonstrates a historical order in which nonlinear measures entered cardiology, and some of them are still very popular in health applications (for example, LLE or Poincaré plots). Besides, this "historical" group demonstrated an average ES more than 2 times higher than any prior conventional approach, to the best of our knowledge.

The forest plot (and the table with Cohen's d_s corrected and 95% CIs) is used to visualize those meta-analytic comparisons of the ESs. Figure 2 represents the overall meta-analytic comparison of the best ESs of the 26 included studies. Figure 3 represents meta-analysis of the DFA group (with 8 studies compared, with an overall ES of 0.364, 95% CI 0.237-0.491). Figure 4 represents meta-analysis of the entropy group (15 studies compared, with an overall strong ES of 1.05, 95% CI 0.572-1.52). Figure 5 represents all other nonlinear methods used in the examined studies (13 studies compared, with an overall ES of 0.702, 95% CI 0.422-0.982). All mentioned comparisons yielded $P<.01$ (as shown the tables in all the figures). As the majority of studies applied several types of nonlinear analysis, we did multiple comparisons of corrected ESs that we separately calculated for each method. The best ES (d_s=7.7, 95% CI 6.4193-8.997) was obtained for the study [60] that used 4 entropy algorithms (ApEn, SampEn, ShanEn, and fuzzy entropy). When calculating ESs for each entropy method separately, ShanEn performed the best, and fuzzy entropy did not yield a significant result.

From overall meta-analytic comparison of ESs, out of 26 studies included, 11 were shown to be statistically insignificant [59-70], and one paper [71] had the lower CI touching the zero line, implying border significance. Thus, 15 (58%) out of 26 studies were statistically significant, with the overall ES of corrected Cohen's d_s=1.03 (elsewhere also reported as Hedges' *g*), which is a large ES [52] and can be translated to more than 1 SD [55].

Figure 1. Flow chart representing the procedure of choosing 26 studies included for this review. ECG: electrocardiogram; HC: healthy controls; HRV: heart rate variability.

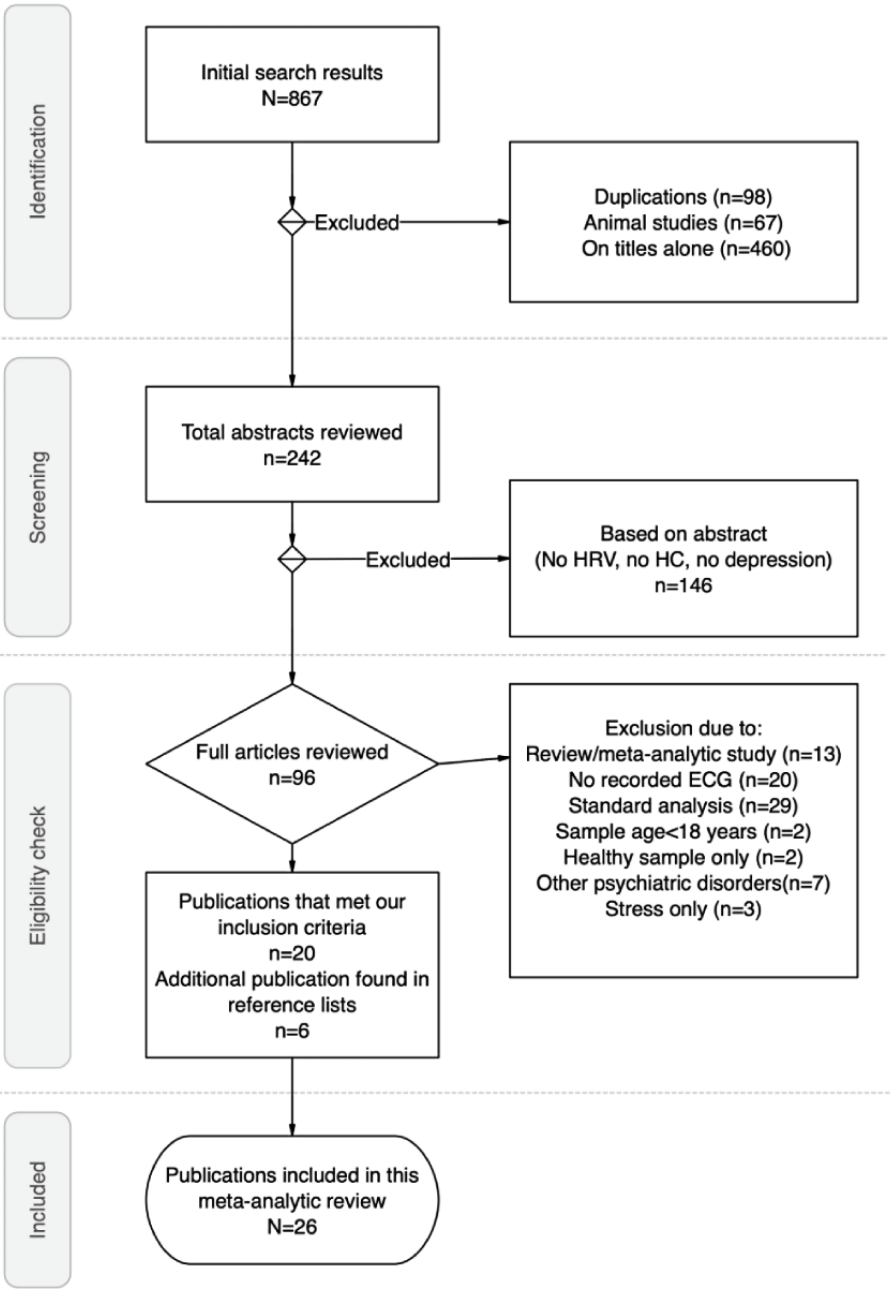


Figure 2. Forest plot and the table of random-effects meta-analysis showing the corrected effect size and CIs, as well as both sample sizes (N1: patients diagnosed with depression; N2: controls). For each study performing more than one method of nonlinear HRV analysis, we presented the largest effect size. Both table and forest plot are generated by the esci module in jamovi software. In all, 15 (58%) out of 26 included studies were shown to have statistically significant results. The overall effect size (unbiased/corrected) is 1.03 (diamond ratio 3.60). HRV: heart rate variability. [14,44,45,57-79].

Meta-Analysis Results

		95 % CI				95 % CI	
Label	Cohen's d	Lower	Upper	p	Diamond Ratio	Lower	Upper
Overall	1.03	0.703	1.35	<.001	3.60	2.70	5.47

Table of Studies

		95 % CI			
Label	Cohen's d	Lower	Upper	N1	N2
Moser 1998	0.1379	-0.40635	0.682	26	26
YeraganiRao 2002	1.3842	0.60779	2.161	14	18
Yeragani 2002a	0.5194	-1.47331	2.512	2	2
Kop 2010	0.2578	0.06007	0.455	776	113
Vigo 2004	0.9958	0.39976	1.592	19	33
Iverson 2005	1.2115	0.61407	1.809	30	22
Pincus 2008	0.5600	-0.41079	1.531	9	8
Boettger 2008	0.6649	-0.00624	1.336	18	18
Bob 2009	0.7918	0.32075	1.263	40	35
Baumert 2009	0.4386	-0.46622	1.344	12	8
Schulz 2010	0.5900	0.21498	0.965	57	57
Jelinek 2011	0.6968	0.38128	1.012	73	93
Voss 2011	0.4155	-0.05144	0.882	36	36
Yang 2011	0.2725	-0.07180	0.617	52	88
Leistedt 2011	1.3754	0.72238	2.028	25	20
Kemp 2012	0.5176	0.20687	0.828	73	94
Berger 2012	0.9348	0.24668	1.623	18	18
Migliorini 2012	0.2123	-0.99092	1.416	4	8
Moon 2013	0.2271	-0.27977	0.734	34	27
Kemp 2014	0.5574	0.24468	0.870	72	94
Khandoker 2017	7.1886	5.58294	8.794	16	29
Valenza 2017	0.0436	-0.35648	0.444	48	48
Chen 2017	0.3130	-0.12798	0.754	40	40
Greco 2017	3.5108	2.69758	4.324	36	24
Byun 2019	7.7079	6.41931	8.997	37	41
Byun 2019a	0.4467	-0.04181	0.935	33	33

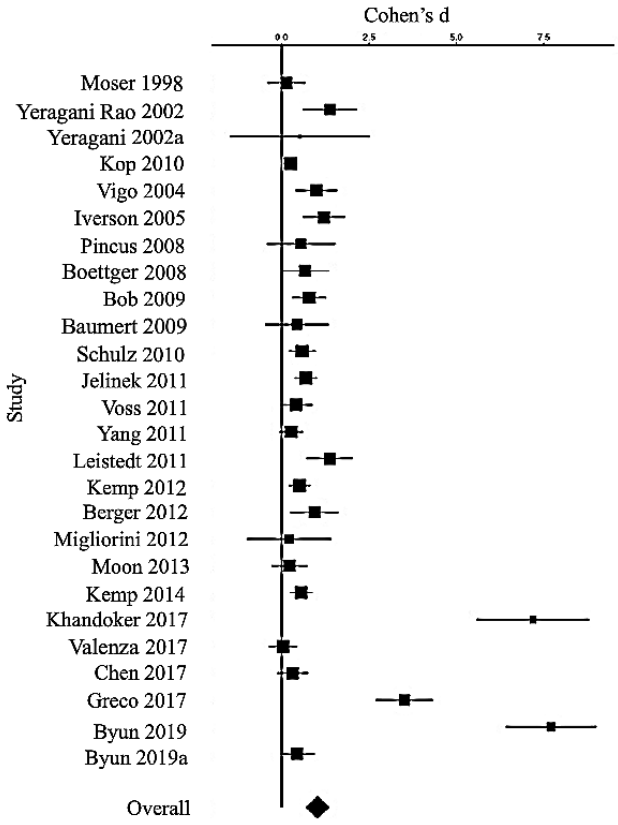


Figure 3. Forest plot (and table) showing the random-effects meta-analysis of a group of selected papers (8 publications [45,58,59,64,73,74,77,79]) that used detrended fluctuation analysis generated by the esci module in jamovi software. In all, 4 out of 8 studies were shown to have statistically nonsignificant results, with the overall effect size (biases corrected) being 0.364 (between small and medium effect, closer to medium, according to Cohen [52]), and the diamond ratio is 1.

Meta-Analysis Results

		95 % CI				95 % CI	
Label	Cohen's d	Lower	Upper	p	Diamond Ratio	Lower	Upper
Overall	0.364	0.237	0.491	<.001	1.00	1.00	8.82

Table of Studies

		95 % CI			
Label	Cohen's d	Lower	Upper	N1	N2
Vigo 2004	0.7288	0.1473	1.310	19	33
Baumert 2009	0.1714	-0.7247	1.068	12	8
Migliorini 2012	0.0277	-1.1726	1.228	4	8
Greco 2017	0.2452	-0.2732	0.764	24	36
Jelinek 2011	0.4977	0.1866	0.809	73	93
Kemp 2012	0.5176	0.2069	0.828	73	94
Kop 2010	0.2578	0.0601	0.455	776	113
Schulz 2010	0.3079	-0.0614	0.677	57	57

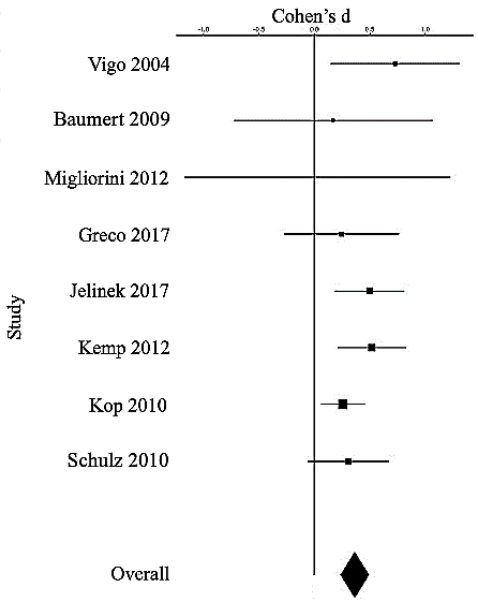


Figure 4. Forest plot (and table) showing the random-effects meta-analysis of a group of selected papers (15 publications [45,57,60,63-67,69,70,72,73,77-79]) that used entropy measures: approximate entropy (ApEn), logarithmic ApEn, sample entropy, fuzzy entropy, Shannon entropy, cross entropy, Renyi entropy, multiscale entropy, and improved refined composite multiscale entropy, generated by the esci module in jamovi software. The overall effect size for this group is 1.05 (which is very large, according to Cohen [52]). In all, 10 out of 18 studies were shown to have statistically nonsignificant results. Interestingly, Byun et al [60] calculated 4 entropy measures and demonstrated that only Shannon entropy yielded a highly useful effect size of 7.7 (which is the best result detected in this entire pool of publications, followed by Khandoker et al [72] with a corrected effect size of 7.3).

Meta-Analysis Results

		95 % CI				95 % CI	
Label	Cohen's d	Lower	Upper	p	Diamond Ratio	Lower	Upper
Overall	1.05	0.572	1.52	<.001	3.62	2.67	5.69

Table of Studies

		95 % CI			
Label	Cohen's d	Lower	Upper	N1	N2
Vigo 2004	0.9958	0.3998	1.592	19	33
Baumert 2009	0.4386	-0.4662	1.344	12	8
Migliorini 2012	0.2123	-0.9909	1.416	4	8
Moon 2013	0.2271	-0.2798	0.734	34	27
Berger 2012	0.9348	0.2467	1.623	18	18
Greco 2017	0.5438	0.0183	1.069	36	24
Byun 2019	7.7079	6.4193	8.997	37	41
Byun 2019a	0.4467	-0.0418	0.935	33	33
Byun 2019a	0.3666	-0.1199	0.853	33	33
Byun 2019a	0.4299	-0.0582	0.918	33	33
Shulz 2010	0.3278	-0.0418	0.697	57	57
Voss 2011	0.4155	-0.0514	0.882	36	36
Pincus 2008	0.0180	-0.9344	0.970	9	8
Pincus 2008	0.3388	-0.6203	1.298	9	8
Yang 2011	0.2725	-0.0718	0.617	52	88
Leistedt 2011	1.3754	0.7224	2.028	25	20
Chen 2017	0.3130	-0.1280	0.754	40	40
Khandoker 2017	7.1886	5.5829	8.794	16	29

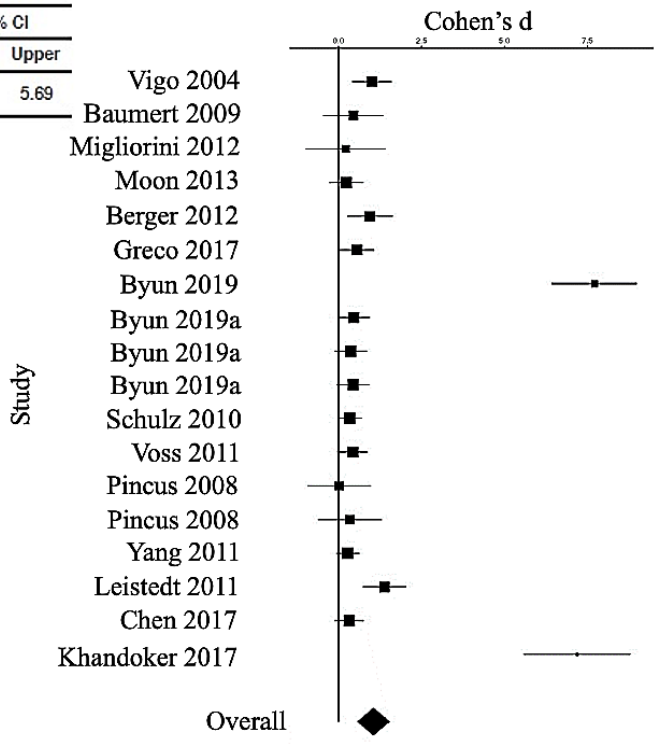


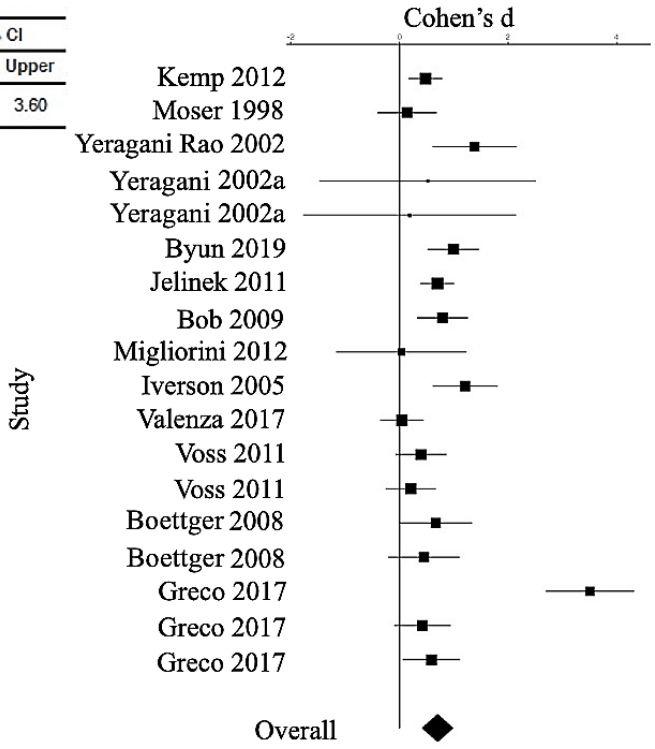
Figure 5. Forest plot (and table) showing the random-effects meta-analysis of group of selected papers that used a number of different nonlinear analyses of HRV in depression, excluding entropy and detrended fluctuation analysis measures. Those are Poincaré plots (SD1 and SD2), largest Lyapunov exponents, symbolic measures, Katz fractal dimension, correlation dimension, Complex Correlated Measure, mutual information, logarithmic respiratory sinus arrhythmia, heart rate turbulence, Lempel-Ziv complexity, recurrence plot analysis (determinism and recurrent rate are the most prominent result in Greco et al [73], ds=3.5), autonomic information flow, and beat decay NN. The plot and table are generated by the esci module in jamovi software. The overall effect size of studies included in this group is 0.7 (diamond ratio 2.3), which is considered large. Eight studies from this group were shown to have practically meaningful results: Kemp et al [74], Yeragani, Rao et al [44], Byun et al [70], Iverson et al [75], Jelinek et al [59], Bob et al [76], and Greco et al [73] (2 out of the 3 applied measures included and represented in this group). [44,45,59,61,62,65,68,70,71,73-76].

Meta-Analysis Results

		95 % CI				95 % CI	
Label	Cohen's d	Lower	Upper	p	Diamond Ratio	Lower	Upper
Overall	0.702	0.422	0.982	<.001	2.30	1.71	3.60

Table of Studies

		95 % CI			
Label	Cohen's d	Lower	Upper	N1	N2
Kemp 2012	0.4778	0.16779	0.788	94	73
Moser 1998	0.1379	-0.40635	0.682	26	26
YeraganiRao 2002	1.3842	0.60779	2.161	18	14
Yeragani 2002a	0.5194	-1.47331	2.512	2	2
Yeragani 2002a	0.1880	-1.77629	2.152	2	2
Byun 2019	0.9931	0.52211	1.464	41	37
Jelinek 2011	0.6968	0.38128	1.012	93	73
Bob 2009	0.7918	0.32075	1.263	35	40
Migliorini 2012	0.0369	-1.16340	1.237	8	4
Iverson 2005	1.2115	0.61407	1.809	22	30
Valenza 2017	0.0436	-0.35648	0.444	48	48
Voss 2011	0.3957	-0.07077	0.862	36	36
Voss 2011	0.2077	-0.25547	0.671	36	36
Boettger 2008	0.6649	-0.00624	1.336	18	18
Boettger 2008	0.4498	-0.21175	1.111	18	18
Greco 2017	3.5108	2.69758	4.324	24	36
Greco 2017	0.4213	-0.10071	0.943	24	36
Greco 2017	0.5863	0.05925	1.113	24	36



Discussion

Our results show that the overall standardized ES of nonlinear measures of HRV in depression overperforms the ESs of conventional measures of HRV reported in the literature. Although the overall ES in our comparison (all 26 studies) was 1.03, the best entropy-based group ES was 1.05, the DFA group ES was 0.36, and the third, miscellaneous group yielded an ES of 0.70. In the latest standard HRV measures-based study [28] (which is very similar to ours by the number of included studies), the ESs of several conventional measures varied up to 0.46. The meta-analysis by Rottenberg [13] reports a small ES (d~0.2), which explains only about 2% of the overall variance for conventional measures. Many of the published papers reported nonsignificant and mild-to-moderate ESs, whereas we found much larger effects: for example, the best ES in our research reached the value of 7.7 [60], and several others demonstrated higher ESs than those reported in conventional analyses (eg, the ES in Khandoker et al [72] was 7.3, which is very large according to Cohen [52]; in Greco et al [73], the ES was 3.5; and several others reached an ES around 1).

As commented in *Results*, almost half of the included studies did not have a significant effect in discriminating patients with depression from controls. This might be because not all patients with depression have disturbed heart function and because of the modest sample sizes. Indeed, the majority of those authors concluded that their initial results were promising but required replication. Later studies (2010 onward) show that researchers started using larger data sets or at least existing databases [58]. In the last 10 years, studies not only started using nonlinear measures but also combining them with some forms of machine learning to discern MDD [60,72,73,77,81,82]. We consider this methodological combination promising, especially since our previous work based on depression detection from electroencephalogram yielded good results [83-86].

Through the list of nonlinear methods presented in *Results*, we can follow the evolution of the understanding of how to interpret the results of nonlinear analysis. First, it was LLE, but since Lyapunov exponents mainly serve to detect chaos in signals, they can hardly be used for the precise delineation of groups. Then, in several papers, symbolic methods were used and reported as being successful, but again, their interpretation was problematic for which they were practically abandoned. The most promising family of nonlinear measures is the one based on entropy, in particular, ShanEn, then SampEn and ApEn (maybe also logarithmic ApEn); MSE seems a little more difficult to interpret. Irregular signals have higher entropies. Increased irregularity can point to a degradation of internal control mechanisms, or as Goldberger [29] puts it, decomplexification that is characteristic of aging and disease. Additionally, DFA, as a fractal method (as well as several other methods of calculating fractal dimension and correlation

dimension) makes sense, since neural control mechanisms are shown to have a fractal nature. In fact, all spectral measures calculated from ECG are a function of RR intervals length and are correlated; they do decrease with aging, but in disease that change is much more pronounced, and the function is lost [87-89]. In that sense, DFA is accurately detecting short- and long-term correlation that are important for healthy heart dynamics but also its synchronization with breathing [82,90]. Among other measures, the Complex Correlated Measure applied by Jelinek et al [59] performed quite well. Some recurrence plot analyses (Poincaré plot analysis and generalized Poincaré analysis) that quantify self-similarity in the processes were also used with good results [91]. In the reviewed literature, there are also combinations and alterations of the mentioned methods of analysis, such as combining the Poincaré analysis with DFA, applying Pearson coefficients on prior Poincaré analysis, or choosing the most prominent coefficients from several analyses and combining them as successful features for classification.

Nonlinear HRV analysis might be used as an aid in differential diagnosing or in indicating comorbidities [14,20,43,92]. For example, Chang et al [82] succeeded in distinguishing between bipolar II depression and unipolar depression, based on SampEn analysis of the HRV of 707 subjects. Kemp et al [14] found that anxiety disorders comorbid to MDD, most of all generalized anxiety disorder, contributed to the reduction of HRV. They elaborated on how nonvagal components of heart rate might further distinguish between subtypes of the disorder [20,43,92].

Cardiac vagal control (CVC) is associated with both physical and mental health. Low CVC is considered to be an indicator of risk of cardiac disease, including myocardial infarction and congestive heart failure [10,93]. Since variability in heart rate that is gated by the respiratory cycle [13,43] reflects the extent of CVC, it is logical to analyze its nonlinear dynamics and its aberrations to detect and treat depression. It could be a link between the polyvagal theory of Porges [94] and the physiological complexity (decomplexification and stereotypy of disease) of Goldberger [95]. In parallel to the polyvagal theory, there is also the neurovisceral integration model [96], both emphasizing the importance of taking into account ANS aberrations, along with the existing need to improve psychiatric nosology.

Important insights about healthy heart dynamics and how it changes with aging and disease were published in the 90s and served well the detection of several pathological entities [41,90,97]. We have learned that the mechanisms of neural control are fractal in nature (scale-free) and that they generate the so-called complex variability (once believed to be a background noise to the signal), which is a characteristic of healthy heart dynamics. In pathological states, one can observe a characteristic loss of complexity (decomplexification) that leads to recognizable oscillatory (predictable) behavior of a complex system, reduced to a single scale or frequency. The aberrated dynamics can be precisely quantified by fractal and nonlinear measures. The standard idea of comparison of healthy and ill organism pertains to the calculation of traditional mean values, SDs and the like, from electrophysiological signal (here ECG). When one compares the recording of a healthy heart with one of a patient diagnosed with congestive heart failure, their calculated means are within the same SD. However, it can be seen even with a naked eye that those signals are different (in dynamics and structure). Traditional methods do not show a significant difference here. The stereotypy of disease, as Goldberger explained [31], is connected with the decomplexification of a dynamical system's output, observed in early complexity studies. Complexity analysis can complement this clinical heuristic with adequate mathematical tools to quantify the changes in a patient's state.

Too aggressive preprocessing of the data can contribute to misleading results due to the loss of the exact order of samples. The history of the system is important in knowing its dynamics: earlier samples—the values of a physical phenomenon we measure (here in microvolts)—affect the later values, and if you shuffle the order, you lose the internal nonlinear structure that is contained in the sequence of those samples; this is called the historicity of data. Thus, it is necessary to analyze raw sequences of the records (broadband signal is the most information rich). This might be the reason why the nonlinear methods are superior to conventional ones.

The data can be easily obtained by novel portable ECG monitoring devices that are approved as medical-grade signal quality equivalent to a Holter monitor but are much more practical and comfortable to use by the patient herself or himself, taking only a couple of minutes. The data can then be processed by a combination of nonlinear analytics and advanced statistical procedures (to control, for example, for comorbidities and other confounding factors or for feature selection for further machine learning). Even better, the analysis can be empowered by machine learning applications that are widely in use due to the high power of computation and cloud computing [83-86,97].

To conclude, the ESs of nonlinear methods are larger than those of standard methods in HRV analysis. Measuring ECG and applying nonlinear analysis of HRV should enter the routine clinical practice for patients with depression. Although Porges [94] states that psychiatrists and psychologists seem not to be sufficiently interested in the use of objective biomarkers in their daily diagnostic work with patients with depression, the real question here is how ethical it is to keep this status quo and apply trial and error protocols in depression treatment without prior objective screening for CVD risks.

Acknowledgments

Part of this work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (DS). MC and JS did not receive any funding support for this work.

Authors' Contributions

MC contributed to conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing–original draft, and writing–review and editing. DS contributed to data curation, investigation, methodology, and writing–review and editing. JS contributed to data curation, investigation, and writing–review and editing.

Conflicts of Interest

None declared.

References

1. The global burden of disease: 2004 update. World Health Organization. 2008. URL: https://apps.who.int/iris/bitstream/handle/10665/43942/9789241563710_eng.pdf?sequence=1&isAllowed=y [accessed 2023-01-05]
2. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: http://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf;jsession [accessed 2023-01-05]
3. Herrman H, Kieling C, McGorry P, Horton R, Sargent J, Patel V. Reducing the global burden of depression: a Lancet-World Psychiatric Association Commission. Lancet 2019 Jun 15;393(10189):e42-e43. [doi: 10.1016/S0140-6736(18)32408-5] [Medline: 30482607]
4. Nicholson A, Kuper H, Hemingway H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. Eur Heart J 2006 Dec;27(23):2763-2774. [doi: 10.1093/eurheartj/ehl338] [Medline: 17082208]
5. Lange-Asschenfeldt C, Lederbogen F. Antidepressant therapy in coronary artery disease. Article in German. Nervenarzt 2011 May;82(5):657-64; quiz 665. [doi: 10.1007/s00115-010-3181-7] [Medline: 21109992]
6. Lett HS, Blumenthal JA, Babyak MA, Sherwood A, Strauman T, Robins C, et al. Depression as a risk factor for coronary artery disease: evidence, mechanisms, and treatment. Psychosom Med 2004;66(3):305-315. [doi: 10.1097/01.psy.0000126207.43307.c0] [Medline: 15184688]
7. Vaccarino V, Badimon L, Bremner JD, Cenko E, Cubedo J, Dorobantu M, ESC Scientific Document Group Reviewers. Depression and coronary heart disease: 2018 position paper of the ESC working group on coronary pathophysiology and microcirculation. Eur Heart J 2020 May 01;41(17):1687-1696 [FREE Full text] [doi: 10.1093/eurheartj/ehy913] [Medline: 30698764]
8. Malzberg B. Mortality among patients with involution melancholia. Am J Psychiatry 1937 Mar;93(5):1231-1238. [doi: 10.1176/ajp.93.5.1231]
9. Musselman DL, Evans DL, Nemeroff CB. The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. Arch Gen Psychiatry 1998 Jul 01;55(7):580-592. [doi: 10.1001/archpsyc.55.7.580] [Medline: 9672048]
10. Dhar AK, Barton DA. Depression and the link with cardiovascular disease. Front Psychiatry 2016 Mar 21;7:33 [FREE Full text] [doi: 10.3389/fpsyt.2016.00033] [Medline: 27047396]
11. Tylee A, Gandhi P. The importance of somatic symptoms in depression in primary care. Prim Care Companion J Clin Psychiatry 2005 Aug 15;7(4):167-176 [FREE Full text] [doi: 10.4088/pcc.v07n0405] [Medline: 16163400]
12. Messerotti Benvenuti S, Buodo G, Mennella R, Palomba D. Somatic, but not cognitive-affective, symptoms are associated with reduced heart rate variability in individuals with dysphoria. Front Psychol 2015 May 07;6:599 [FREE Full text] [doi: 10.3389/fpsyg.2015.00599] [Medline: 25999905]
13. Rottenberg J. Cardiac vagal control in depression: a critical analysis. Biol Psychol 2007 Feb;74(2):200-211. [doi: 10.1016/j.biopsycho.2005.08.010] [Medline: 17045728]
14. Kemp AH, Quintana DS, Quinn CR, Hopkinson P, Harris AWF. Major depressive disorder with melancholia displays robust alterations in resting state heart rate and its variability: implications for future morbidity and mortality. Front Psychol 2014 Nov 27;5:1387 [FREE Full text] [doi: 10.3389/fpsyg.2014.01387] [Medline: 25505893]
15. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017 Sep 28;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
16. Rugulies R. Depression as a predictor for coronary heart disease. a review and meta-analysis. Am J Prev Med 2002 Jul;23(1):51-61. [doi: 10.1016/s0749-3797(02)00439-7] [Medline: 12093424]
17. van Zyl LT, Hasegawa T, Nagata K. Effects of antidepressant treatment on heart rate variability in major depression: a quantitative review. Biopsychosoc Med 2008 Jun 30;2(1):12 [FREE Full text] [doi: 10.1186/1751-0759-2-12] [Medline: 18590531]
18. Carney RM, Freedland KE. Depression and coronary heart disease. Nat Rev Cardiol 2017 Mar;14(3):145-155. [doi: 10.1038/nrcardio.2016.181] [Medline: 27853162]
19. Kemp AH, Quintana DS, Gray MA, Felmingham KL, Brown K, Gatt JM. Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. Biol Psychiatry 2010 Jun 01;67(11):1067-1074. [doi: 10.1016/j.biopsych.2009.12.012] [Medline: 20138254]

20. Taylor CB, Conrad A, Wilhelm FH, Strachowski D, Khaylis A, Neri E, et al. Does improving mood in depressed patients alter factors that may affect cardiovascular disease risk? J Psychiatr Res 2009 Dec;43(16):1246-1252 [FREE Full text] [doi: 10.1016/j.jpsychires.2009.05.006] [Medline: 19577757]
21. Nemeroff CB, Goldschmidt-Clermont PJ. Heartache and heartbreak--the link between depression and cardiovascular disease. Nat Rev Cardiol 2012 Sep 26;9(9):526-539. [doi: 10.1038/nrcardio.2012.91] [Medline: 22733213]
22. Bassett D. A literature review of heart rate variability in depressive and bipolar disorders. Aust N Z J Psychiatry 2016 Jun 23;50(6):511-519. [doi: 10.1177/0004867415622689] [Medline: 26698824]
23. Gan Y, Gong Y, Tong X, Sun H, Cong Y, Dong X, et al. Depression and the risk of coronary heart disease: a meta-analysis of prospective cohort studies. BMC Psychiatry 2014 Dec 24;14(1):371 [FREE Full text] [doi: 10.1186/s12888-014-0371-z] [Medline: 25540022]
24. de la Torre-Luque A, Bornas X, Balle M, Fiol-Veny A. Complexity and nonlinear biomarkers in emotional disorders: a meta-analytic study. Neurosci Biobehav Rev 2016 Sep;68:410-422. [doi: 10.1016/j.neubiorev.2016.05.023] [Medline: 27267791]
25. Carr O, Andreotti F, Saunders KEA, Bilderbeck AC, Goodwin GM, de Vos M. Linking changes in heart rate variability to mood changes in daily life. Computing Cardiol 2017;44:1-4. [doi: 10.22489/cinc.2017.319-334]
26. Brown L, Karmakar C, Gray R, Jindal R, Lim T, Bryant C. Heart rate variability alterations in late life depression: a meta-analysis. J Affect Disord 2018 Aug 01;235:456-466. [doi: 10.1016/j.jad.2018.04.071] [Medline: 29679898]
27. Koch C, Wilhelm M, Salzmann S, Rief W, Euteneuer F. A meta-analysis of heart rate variability in major depression. Psychol Med 2019 Sep;49(12):1948-1957. [doi: 10.1017/S0033291719001351] [Medline: 31239003]
28. Klonowski W. From conformons to human brains: an informal overview of nonlinear dynamics and its applications in biomedicine. Nonlinear Biomed Phys 2007 Jul 05;1(1):5 [FREE Full text] [doi: 10.1186/1753-4631-1-5] [Medline: 17908344]
29. Goldberger AL. Fractal variability versus pathologic periodicity: complexity loss and stereotypy in disease. Perspect Biol Med 1997;40(4):543-561. [doi: 10.1353/pbm.1997.0063] [Medline: 9269744]
30. Vargas B, Cuesta-Frau D, Ruiz-Esteban R, Cirugeda E, Varela M. What can biosignal entropy tell us about health and disease? applications in some clinical fields. Nonlinear Dynamics Psychol Life Sci 2015 Oct;19(4):419-436. [Medline: 26375934]
31. Goldberger AL, Peng CK, Lipsitz LA. What is physiologic complexity and how does it change with aging and disease? Neurobiol Aging 2002;23(1):23-26. [doi: 10.1016/s0197-4580(01)00266-4] [Medline: 11755014]
32. Pincus SM, Goldberger AL. Physiological time-series analysis: what does regularity quantify? Am J Physiol 1994 Apr;266(4 Pt 2):H1643-H1656. [doi: 10.1152/ajpheart.1994.266.4.H1643] [Medline: 8184944]
33. Pincus S, Singer BH. Randomness and degrees of irregularity. Proc Natl Acad Sci U S A 1996 Mar 05;93(5):2083-2088 [FREE Full text] [doi: 10.1073/pnas.93.5.2083] [Medline: 11607637]
34. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000 Jun;278(6):H2039-H2049 [FREE Full text] [doi: 10.1152/ajpheart.2000.278.6.H2039] [Medline: 10843903]
35. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Phys Rev Lett 2002 Aug 05;89(6):068102 [FREE Full text] [doi: 10.1103/PhysRevLett.89.068102] [Medline: 12190613]
36. Costa M, Cygankiewicz I, Zareba W, Lobodzinski S. Multiscale complexity analysis of heart rate dynamics in heart failure: preliminary findings from the music study. 2006 Presented at: 2006 Computers in Cardiology; September 17-20, 2006; Valencia, Spain p. 101-103 URL: https://ieeexplore.ieee.org/document/4511798
37. Glass L, Mackey M. From Clocks to Chaos: the Rhythms of Life. Princeton, NJ: Princeton University Press; 1988.
38. Goldberger A, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000 Jun 13;101(23):E215-E220. [doi: 10.1161/01.cir.101.23.e215] [Medline: 10851218]
39. Pincus SM, Gladstone IM, Ehrenkranz RA. A regularity statistic for medical data analysis. J Clin Monit 1991 Oct;7(4):335-345. [doi: 10.1007/BF01619355] [Medline: 1744678]
40. Pincus SM. Quantitative assessment strategies and issues for mood and other psychiatric serial study data. Bipolar Disord 2003 Aug;5(4):287-294. [doi: 10.1034/j.1399-5618.2003.00036.x] [Medline: 12895206]
41. Pincus SM, Cummins TR, Haddad GG. Heart rate control in normal and aborted-SIDS infants. Am J Physiol 1993 Mar;264(3 Pt 2):R638-R646. [doi: 10.1152/ajpregu.1993.264.3.R638] [Medline: 8457020]
42. Peng CK, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 1995;5(1):82-87. [doi: 10.1063/1.166141] [Medline: 11538314]
43. Kemp AH, Quintana DS, Malhi GS. Effects of serotonin reuptake inhibitors on heart rate variability: methodological issues, medical comorbidity, and clinical relevance. Biol Psychiatry 2011 Apr 15;69(8):e25-6; author reply e27. [doi: 10.1016/j.biopsych.2010.10.035] [Medline: 21353666]
44. Yeragani VK, Rao KARK, Smitha MR, Pohl RB, Balon R, Srinivasan K. Diminished chaos of heart rate time series in patients with major depression. Biol Psychiatry 2002 May 01;51(9):733-744. [doi: 10.1016/s0006-3223(01)01347-6] [Medline: 11983187]

45. Migliorini M, Mendez MO, Bianchi AM. Study of heart rate variability in bipolar disorder: linear and non-linear parameters during sleep. Front Neuroeng 2012 Jan 10;4:22 [FREE Full text] [doi: 10.3389/fneng.2011.00022] [Medline: 22291638]
46. Schumann A, Andrack C, Bär KJ. Differences of sympathetic and parasympathetic modulation in major depression. Prog Neuropsychopharmacol Biol Psychiatry 2017 Oct 03;79(Pt B):324-331. [doi: 10.1016/j.pnpbp.2017.07.009] [Medline: 28710030]
47. Wagner CD, Persson PB. Chaos in the cardiovascular system: an update. Cardiovasc Res 1998 Nov;40(2):257-264. [doi: 10.1016/s0008-6363(98)00251-x] [Medline: 9893718]
48. Sgoifo A, Carnevali L, Alfonso ML, Amore M. Autonomic dysfunction and heart rate variability in depression. Stress 2015 May 25;18(3):343-352. [doi: 10.3109/10253890.2015.1045868] [Medline: 26004818]
49. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
50. Rohatgi A. WebPlotDigitalizer: HTML5 based online tool to extract numerical data from plot images. Version 4.1. arohatgi.info. URL: http://arohatgi.info/WebPlotDigitizer/app/2012 [accessed 2023-01-05]
51. esci in jamovi. The New Statistics. URL: https://thenewstatistics.com/itns/esci/jesci/ [accessed 2023-01-05]
52. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ, England: Lawrence Erlbaum Associates; 1988.
53. McGrath RE, Meyer GJ. When effect sizes disagree: the case of r and d. Psychol Methods 2006 Dec;11(4):386-401. [doi: 10.1037/1082-989X.11.4.386] [Medline: 17154753]
54. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. San Diego, CA: Academic Press; 1985.
55. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
56. Cumming G. Understanding the New Statistics: Effect sizes, Confidence Intervals, and Meta-Analysis. New York, NY: Routledge; 2012.
57. Leistedt SJ, Linkowski P, Lanquart J, Mietus JE, Davis RB, Goldberger AL, et al. Decreased neuroautonomic complexity in men during an acute major depressive episode: analysis of heart rate dynamics. Transl Psychiatry 2011 Jul 26;1(7):e27 [FREE Full text] [doi: 10.1038/tp.2011.23] [Medline: 22832529]
58. Kop WJ, Stein PK, Tracy RP, Barzilay JI, Schulz R, Gottdiener JS. Autonomic nervous system dysfunction and inflammation contribute to the increased cardiovascular mortality risk associated with depression. Psychosom Med 2010 Sep;72(7):626-635 [FREE Full text] [doi: 10.1097/PSY.0b013e3181eadd2b] [Medline: 20639389]
59. Jelinek HF, Khandoker AH, Quintana DS, Imam MH, Kemp AH. Complex Correlation Measure as a sensitive indicator of risk for sudden cardiac death in patients with depression. 2011 Presented at: 2011 Computing in Cardiology; September 18-21, 2011; Hangzhou, China p. 809-812.
60. Byun S, Kim AY, Jang EH, Kim S, Choi KW, Yu HY, et al. Entropy analysis of heart rate variability and its application to recognize major depressive disorder: a pilot study. Technol Health Care 2019 Jun 18;27(S1):407-424 [FREE Full text] [doi: 10.3233/THC-199037] [Medline: 31045557]
61. Yeragani VK, Pesce V, Jayaraman A, Roose S. Major depression with ischemic heart disease: effects of paroxetine and nortriptyline on long-term heart rate variability measures. Biol Psychiatry 2002 Sep 01;52(5):418-429. [doi: 10.1016/s0006-3223(02)01394-x] [Medline: 12242058]
62. Moser M, Lehofer M, Hoehn-Saric R, McLeod DR, Hildebrandt G, Steinbrenner B, et al. Increased heart rate in depressed subjects in spite of unchanged autonomic balance? J Affect Disord 1998 Mar;48(2-3):115-124 [FREE Full text] [doi: 10.1016/s0165-0327(97)00164-x] [Medline: 9543200]
63. Pincus SM, Schmidt PJ, Palladino-Negro P, Rubinow DR. Differentiation of women with premenstrual dysphoric disorder, recurrent brief depression, and healthy controls by daily mood rating dynamics. J Psychiatr Res 2008 Apr;42(5):337-347. [doi: 10.1016/j.jpsychires.2007.01.001] [Medline: 17336329]
64. Baumert M, Lambert GW, Dawood T, Lambert EA, Esler MD, McGrane M, et al. Short-term heart rate variability and cardiac norepinephrine spillover in patients with depression and panic disorder. Am J Physiol Heart Circ Physiol 2009 Aug;297(2):H674-H679 [FREE Full text] [doi: 10.1152/ajpheart.00236.2009] [Medline: 19502559]
65. Voss A, Boettger MK, Schulz S, Gross K, Bär KJ. Gender-dependent impact of major depression on autonomic cardiovascular modulation. Prog Neuropsychopharmacol Biol Psychiatry 2011 Jun 01;35(4):1131-1138. [doi: 10.1016/j.pnpbp.2011.03.015] [Medline: 21453741]
66. Yang AC, Tsai S, Yang C, Kuo C, Chen T, Hong C. Reduced physiologic complexity is associated with poor sleep in patients with major depression and primary insomnia. J Affect Disord 2011 Jun;131(1-3):179-185. [doi: 10.1016/j.jad.2010.11.030] [Medline: 21195485]
67. Moon E, Lee SH, Kim DH, Hwang B. Comparative study of heart rate variability in patients with schizophrenia, bipolar disorder, post-traumatic stress disorder, or major depressive disorder. Clin Psychopharmacol Neurosci 2013 Dec 28;11(3):137-143 [FREE Full text] [doi: 10.9758/cpn.2013.11.3.137] [Medline: 24465250]
68. Valenza G, Citi L, Garcia RG, Taylor JN, Toschi N, Barbieri R. Complexity variability assessment of nonlinear time-varying cardiovascular control. Sci Rep 2017 Feb 20;7(1):42779 [FREE Full text] [doi: 10.1038/srep42779] [Medline: 28218249]

69. Chen X, Yang R, Kuang D, Zhang L, Lv R, Huang X, et al. Heart rate variability in patients with major depression disorder during a clinical autonomic test. Psychiatry Res 2017 Oct;256:207-211. [doi: 10.1016/j.psychres.2017.06.041] [Medline: 28646783]
70. Byun S, Kim AY, Jang E, Kim S, Choi K, Yu HY, et al. Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol. Comput Biol Med 2019 Sep;112:103381 [FREE Full text] [doi: 10.1016/j.compbiomed.2019.103381] [Medline: 31404718]
71. Boettger S, Hoyer D, Falkenhahn K, Kaatz M, Yeragani VK, Bär KJ. Nonlinear broad band dynamics are less complex in major depression. Bipolar Disord 2008 Mar;10(2):276-284. [doi: 10.1111/j.1399-5618.2007.00503.x] [Medline: 18271907]
72. Khandoker AH, Luthra V, Abouallaban Y, Saha S, Ahmed KI, Mostafa R, et al. Predicting depressed patients with suicidal ideation from ECG recordings. Med Biol Eng Comput 2017 May 18;55(5):793-805. [doi: 10.1007/s11517-016-1557-y] [Medline: 27538398]
73. Greco A, Messerotti Benvenuti S, Gentili C, Palomba D, Scilingo EP, Valenza G. Assessment of linear and nonlinear/complex heartbeat dynamics in subclinical depression (dysphoria). Physiol Meas 2018 Mar 29;39(3):034004. [doi: 10.1088/1361-6579/aaaeac] [Medline: 29595146]
74. Kemp AH, Quintana DS, Felmingham KL, Matthews S, Jelinek HF. Depression, comorbid anxiety disorders, and heart rate variability in physically healthy, unmedicated patients: implications for cardiovascular risk. PLoS One 2012 Feb 15;7(2):e30777 [FREE Full text] [doi: 10.1371/journal.pone.0030777] [Medline: 22355326]
75. Iverson GL, Gaetz MB, Rzempoluck EJ, McLean P, Linden W, Remick R. A new potential marker for abnormal cardiac physiology in depression. J Behav Med 2005 Dec 13;28(6):507-511. [doi: 10.1007/s10865-005-9022-7] [Medline: 16222413]
76. Bob P, Susta M, Gregusova A, Jasova D. Dissociation, cognitive conflict and nonlinear patterns of heart rate dynamics in patients with unipolar depression. Prog Neuropsychopharmacol Biol Psychiatry 2009 Feb 01;33(1):141-145. [doi: 10.1016/j.pnpbp.2008.11.005] [Medline: 19041359]
77. Vigo D, Nicola Siri L, Ladrón de Guevara MS, Martínez-Martínez JA, Fahrer R, Cardinali DP, et al. Relation of depression to heart rate nonlinear dynamics in patients ≥60 years of age with recent unstable angina pectoris or acute myocardial infarction. Am J Cardiol 2004 Mar 15;93(6):756-760 [FREE Full text] [doi: 10.1016/j.amjcard.2003.11.056] [Medline: 15019886]
78. Berger S, Kliem A, Yeragani V, Bär KJ. Cardio-respiratory coupling in untreated patients with major depression. J Affect Disord 2012 Jul;139(2):166-171. [doi: 10.1016/j.jad.2012.01.035] [Medline: 22386048]
79. Schulz S, Koschke M, Bär KJ, Voss A. The altered complexity of cardiovascular regulation in depressed patients. Physiol Meas 2010 Mar;31(3):303-321. [doi: 10.1088/0967-3334/31/3/003] [Medline: 20086275]
80. Byun S, Kim AY, Jang E, Kim S, Choi K, Yu H, et al. Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol. Comput Biol Med 2019 Sep;112:103381 [FREE Full text] [doi: 10.1016/j.compbiomed.2019.103381] [Medline: 31404718]
81. Kuang D, Yang R, Chen X, Lao G, Wu F, Huang X, et al. Depression recognition according to heart rate variability using Bayesian Networks. J Psychiatr Res 2017 Dec;95:282-287. [doi: 10.1016/j.jpsychires.2017.09.012] [Medline: 28926794]
82. Chang HA, Chang CC, Kuo TBJ, Huang SY. Distinguishing bipolar II depression from unipolar major depressive disorder: differences in heart rate variability. World J Biol Psychiatry 2015 Mar 24;16(5):351-360. [doi: 10.3109/15622975.2015.1017606] [Medline: 25800950]
83. Čukić M, Stokić M, Simić S, Pokrajac D. The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification method. Cogn Neurodyn 2020 Aug;14(4):443-455 [FREE Full text] [doi: 10.1007/s11571-020-09581-x] [Medline: 32655709]
84. Čukić M, Stokić M, Radenković S, Ljubisavljević M, Simić S, Savić D. Nonlinear analysis of EEG complexity in episode and remission phase of recurrent depression. Int J Methods Psychiatr Res 2020 Jun 09;29(2):e1816 [FREE Full text] [doi: 10.1002/mpr.1816] [Medline: 31820528]
85. Čukić M, López V, Pavón J. Classification of Depression Through Resting-State Electroencephalogram as a Novel Practice in Psychiatry: Review. J Med Internet Res 2020 Nov 03;22(11):e19548 [FREE Full text] [doi: 10.2196/19548] [Medline: 33141088]
86. Čukić M, Pokrajac D, Lopez D. On mistakes we made in prior computational psychiatry data driven approach projects and how they jeopardize translation of those findings in cinical practice. 2020 Aug 25 Presented at: IntelliSys 2020: Intelligent Systems and Applications; September 3-4, 2020; London, UK p. 493-510. [doi: 10.1007/978-3-030-55190-2_37]
87. Zeković J, Madžgalj Š, Platiša MM. Detrended fluctuation analysis of heart and respiratory rhythm in atrial fibrillation. 2018 Presented at: 2018 Computing in Cardiology Conference (CinC); September 23-26, 2018; Maastricht, the Netherlands p. 1-4. [doi: 10.22489/cinc.2018.300]
88. Platiša MM, Radovanović NN, Milašinović G, Pavlović SU. Sample entropy approach to the examination of cardio-respiratory coupling in response to cardiac resynchronization therapy. 2021 Presented at: Entropy 2021: The Scientific Tool of the 21st Century; May 5-7, 2021; Basel, Switzerland. [doi: 10.3390/entropy2021-09770]
89. Platisa MM, Gal V. Dependence of heart rate variability on heart period in disease and aging. Physiol Meas 2006 Oct;27(10):989-998. [doi: 10.1088/0967-3334/27/10/005] [Medline: 16951458]

90. Peng CK, Hausdorff JM, Goldberger AL. Fractal mechanisms in neural control: human heartbeat and gait dynamics in health and disease. In: Walleczek J, editor. Self-Organized Biological Dynamics and Nonlinear Control. Cambridge, MA: Cambridge University Press; Aug 14, 1999.
91. Kemp AH, Quintana DS, Malhi GS. Effects of serotonin reuptake inhibitors on heart rate variability: methodological issues, medical comorbidity, and clinical relevance. Biol Psychiatry 2011 Apr 15;69(8):e25-6; author reply e27. [doi: 10.1016/j.biopsych.2010.10.035] [Medline: 21353666]
92. Kemp A. Depression, antidepressant treatment and the cardiovascular system. Acta Neuropsychiatr 2014 Jun 24;23(2):82-83. [doi: 10.1111/j.1601-5215.2011.00535.x]
93. Tsuji H, Larson MG, Venditti FJ, Manders ES, Evans JC, Feldman CL, et al. Impact of reduced heart rate variability on risk for cardiac events: the Framingham Heart study. Circulation 1996 Dec;94(11):2850-2855. [doi: 10.1161/01.cir.94.11.2850]
94. Porges SW. The Polyvagal Theory: Neurophysiological Foundations of Emotions, Attachment, Communication, Self-Regulation. 1st ed. New York, NY: W.W. Norton & Company; 2011.
95. Goldberger AL, Amaral LAN, Hausdorff JM, Ivanov PC, Peng C, Stanley HE. Fractal dynamics in physiology: alterations with disease and aging. Proc Natl Acad Sci U S A 2002 Feb 19;99 Suppl 1(Suppl 1):2466-2472 [FREE Full text] [doi: 10.1073/pnas.012579499] [Medline: 11875196]
96. Thayer JF, Lane RD. A model of neurovisceral integration in emotion regulation and dysregulation. J Affect Disord 2000 Dec;61(3):201-216. [doi: 10.1016/s0165-0327(00)00338-4] [Medline: 11163422]
97. Saad M, Ray LB, Bujaki B, Parvaresh A, Palamarchuk I, De Koninck J, et al. Using heart rate profiles during sleep as a biomarker of depression. BMC Psychiatry 2019 Jun 07;19(1):168 [FREE Full text] [doi: 10.1186/s12888-019-2152-1] [Medline: 31174510]

Abbreviations

ApEn: approximate entropy
ANS: autonomous nervous system
CVC: cardiac vagal control
CVDs: cardiovascular diseases
DFA: detrended fluctuation analysis
DSM: Diagnostic and Statistical Manual of Mental Disorders
ECG: electrocardiogram
ES: effect size
HRV: heart rate variability
LLE: largest Lyapunov exponents
MDD: major depressive disorder
MSE: multiscale entropy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
SampEn: sample entropy
ShanEn: Shannon entropy

Edited by J Torous; submitted 16.06.22; peer-reviewed by C Massaroni, S Jonas; comments to author 09.11.22; revised version received 28.11.22; accepted 06.12.22; published 17.01.23.

Please cite as:
Čukić M, Savić D, Sidorova J
When Heart Beats Differently in Depression: Review of Nonlinear Heart Rate Variability Measures
JMIR Ment Health 2023;10:e40342
URL: https://mental.jmir.org/2023/1/e40342
doi:10.2196/40342
PMID:36649063

Review

Digital Phenotyping for Differential Diagnosis of Major Depressive Episode: Narrative Review

Eric Ettore[1], MD; Philipp Müller[2], PhD; Jonas Hinze[3], MSc; Matthias Riemenschneider[3], MD, PhD; Michel Benoit[4], MD, PhD; Bruno Giordana[4], MD; Danilo Postin[5], MSc; Rene Hurlemann[5], MD, PhD; Amandine Lecomte[6], MSc; Michel Musiol[6], PhD; Hali Lindsay[2], MSc; Philippe Robert[7], MD, PhD; Alexandra König[8], PhD

[1]Department of Psychiatry and Memory Clinic, University Hospital of Nice, Nice, France

[2]Research Department Cognitive Assistants, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Saarbrücken, Germany

[3]Department of Psychiatry and Psychotherapy, Saarland University Medical Center, Hombourg, Germany

[4]Department of Psychiatry, Hopital Pasteur, University Hospital of Nice, Nice, France

[5]Department of Psychiatry, School of Medicine and Health Sciences, Carl von Ossietzky University of Oldenburg, Bad Zwischenahn, Germany

[6]Research Department Sémagramme Team, Institut national de recherche en informatique et en automatique, Nancy, France

[7]Research Department, Cognition-Behaviour-Technology Lab, University Côte d'Azur, Nice, France

[8]Research Department Stars Team, Institut national de recherche en informatique et en automatique, Sophia Antipolis - Valbonne, France

Corresponding Author:
Eric Ettore, MD
Department of Psychiatry and Memory Clinic
University Hospital of Nice
30 Voie Romaine
Nice, 06000
France
Phone: 33 633306394
Email: ettore.e@chu-nice.fr

Related Article:

This is a corrected version. See correction statement: https://mental.jmir.org/2023/1/e46831

Abstract

Background: Major depressive episode (MDE) is a common clinical syndrome. It can be found in different pathologies such as major depressive disorder (MDD), bipolar disorder (BD), posttraumatic stress disorder (PTSD), or even occur in the context of psychological trauma. However, only 1 syndrome is described in international classifications (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [DSM-5]/International Classification of Diseases 11th Revision [ICD-11]), which do not take into account the underlying pathology at the origin of the MDE. Clinical interviews are currently the best source of information to obtain the etiological diagnosis of MDE. Nevertheless, it does not allow an early diagnosis and there are no objective measures of extracted clinical information. To remedy this, the use of digital tools and their correlation with clinical symptomatology could be useful.

Objective: We aimed to review the current application of digital tools for MDE diagnosis while highlighting shortcomings for further research. In addition, our work was focused on digital devices easy to use during clinical interview and mental health issues where depression is common.

Methods: We conducted a narrative review of the use of digital tools during clinical interviews for MDE by searching papers published in PubMed/MEDLINE, Web of Science, and Google Scholar databases since February 2010. The search was conducted from June to September 2021. Potentially relevant papers were then compared against a checklist for relevance and reviewed independently for inclusion, with focus on 4 allocated topics of (1) automated voice analysis, behavior analysis by (2) video and physiological measures, (3) heart rate variability (HRV), and (4) electrodermal activity (EDA). For this purpose, we were interested in 4 frequently found clinical conditions in which MDE can occur: (1) MDD, (2) BD, (3) PTSD, and (4) psychological trauma.

Results: A total of 74 relevant papers on the subject were qualitatively analyzed and the information was synthesized. Thus, a digital phenotype of MDE seems to emerge consisting of modifications in speech features (namely, temporal, prosodic, spectral,

source, and formants) and in speech content, modifications in nonverbal behavior (head, hand, body and eyes movement, facial expressivity, and gaze), and a decrease in physiological measurements (HRV and EDA). We not only found similarities but also differences when MDE occurs in MDD, BD, PTSD, or psychological trauma. However, comparative studies were rare in BD or PTSD conditions, which does not allow us to identify clear and distinct digital phenotypes.

Conclusions: Our search identified markers from several modalities that hold promise for helping with a more objective diagnosis of MDE. To validate their potential, further longitudinal and prospective studies are needed.

(JMIR Ment Health 2023;10:e37225) doi:10.2196/37225

KEYWORDS

depression; bipolar disorder; posttraumatic stress disorder; differential diagnosis; digital phenotyping; speech analysis; nonverbal behavior; physiological measures; posttraumatic stress disorder; mental health; clinical interview; diagnosis; mental disorder; interview; digital health; psychotrauma; digital; information

Introduction

Major depression is a frequent syndrome affecting more than 264 million people worldwide [1]. Major depressive episode (MDE), as defined by the current criteria [2], describes a large heterogeneous clinical syndrome comprising more than 1490 combinations of symptoms [3]. The possibility of fulfilling MDE criteria with opposite symptoms, such as insomnia and hypersomnia, decrease/increase in appetite or agitation, and psychomotor retardation, highlights this heterogeneity. This heterogeneity could explain why studies with large samples of patients with major depression show that only 30% of them remitted with a first-line antidepressant and another 30% did not remit after 4 consecutive treatment trials [4].

These various “profiles” of MDE may require different therapeutic approaches. In fact, according to international classifications [2], MDE symptoms are the same whether in major depressive disorder (MDD) or in bipolar disorder (BD). In clinical practice, hypomanic or manic episodes have been the most recognizable characteristics of BD, while depression seems to be most frequent clinical manifestation [5,6]. Performing differential diagnosis between MDD and BD during an MDE is challenging. Approximately 20% of people with MDE would be misdiagnosed as MDD when the correct diagnosis is BD [7]. Thereby, the delay for a correct BD diagnosis can vary from almost 7 to 10 years after the first mood symptoms [8]. These misdiagnoses can have several damaging consequences, for example, induction of manic, hypomanic, or mixed states; development of treatment resistance; or cycle acceleration [9].

Furthermore, Rytwinski et al [10] showed in their meta-analysis that almost 52% of patients with posttraumatic stress disorder (PTSD) have comorbidity with MDD. In fact, in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), several symptoms are shared between MDD and PTSD [11] and a study by Gros et al [12] showed that in patients classified by DSM-5 as having MDD and PTSD, only the presence of trauma (criterion A of DSM-5) could differentiate MDD from PTSD efficiently. Moreover, patients classified as comorbid between MDD and PTSD showed the highest severity of symptoms, more cognitive deterioration, higher suicidality, and worse prognosis compared with noncomorbid patients [12]. One possible explanation could be the time-dependent fluctuation of PTSD symptoms [13]. In a dynamic model, MDD could be a progression or an evolution of psychotraumatism [14]. Although there is a frequent association between MDD and PTSD, the link between them remains unclear. However, it points to the importance of exploring the presence of psychological trauma and PTSD in patients with MDE because it carries the potential to drastically impact further therapeutic care. Misdiagnosing PTSD or the presence of psychological trauma could lead to a worse prognosis due to inappropriate drug treatment or psychotherapy.

In this article, we focus on the use of new digital markers in MDE, given their potential to serve as an additional and objective diagnostic support. For this, we explore studies in MDE, especially when it is presented in the context of either MDD or BD, as well as when it is presented alongside the presence of psychological trauma or PTSD.

Digital phenotyping refers to the moment-to-moment quantification of human behavior in everyday life using data from digital devices [15]. It suggests the possibility of revealing clinically relevant information by a continuous and nonintrusive monitoring of behavioral and mental states. Two categories of data can be captured by digital phenotyping. First, active data, which require the input of the participants being studied. It includes, for example, recordings, responses to surveys, and social media activity. Passive data, however, do not require the individual’s participation to be captured. They include, for example, accelerometer-derived data or GPS coordinates [16]. Digital markers correspond to disease indicators obtained digitally, which can be used to define a digital phenotype [17]. It refers to the possibility of capturing, through computerized measurement tools, certain symptoms or behavior specific to a psychiatric disorder.

There is a specific interest in psychiatry in which symptoms and clinical states are mainly measured using question-based scales and without biological markers. Thus, the interest of digital phenotyping would be to obtain objective and quantifiable measurements of these symptoms or behaviors [16]. Moreover, diagnostic categories do not capture the heterogeneity of symptoms, and variability between patients can lead to misdiagnosis and incorrect treatments [18]. Identifying objective markers of clinical states, including trans-diagnostic symptoms, could improve disease classification and treatment [19].

Thus, digital phenotyping should play a role in routine clinical practice, especially by improving clinical diagnosis and treatment by an early detection of condition onset, by assessing treatment response, or even by detecting relapse [20]. Therefore, we assume that new digital measures may provide indicators for the heterogeneous characteristics of MDE and could help to better distinguish between its potential different clinical profiles. Based on this rationale, we performed a narrative review including studies on several technologies, such as speech and video analysis and physiological measures, namely, heart rate variability (HRV) and electrodermal activity (EDA) for the assessment of MDE.

Methods

Search Strategy

We conducted this narrative review from June to September 2021, which mainly concerned the current application of digital tools for MDE diagnosis. The following electronic databases were searched: PubMed/MEDLINE, Web of Science, and Google Scholar. The review was limited to articles in English or French and because we aimed to establish an overview of the most recent advances in these domains, we restricted our search to studies published after January 2010. For this paper, we decided to concentrate on digital tools that are easy to use during a clinical interview, minimally invasive, and least dependent on compliance. Thus, we focused on 4 types of digital markers: (1) automated voice analysis, behavior analysis by (2) video and physiological measures, (3) HRV, and (4) EDA. We then used broad search terms to capture as many studies as possible that are specifically related to these technologies and we associated them with terms related to specific psychiatric issue. We concentrated on MDE and on several psychiatric conditions where it may occur, especially on (1) MDD, (2) BD, (3) PTSD, and (4) psychological trauma. The search strings are included in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: primary articles or reviews dealing with depression, BD, bipolar depression, PTSD, or psychological trauma using digital tools for speech or video analysis or analysis of physiological parameters such as HRV or skin conductance. The exclusion criteria were as follows: articles related to activity monitoring on social networks or by SMS text messages or phone and articles related to ecological momentary assessment. There is an abundance of literature on these topics, but we wanted to focus here on digital tools that can be used specifically during a social interaction such as an interview with the clinician. Unrelated and redundant articles or studies using technologies for therapeutic purposes were also excluded. Theoretical papers, study protocols, letters, books or book chapters, statistical reviews, and dissertations were also excluded.

Study Selection

Three authors (EE, AK, and PR) independently screened relevant titles and abstracts for this narrative review. Then, all 3 authors screened relevant papers for eligibility. Finally, according to inclusion and exclusion criteria, full texts of eligible articles were obtained (EE).

Results

Overview

We selected 74 articles dealing with the use of the 4 aforesaid digital tools for diagnosis assessment under different psychiatric conditions. A total of 39 articles were selected for MDD, 18 for BD, and 17 for PTSD and psychological trauma. Results are summarized on 3 tables. See Table 1 for MDD, and Multimedia Appendix 2 for BD, PTSD and psychological trauma.

Table 1. Summary of speech analysis, nonverbal behavioral analysis, heart rate variability (HRV), and electrodermal activity studies (EDA) in patients with major depressive disorder (MDD).

Study	Participants	Recording setting	Principal findings	Principal features
Speech analysis studies				
Alghowinem et al [21]	• 30 with MDD and 30 in HC[a]	• Reading tasks with negative and positive meaning • Free speech by answering questions, where the patients describe events that had aroused significant emotions	• MFCC[b], jitter, shimmer, energy, and loudness features were robust in getting the general characteristic of depressive speech	• Prosodic, spectral, and source features
Arevian et al [22]	• 15 with MDD, 14 with BD[c], 14 with schizophrenia, and 14 with affective schizophrenia	• Free speech (answering questions from an app with phone call)	• Correlation with providers' global assessment: • ↓ Negative emotional language • ↑ Positive emotional language • ↑ Complex word use	• Type of words used
Hönig et al [23]	• 219 participants with BDI (BD type I) assessment	• Read speech, spontaneous speech, telling an imagined story	• Depression severity: • ↓ Average and SD MFCC 2 • ↓ SD MFCC 3 • ↓ Average pitch • ↓ Shimmer • ↓ Spectral harmonicity • ↓ Speech rate (↑ Average syllables duration)	• Prosodic, spectral, source, and filter features
Horwitz et al [24]	• 35 participants with Hamilton Rating Scale for Depression and Quick Inventory of Depressive Symptomatology assessment	• Free speech, reading, and distinct vowels	• Depression severity: • ↑ Jitter and shimmer • ↓ Harmonic-to-noise ratio • ↓ First and third formants • ↓ Speech rate • ↓ (Read) and (free speech) pitch-based features	• Prosodic, spectral, source, and filter features
Mundt et al [25]	• 105 with MDD	• Free and with tasks: reciting the alphabet, counting, reading, and sustained vowels	• Correlation with depression severity: • ↑ Total speech time • ↑ Total pause time • ↑ Variable pause length • ↑ Percentage pause time • ↓ Speech pause ratio • ↓ Speaking rate	• Temporal and prosodic features
Quatieri et al [26]	• 35 with MDD	• Free speech for prosodic measurements and distinct vowels for shimmer, jitter, and aspiration measurements	• Depression severity and psychomotor retardation: • ↑ Shimmer • ↑ Aspiration • ↓ Harmonic-to-noise ratio • Depression severity • ↑ Jitter • ↓ Pitch variance • ↓ Average velocity	• Prosodic, source, and filter features
Shinohara et al [27]	• 30 with MDD and 14 in HC	• Free speech and reading phrases aloud • Detection of emotional elements (ie, "anger" and "joy") • 2 indices extracted: "vitality" and "mental activity"	• MDD correlated with: • ↓ Vitality scores (effect size 1.03 and AUC[d] 0.76, sensitivity 0.93, and specificity 0.55)	• Temporal and "emotional" prosodic features

Study	Participants	Recording setting	Principal findings	Principal features
Taguchi et al [28]	• 36 with MDD and 36 in HC	• 3 tasks: reading task, verbal fluency task, and reading task again	• Significant discrimination between MDD and HC with MFCC 2 (sensitivity 77.8% and specificity 86.1%)	• Spectral features
Wang et al [29]	• 47 with MDD and 57 in HC	• 4 tasks: video watching, question answering (natural speech and interview), text reading, and picture describing • Each task involved 3 emotional materials: positive, negative, and neutral	• Patients with MDD (no matter which emotion or task was involved, with a large effect size): • ↓ Loudness • ↓ MFCC 5 • ↓ MFCC 7 • ↓ Fundamental frequency (F0) and MFCC 3 (in some scenarios with moderate effect size)	• Prosodic and spectral features
Xu et al [30]	• 45 with MDD, 43 with schizophrenia, and 41 in HC	• Semistructured face-to-face interviews	• Patients with MDD: • ↑ % of words related to the past and sadness • ↑ Conversational interruption • ↑ Response time	• Temporal and prosodic features, as well as the type of words used
Yamamoto et al [31]	• 97 with MDD, 68 with BD, and 76 in HC	• 10 minutes of free speech	• Depression severity: • ↓ Speech rate • ↑ Pause time • ↑ Response time	• Temporal features
Nonverbal behavioral analysis studies				

Study	Participants	Recording setting	Principal findings	Principal features
Alghowinem et al [32]	• Database with depression status (Quick Inventory of Depressive Symptomatology, 16 Items)	• Free interview	• Eye activity • Head pose	• Average recalls (mean of sensitivity and specificity) was 70% for detecting depression
Bhatia et al [33]	• 47 with MDD	• No task • Symptom severity was evaluated at 1, 7, 13, and 21 weeks	• Head movement	• Head movement synchrony did not change over the course of treatment with a change in depression severity
Dibeklioglu et al [34]	• 49 with MDD	• No task • Symptom severity was evaluated at 1, 7, 13, and 21 weeks	• Facial movement • Head movement • Vocal prosody	• For depression recognition: AUC was 67.25% for the fusion of head movement dynamics and vocal prosody • AUC was 73.16% for facial movement dynamics and vocal prosody • AUC was 77.77% for a combination of facial and head movement dynamics • AUC was 78.67% by fusion of all modalities
Fiquer et al [35]	• 40 with MDD	• Evaluation at baseline (T0) and after a 2 weeks' tDCS[e] treatment (T1) • Describing their current mood during a 15-minute interview • Face and trunk recording	• Ethogram (21 questions), 20 nonverbal categories: 10 indicative of high-energy and favorable disposition to social interaction (eye contact, illustrative gestures, symmetric smile, raised eyebrows, yes/no nodding, head up, head to side, verbal backchannel, body posture toward the interlocutor); 10 indicative of low energy, negative feelings, or social disinterest (folded arms, head down, shrug, asymmetric smile, adaptive gestures, crying, frown, tight lips, lips down, silence); 1 verbal category: speaking	• Clinical improvement: • ↓ Head down, lips down, frown, and crying • ↑ Yes nodding and eye contact • Facial, head, and hand expressive movements were associated with the severity of depression
Fiquer et al [36]	• 78 patients with MDD: 50 from the Netherlands and 28 from Brazil	• During the HDRS[f] interview, face and trunk recording of both the patient and interviewer; scoring by a blinded observator	• Speaking effort: patients' behaviors • Encouragement: interviewers' behaviors	• No association between behavioral variables and baseline severity of depression • Patients who did not respond to treatment or did not remit: speaking effort from before to after treatment
Fiquer et al [37]	• 100 with MDD and 83 in HC • 41 with MDD treated by sertraline (25-200 mg) at the first hospital • 59 with MDD; 23 treated by escitalopram (10-20 mg) and 36 treated by tDCS in a second hospital	• Semistructured 15-minute interview with general questions, face and trunk recording • 2 assessments: before treatment (T0) and after 8 weeks of treatment (T1) • Scoring by a blinded observator	• Ethogram (21 questions) from Fiquer et al [35]	

Study	Participants	Recording setting	Principal findings	Principal features
				• Patients with MDD: • ↑ Shrug, head, and lips down; adaptive hand gestures; frown, head, and lips down and cry • ↓ Asymmetric smile, eye contact, and smile • NVB[g] was not associated with depression severity and did not significantly change after depression treatment • Treatment responders at baseline: • ↑ Interpersonal proximity; head down; adaptive hand gestures; frown, cry, and folded arms; head to side; no nodding • ↓ Eye contact
Girard et al [38]	• 18 with MDD	• Recording during the first 3 interview questions (depressed mood, feelings of guilt, and suicidal ideation)	• Facial expressivity defined by the Facial Action Coding System in terms of individual muscle movements called AUs[h]	• With symptom severity: • ↓ AU 12 activity (smile and signal affiliative intent) • ↑ AU 14 activity (contempt, negative affect, and signal nonaffiliative intent) • ↑ Facial expressions associated with contempt
Girard et al [39]	• 33 with MDD	• Recording during the first 3 interview questions (depressed mood, feelings of guilt, and suicidal ideation)	• Facial expressivity defined by the Facial Action Coding System in terms of individual muscle movements called AUs	• With symptom severity • ↓ Affiliative facial expressions (AUs 12 and 15) • ↑ Nonaffiliative facial expressions (AU 14) • ↓ Head motion (ie, amplitude and velocity)
Guo et al [40]	• 52 males and 52 females in the depressed group (Patient Health Questionnaire–9 >5) • 52 males and 52 females in HC (BDI <5)	• Stimuli tasks of 3 emotional valences (watching film clips, replying to 9 free-response questions, reading 3 phonetically balanced passages containing affective content, and describing pictures)	• Facial expressivity • Audio	• For depression recognition: • Watching film clips showed the highest recognition rates (AUC up to 0.798 and 0.807) • Positive emotional stimuli greater than negative emotional stimuli
Jiang et al [41]	• 12 with MDD • Evaluated before and after deep brain stimulation	• Free interview	• Facial expressivity (7 basic emotions)	• AUC 0.75 detecting response to treatment
Kacem et al [42]	• 49 with MDD	• No task • Symptom severity was evaluated at 1, 7, 13, and 21 weeks	• Facial movement • Head movement	• For depression recognition: • Facial movement was greater than head movement (AUC 66.19% vs 61.43%) • With 2 modalities combined, AUC 70.83%
HRV studies				

Study	Participants	Recording setting	Principal findings	Principal features

Adolph et al [43]	• 85 outpatients with suicidal ideation	• ECG[i]:
		  • At rest quietly for 3 minutes: "Resting HRV"
		  • After watching a sad film: "HRV reactivity"
			• HF[j]-HRV reactivity but not HF-HRV at rest was predictive of higher scores on suicidal ideation	• Frequency domain: HF

Fernandes et al [44]	• 50 with MDD	• ECG at rest • Video for positive and negative NVB	• HF and RMSSD[k] were positively correlated with positive NVB • Negative NVB was not associated with HRV	• Frequency domain: HF and LF[l] • Time domain: RMSSD • Ethogram (Fiquer et al [35])
Giurgi-Oncu et al [45]	• 78 with MDD	• 24-hour Holter ECG • Evaluation at 1 and 6 months after therapy (sertraline)	• ↑ HRV at 1 and 6 months after selective serotonin reuptake inhibitors	• Frequency domain: HF • Time domain: SDNN[m] and RMSSD
Hartmann et al [46]	• 62 with MDD and 65 in HC	• 2 time measures: before and after 2 weeks of antidepressant treatment • 15 minutes of resting ECG 1-2 days before treatment	• At baseline compared with HC: ↓ HRV (HF, LF, SD1[n], and RMSSD) • After treatment: HRV normalized in MDD for HF, LF, SD2[o], and SD1/SD2 ratio	• Frequency domain: LF, HF, and LF-to-HF ratio • Time domain: SDNN, RMSSD, pNN50[p] • Nonlinear: SD1, SD2, SD1-to-SD2 ratio
Kircanski et al [47]	• 722 with MDD, 309 with anxious depression (according to the HDRS interview), and 413 without • Outcomes at 8 weeks	• 2-minute seated ECG recording, first with eyes open and second with eyes closed	• In anxious depression, better treatment response if: ↑ HRV in pretreatment • In nonanxious depression, better treatment response if: ↓ HRV in pretreatment	• Time domain: RMSSD • Heart rate

Lee et al [48]	• 34 with MDD
	• Outcome at 12 weeks of antidepressant treatment
		• 5-minute ECG	• Positive correlation at baseline:
			• HDRS items 14 and 15 with LF-to-HF ratio
			• Positive correlation at the endpoint:
			  • HDRS item 5 and LF
			  • HDRS items 7 and 13 (fatigue-related item) and LF
			  • HDRS item 8 and LF, SDNN, and RMSSD
			  • HDRS total, LF, and HF
				• Frequency domain: VLF[q], LF, HF, LF-to-HF ratio
				• Time domain: SDNN, RMSSD

Neyer et al [49]	• 50 with MDD	• ECG at rest for the HRV measure • Measures before and after treatment	• Depressive symptoms improved without any change in HRV	• Frequency domain: HF • Time domain: RMSSD
Sarlon et al [50]	• 89 with MDD		• No association was found between HRV and symptom severity	• Frequency domain: VLF, LF, HF, LF-to-HF ratio • Time domain: SDNN, RMSSD, beat-to-beat interval • Skin conductance • Skin temperature

Study	Participants	Recording setting	Principal findings	Principal features
		• HRV: blood volume pulse finger clip sensor • Skin conductance: Velcro tape with integrated Ag/AgCl electrodes • Skin temperature: Nexus temperature • Beat-to-beat interval: elastic belt with a breathing sensor • 3 conditions: baseline sitting for 5 minutes, with emotion-induced stressors (recall an unpleasant stressful experience), and relaxed state after 300 seconds		
EDA studies				

Study	Participants	Recording setting	Principal findings	Principal features
Kim et al [51]	• 30 with MDD and 37 in HC	• ProComp Infiniti (SA7500, computerized biofeedback system, thought technology) • EDA during 5 experimental phases: baseline, mental arithmetic task, recovery from the stress task, relaxation task, recovery from the relaxation task	• Classifying participants with MDD versus controls: 74% accuracy, 74% sensitivity, 71% specificity • Stress and relaxation tasks were the most relevant	• MSCL[r], SDSCL[s], SKSCL[t], and NSSCR[u]
Kim et al [52]	• 30 with MDD and 31 in HC	• As per Kim et al [51]	• Classifying participants with MDD versus controls: 70% accuracy, 70% sensitivity, 71% specificity	• MSCL, SDSCL, SSCL[v], MSCR[w], NNSCR[x], and poststroke depression
Litwińska-Bołtuć et al [53]	• 97 with MDD • Follow-up at 1 year	• EDOR (Electro Dermal Orienting Reactivity) test (Emotra AB)	• Hyporeactive patients: relapse or recurrence of depression was nearly 5 times higher	• MSCR

Pedrelli et al [54]	• 31 with MDD
	• Follow-up at 8 weeks
		• 2 E4 Empatica wristbands, one on each wrist for 22 hours/day, 7 days/ week
		• Smartphone sensor data: movisensXS phone app (movisens GmbH)
			• Correlations between the models' estimate of HDRS scores and clinician-rated HDRS: from OR[y] 0.46 (CI 0.42-0.74) to OR 0.7 (CI 0.66-0.74)	• Empatica:
				  • EDA: MSCR, PSCR[z], AASCR[aa]
				  • Peripheral skin temperature
				  • Heart rate
				  • 3-axis accelerometer and sleep characteristics
				• Smartphone data: mobile-based social interactions (number of calls, SMS text messages), activity patterns (walking), number of apps used

Smith et al [55]	• 11 with MDD • 16 in HC	• Shimmer3 GSR+ unit with Shimmer Optical Pulse Sensing Probe; at rest with eyes closed for 3 minutes	• 81% accuracy for detecting depression with HRV • No benefit to including skin conductance response to improve accuracy	• HRV • Time domain: SDNN, beat-to-beat interval • Nonlinear: SD1, SD2 • Skin conductance response: MSCR, SDSCR[ab], PSCR, and AASCR

[a]HC: healthy control.

[b]MFCC: mel-frequency cepstral coefficient.

[c]BD: bipolar disorder.

[d]AUC: area under the curve.

[e]tDCS: transcranial direct current stimulation.

[f]HDRS: Hamilton Depression Rating Scale.

[g]NVB: nonverbal behavior.

[h]AU: action unit.

[i]ECG: electrocardiogram.

[j]HF: high frequency

[k]RMSSD: root-mean-square surface distance.

[l]LF: low frequency

[m]SDNN: SD of the NN (R-R) intervals.

[n]SD1: SD of points perpendicular to the major axis of the Poincaré plot.

[o]SD2: SD of points along the major axis of the Poincaré plot.

[p]PNN50: proportion of NN50 divided by the total number of NN (R-R) intervals.

[q]VLF: very-low frequency.

[r]MSCL: mean amplitude of the skin conductance level.

[s]SDSCL: standard deviations of the skin conductance level.
[t]SKSCL: skewness of the of the skin conductance level.
[u]NSSCR: nonspecific skin conductance response.
[v]SSCL: slope of the skin conductance level.
[w]MSCR: mean amplitude of the skin conductance response.
[x]NNSCR: number of nonspecific skin conductance responses.
[y]OR: odds ratio.
[z]PSCR: peak skin conductance response.
[aa]AASCR: average amplitude skin conductance response.
[ab]SDSCR: standard deviations of the skin conductance response.

Major Depressive Disorder

Speech Analysis

Prosodic abnormalities in patients with depression are well known and a monotonous speech or reduced prosody can be easily and frequently heard. It is assumed that cognitive and physiological changes in depression could affect speech production, and changes in the automatic and somatic nervous system cause disturbances in muscle tension and respiratory rate. These changes will not only influence vocal folds and vocal tract dynamics, but also constrain articulatory movement [56].

Speech analysis, which consists of automatically extracting vocal and linguistic features from audio signals, offers the possibility to detect and measure these changes and serves as an additional objective assessment of depression. Traditionally, acoustic features can be divided into 5 types: temporal, prosodic, spectral, source, and filter features. Temporal features, defined as speech prosodic timing measures, appear to be a promising measure of depression and its severity [57].

Yamamoto et al [31] used 3 temporal features in a prospective follow-up study: speech rate, pause time, and response time. They found a correlation with scores on the Hamilton Rating Scale for Depression (HAM-D). Mundt et al [25] found 5 prosodic timing measures to be significantly correlated with depression severity: total speech time, total pause time, percentage pause time, speech pause ratio, and speaking rate. Xu et al [30] found that people with depression have a longer response time and interrupt their communication partner more often. To summarize, people with depression used a much larger percentage of words related to the past and sadness emotions. Similarly, Shinohara et al [27] used a "vitality" score based on emotional elements of speech (ie, *joy*, *anger*) and showed a negative correlation between the *vitality* score and HAM-D scores. Arevian et al [22] analyzed the types of word used and found, among others, more negative emotion and less complex word use in patients with depression.

Finally, Alghowinem et al [21] and Hönig et al [23] found that the average syllable duration could be positively correlated with depression severity.

With prosodic features in people with depression, reduced fundamental frequency (F0) range and average are frequently found [23,24,26,58]. F0 corresponds to the lowest frequency of the speech signal perceived as pitch. It could explain the depressive monotonous speech and could be a consequence of disturbances in laryngeal muscle tension due to psychomotor retardation. The latter may also explain reduction in F0 variability [56,57]. However, some studies also report no significant correlation between F0 variables and depression [25,28].

Several studies on depression showed a shift in spectral energy, and mel-frequency cepstral coefficients (MFCCs) are often used as spectral features. Spectral features characterize the spectrum of speech, which at a given time correspond to the frequency distribution of the speech signal. Taguchi et al [28] showed that the second dimension of MFCC allowed a significant discrimination between patients with depression and controls. Further, Wang et al [29] found that MFCC 5, MFCC 7, and loudness were consistently lower in people with depression. According to the task designed during speech recording, MFCC 3 and F0 were significantly lower than in people with depression than in healthy people.

Finally, source features such as harmonic-to-noise ratio (HNR), jitter, and shimmer also tend to increase with depression severity, supporting the hypothesis of a more breathy phonation in depressed speech [26]. This could be explained by a more open and turbulent glottis, linked to a reduction in laryngeal muscle tension. The study by Hönig et al [23] also supports this hypothesis. Alghowinem et al [21] showed that not only shimmer and jitter but also loudness and MFCC features are part of the general characteristics of depressive speech.

Nonverbal Behavior Analysis

Traditionally, evaluation of clinical depression is based on patients' verbal information from psychological interviews and rating scales or self-report questionnaires. However, verbal analysis has several limitations. First, patients may underreport or overreport depressive symptoms due to different reasons (eg, social stigma). Second, it can be assumed that most human communication is taking place through nonverbal behavior. Indeed, a large part of nonverbal behavior is mostly outside of conscious control, and these cues and signals may differ from verbal reports. Facial expression, gestures, and body postures are mainly involuntary, and therefore represent a privileged way of expressing feelings and emotions [35,59]. Girard et al [38,39] showed that when the severity of depressive symptoms was high, participants made fewer affiliative facial expression such as smiling, and more nonaffiliative facial expression such as contempt. They also showed that patients exhibited diminished head motion (amplitude and velocity). These results support the social withdrawal hypothesis in depression. In fact, rather than affective valence with an increase or decrease in facial expression, it is the social-communicative value with affiliative

expression that seems to be affected in depression. Thus, nonverbal behavior may serve to maintain interpersonal distance while facilitating social withdrawal.

Fiquer et al [35] supported the assumption that depression involves behaviors related to social withdrawal and negative feelings. Crying, asymmetric smiling, and motionless head/eyebrows were associated with higher levels of depression severity, while patients' illustrative gestures, defined as "the hand and arm movements used to support the speech," were associated with a lower level of depression severity.

Moreover, decrease of negative facial and head indicators and increase of eye contact and yes nodding accompanied clinical improvement, whereas illustrative gestures did not [35]. In another study, Fiquer et al [37] compared nonverbal behavior in patients with depression versus controls. They showed that patients with depression displayed higher levels of negative nonverbal behaviors (shrug, head and lips down, adaptative hand gesture, frown, and cry) and lower levels of positive nonverbal behaviors (eye contact and smile). The authors raise the important role of these behaviors in social withdrawal [37]. Nevertheless, the aforementioned nonverbal behaviors was not associated with depression severity before treatment, which suggests its independence.

The authors concluded that certain typically found nonverbal behaviors could represent a predisposition to depression possibly influenced by personality features. These social symptoms expressed through nonverbal behavior may be the last ones to vanish [37].

On the contrary, Fiquer et al [36] studied nonverbal variables such as speaking effort, encouragement, attunement, and change in attunement, and found no association between these behaviors and the severity of depression. Finally, several recent studies achieved classifying depression with high accuracy using video analysis with head and facial movements [34,42], only head movement [33], facial expressivity [41], or with head pose and eye activity [32]. Guo et al [40] found that the highest accuracy of depression recognition was when patients were watching film clips of positive and negative emotional stimuli compared with other tasks.

In a recent paper, Alghowinem et al [60] used feature selection methods and found that some features have a high capability for distinguishing between depression severities. For speech analysis, the strongest features were the temporal features, F0, HNR, formants, and MFCC; and for behavioral analysis, it was the left-right eye movement and gaze direction and the yaw head movement. The authors showed that these features outperformed all other features in depression detection.

The studies mentioned so far demonstrate the usefulness of behavioral analysis in understanding the mechanisms that underlie the onset and course of depression. These analyses are less influenced by conscious control compared with verbal communication [59], which could make them a reliable objective measure.

Physiological Measures

Heart Rate Variability

HRV corresponds to the beat-to-beat variations in the heart rate over a given period. HRV appears to be a good reflection of the heart's ability to modulate its rhythm in response to external and internal stimuli. HRV is regulated by the autonomic nervous system and, as a result, can inform of its functioning. HRV is related to the influence of the parasympathetic nervous system (PNS) and the sympathetic nervous system (SNS) [61]. Three parameters are commonly used: high-frequency HRV (HF-HRV; 0.15-0.40 Hz), low frequency HRV (LF-HRV; 0.04-0.15 Hz), and the LF-to-HF ratio. HF-HRV is mainly under the control of the PNS, whereas LF-HRV is under the influence of both PNS and SNS, but mostly the SNS. The LF-to-HF ratio is considered an index of sympathovagal balance [61].

In addition, time-domain parameters such as the root-mean-square of successive differences (RMSSD) and the SD of normal-to-normal intervals (SDNN) are frequently used [62]. Normal-to-normal is the interval between 2 heartbeats. HRV can be an indicator of the ability to regulate stress and emotions, which is particularly relevant in psychiatric disorders [63]. Furthermore, a good cardiovascular adaptability can be reflected by a high HRV [64], while LF can be a risk factor of cardiovascular diseases [65]. A meta-analysis in depression [66] showed significant reductions in frequency domain parameters such as HF, LF, and very-low frequency (VLF) HRV as well as a significantly higher LF-to-HF ratio compared with healthy controls. Significant reductions in time domain parameters were further found with reduced RMSSD, SDNN, and interbeat intervals.

Some recent studies have focused on different clinical subdimensions of MDD. Lee et al [48] showed a positive correlation between fatigue-related items of the Hamilton Depression Rating Scale (HDRS) and LF-HRV. The latter was positively correlated with the items "midnight insomnia" and "slower thinking." In the same line, Adolph et al [43] found that HF-HRV in reaction to watching a sad film was predictive of higher scores on suicidal ideation, whereas HF-HRV at rest was not predictive of suicidal ideation. Further studies investigated the relationship between HRV and treatment response. First, Hartmann et al [46] found that patients with MDD normalized their HRV after treatment, especially for HF-HRV and LF-HRV. Further, Giurgi-Oncu et al [45] showed a normalization of HRV parameters (including HF, RMSSD, and SDNN) with the depression treatment by therapy and medication. Kircanski et al [47] predicted treatment success among patients with depression with or without anxiety. The authors found that a higher HF-HRV in pretreatment had better outcomes than lower HF-HRV in anxious depression, whereas lower HF-HRV had better outcomes than higher HF-HRV in nonanxious depression.

By contrast, Sarlon et al [50] found no association between HRV and depressive symptoms severity. Neyer et al [49] showed in a prospective study with HRV measures before and after treatment that depressive symptoms improved without change in HRV, suggesting a more complex relationship.

Finally, Fernandes et al [44] explored the association between HRV and nonverbal behavior using a validated ethogram. Results showed a positive correlation among HRV (HF and RMSSD) and positive nonverbal behaviors, but no association between HRV and negative affect. The authors suggested that HRV and nonverbal behaviors could be regulated by vagal activity, which may be related to the social engagement system.

Electrodermal Activity

The most basic indicators of the state of the autonomic nervous system are heart rate and EDA. EDA can measure electrical conductance of the skin, which depends on the quantity of sweat secreted by glands in the hypodermis and reflects sympathetic nervous activity.

EDA has a tonic and a phasic component. The former is measured in the skin conductance level (SCL) and the latter with the skin conductance response (SCR) [67]. EDA can be used as an indicator for emotional reactivity [68]. In fact, rather than the valence of emotion, it is the intensity that seems related to EDA [69] and negative stimulation induces more extensive reactions than positive [70]. Concerning emotional reactivity, we distinguish labile and stable patients for EDA. EDA labile patients could be described as calm and deliberative, whereas EDA stable patients tend to be more irritable, emotionally expressive, and active. EDA labile patients would be dependent on anxiety traits and better control for potentially threatening stimuli [71].

Several studies show that EDA could be an indication of nonconscious emotional processes [72-74]. A recent systematic review [67] showed lower EDA, especially lower SCL and SCR, in patients with depression compared with healthy controls. Others studies report that, compared with patients with depression also experiencing agitation, those with psychomotor retardation or symptoms of inhibition have lower EDA levels [75-77]. Moreover, patients with “psychotic ” and “endogenous” depression could have lower EDA than “nonendogenous” depression [77-79].

Other recent studies showed moderate to high accuracy in classifying patients with MDD and healthy controls [51,52], such as that by Smith et al [55], but the latter used HRV and SCR for depression detection and found no benefit in including SCR to improve accuracy. In addition, Pedrelli et al [54] showed that EDA and HRV features from wearables as well as activity level and sleep parameters could provide an estimate of changes in severity of depressive symptoms.

In a cohort of patients with MDD at 1-year follow-up, Litwińska-Bołtuć et al [53] concluded that SCR hyporeactivity was associated with recurrent episodes and relapse, and recurrence of depression was almost 5 times higher than that in the reactive patients.

Concluding Remarks

Based on these aforementioned studies, the digital phenotype of MDD could be identified as follows: a reduction in certain speech temporal features (ie, speech rate or speech time), an increase in others (ie, pause time or response time) [25,31], and a change in the type of words used (ie, less complex and more emotionally negative) [22,27]. In addition, prosodic features (ie, F0 range and average) [23,24,26,58], variability [56,57], spectral features [28,29], and source features (ie, HNR, jitter, shimmer [21,26]) are changed.

The behavior of patients with depression could be characterized by a more negative nonverbal behavior (ie, head motion, facial expressivity, and hand or body gestures) [35,38,39]. But more than the valence, it seems that it is the affiliative behaviors that are impacted, supporting a social withdrawal [38,39]. In fact, certain patterns of behavior when interacting with others may predict the severity of symptoms or treatment response [36,59,80-83]. Finally, physiological measures are mainly lowered for both HRV (ie, HF, LF, VLF, and time domains) [66] and EDA (ie, SCR and SCL) [67].

Bipolar Disorder

Speech Analysis

Several studies managed to classify the course of mood episodes or relapses with high confidence. For instance, speech pause duration [84] and number of longer pauses [85] are significantly higher in depressive states than in a hypomanic or euthymic state. Concerning prosodic and source features, Guidi et al [85-87] repeatedly found an increase of F0 in hypomanic states compared with euthymic or depressive states. Similar results were also obtained by Vanello et al [88]. Likewise, filter features seem to increase with manic mood [89], especially F1 and F2 formants, which correspond to the first and the second peak in the spectrum that results from a resonance of the vocal tract.

Finally, many studies extracted several speech features (ie, the openSMILE toolkit), achieving good classification accuracies [90,91]. Classification of depressive states could be obtained with an area under the curve (AUC) value of 0.78 and manic and mixed states with an AUC value of 0.89 [92].

Physiological Measures

Heart Rate Variability

A meta-analysis [93] found reduced HRV compared with healthy controls. In particular, LF-HRV was reduced but no differences in either HF or LF-to-HF ratio were found. However, these analyses did not consider different mood phases. Recent studies were interested in the exploration of HRV during a specific phase of BD. First, Wazen et al [94] showed that several HRV parameters (HF-HRV, time-domain, and non-linear domain measures) could increase when patients go from the mania phase to the euthymic phase. Conversely, Faurholt-Jepsen et al [93] reported an increased HRV in manic states compared with depressive and euthymic states, but no difference between the depressive and euthymic states. Hage et al [95] found that LF-HRV and heart period were significantly lower in patients with BD compared with healthy controls. However, after 8 weeks of treatment, there were no significant changes in HRV parameters.

The link between disease severity and HRV has also been explored. Benjamin et al [96] found a significant association between HF-HRV and disease severity. In the same way, Ortiz et al [97] found that longer illness duration, higher number of depressive episodes, longer duration of most severe

manic/hypomanic episode, comorbid anxiety disorders, and a family history of suicide were associated with a reduced HRV. Moreover, the severity during a depressive episode was associated with lower HRV. Finally, Freyberg et al [98] compared 20 newly diagnosed BD cases, 20 unaffected first-degree relatives, and 20 healthy controls, and demonstrated that HRV did not differ in any measures between the 3 groups.

Lastly, 2 studies have tried to distinguish bipolar depression and MDD, in terms of HRV. First, Chang et al [99] studied a population of patients with depression and BD type II and found that, compared with patients with MDD, they had lower total HRV, with lower HF and LF-HRV but higher LF-to-HF ratio. The authors concluded that HRV may aid in the differential diagnosis of bipolar depression type II and unipolar depression. Second, Hage et al [100] supported these findings by comparing patients with MDD with those with BD depression including BD type I, type II, and "not otherwise specified." Indeed, patients with MDD had significantly higher baseline sinus arrythmia (ie, HF-HRV) and LF-HRV in comparison to patients with BD.

Electrodermal Activity

Greco et al [101] showed that EDA levels could differentiate the phases of BD (depressive, euthymic, and mixed states). Indeed, electrodermal hypoactivity could be a rather stable trait of patients with depression and may reflect euthymia or remission when EDA increases [101]. Lemaire et al [102] measured the intensity and duration of physiological responses to affective stimuli. For this purpose, they used EDA and the heart interbeat intervals. The authors concluded that affective dysregulation, a key dimension of BD, could be measured as an increased affective reactivity to neutral pictures and reduced maintenance of subjective affective responses to all type of pictures, irrespective of the clinical state.

Concluding Remarks

For the digital phenotype of BD, we found that, in the depressed phase, speech changed in temporal (ie, decreased speech pause) and prosodic features (ie, decreased F0), but both increased when patients go into the hypomanic phase [84-88].

Independent of the phase, HRV appears reduced (ie, LF) [93], but the change in HRV between phases remains unclear [94,95,103]. For EDA, an electrodermal hypoactivity in a depressive state is reported, which increases when patients move toward euthymia [101].

PTSD and Psychological Trauma

Speech Analysis

The link between trauma and voice has been likewise explored. Monti et al [104] reported a significant relationship between voice fundamental frequency and the "total trauma" reported with the Childhood Trauma Questionnaire. Interestingly, after trauma recall, vocal jitter and voice irregularity (roughness) were strongly related to the existence of childhood trauma [104]. This suggests that variations in vibratory patterns of the vocal folds could have a relationship with childhood trauma recall. In another study, Monti et al [105] also found an association of anxious attachment and childhood neglect with intensity/loudness measures in singer's voice. Anxious attachment was also positively correlated with jitter and irregularity. Moreover, in speech content, trauma narrative shows an increased use of first-person singular pronouns [106,107] and changes in narrative coherence [108,109].

As in depression, voice markers including F0, jitter, shimmer, and HNR are also found in PTSD [110,111]. Speech rate in both depression and PTSD was shown to be negatively correlated with severity of symptoms. Reduced tonality in vowel space is found in self-reported symptoms of both PTSD and depression [112]. However, in PTSD without MDD, voice markers showed slower and more monotonous speech and less change in tonality [113]. Finally, a recent study showed that audio intensity and reduced pitches per frame could predict PTSD, whereas reduced frequency of positive words seemed to rather predict depression [106].

Nonverbal Behavior Analysis

A core symptom of PTSD is an exaggerated startle response [2]. In their study on patients with PTSD, McTeague et al [114] focused on the eye-blink startle response to narrative texts corresponding to traumatic stories. The authors found a heightened startle reflex response during trauma imagery, which was also evident in reactivity to anger, panic, or physical danger compared with controls. Surprisingly, participants with multiple traumas showed a blunted reactivity compared with participants with a single trauma. Furthermore, participants with multiple traumas had more chronic and severe PTSD with more anxiety and mood comorbidity [114].

In the same paradigm, Blechert et al [115] studied the bodily startle response to electric stimuli in individuals with trauma, with and without PTSD. The results implied a primary response characterized by reflexlike facial and torso movement and a secondary response characterized by an emotional facial response. Besides, participants with PTSD had a stronger initial startle response and a more negatively valenced secondary facial expression compared with controls with trauma. Beyond the startle response to threat, several studies with eye tracking showed that higher levels of PTSD symptoms were associated with an increased attentional bias toward threat [116]. Some authors assumed that negative affect in PTSD could have an important role in maintaining this attention bias [117].

More generally, Katz et al [118] emphasized the importance of trauma disclosure. The authors analyzed 40 video interviews of children who have suffered an abuse, but only half of them disclosed it. Children who did not disclose the abuse showed specific nonverbal cues, such as more physical disengagement, than children who disclosed. Schultebraucks et al [106] performed deep learning audio- and video-based classification of depression versus PTSD for 81 participants, 1 month after their admission to a trauma center following a traumatic exposure. The authors successfully classified between PTSD and MDD status with a high discriminatory accuracy. Likewise, several voice and video markers have been identified as predictive of either PTSD or MDD. In a video analysis, higher fear and anger expressivity were predictive of PTSD, while higher contempt expressivity was important for the classification of MDD. In an audio analysis, an increased use of first-person

singular pronouns, lower audio intensity, and reduced pitch were important features for the PTSD classification, whereas a reduced frequency of positive words was predictive of depression.

Physiological Measures

Heart Rate Variability

Several studies have investigated HRV in PTSD and suggested reduced parasympathetic activity in PTSD. Indeed, 2 meta-analyses found that HRV is lower both at rest and during stress in PTSD compared with controls. Particularly, RMSSD and HF-HRV were reduced, while SDNN and LF-HRV were also reduced, but with a smaller effect. Furthermore, a higher LF-to-HF ratio was found in PTSD [119,120]. Another meta-analysis [121] investigated the relationship between early life maltreatment (ELM) and resting-state HRV. Although no evidence for a relationship was found in comparative and correlational meta-analyses, the meta-regression analysis disclosed interesting results regarding both age and presence of psychopathology. In fact, studies including older-aged samples reported greater reductions in resting-state HRV in association with more severe ELM. Moreover, in clinical samples, compared with nonclinical samples, the authors found stronger reductions in resting-state HRV in ELM-exposed individuals compared with those nonexposed. The authors assumed that ELM could lead to alterations in the development of the autonomic nervous system, especially to an insufficient increase in vagal activity during childhood and adolescence, leading particularly to deficiencies in affective regulation that could increase the risk for psychopathology [121].

Finally, Stone et al [122] compared HF-HRV in women with depression with and without a history of childhood emotional abuse (CEA) and in controls without CEA, and reported that women with depression with CEA exhibited lower HF-HRV than both groups without CEA.

Electrodermal Activity

In patients with PTSD, higher subjective levels of intrusive reexperiencing have been linked to physiological reactivity during trauma recall, while acute dissociative symptoms could show reduced physiological reactivity to trauma memories. Actually, increased physiological reactivity was not associated with the self-report of memory avoidance or dissociation but rather with the degree of memory intrusiveness felt [123,124]. Thus, it can be hypothesized that persistently high physiological reactivity may facilitate chronic arousal dysregulation that disrupts neural systems and could lead to maintain the intrusiveness of remembering [125,126].

Patients with PTSD showed an increased physiological reactivity during trauma recall, while patients with depression had a low physiological response to the memory of an event that they associated with the onset of their depression [126]. Although some researchers suggest a possible attenuation of physiological reactivity in patients with PTSD by depressive symptoms [114], when comparing trauma recall between PTSD with and without depression, no blunting effect was found by depressive mood on physiological reactivity [126]. D'Andrea et al [127] studied individuals exposed to trauma and showed different autonomic reactivities in response to startling sounds according to the type and symptomatology of trauma. Indeed, blunted autonomic reactivity was found in participants with significant symptoms and early exposure to multiple traumas. Conversely, more attenuated trauma exposure had large heart rate acceleration and SCRs.

Concluding Remarks

In PTSD, modifications in speech prosodic and source features are found as in depression [104,110-113], but specific changes are reported in the speech content, especially during trauma narrative or trauma recall, that is, use of first-person singular pronouns [105,106], and changes in narrative coherence [108,109].

In the same way, behavior analysis shows a specific startle response during trauma recall (ie, eye blink or bodily startle) [114,115]. Besides, facial expressivity with higher fear or anger could be predictive of the presence of PTSD [106].

In physiological measures, for HRV, PTSD studies indicated a reduced parasympathetic activity (ie, HF, LF, and time domains) [119,120] as in depression. However, the presence of trauma in people with depression could further decrease HRV (ie, HF) [122]. For EDA, the presence of PTSD, especially with memory intrusiveness felt, leads to an increased EDA reactivity during trauma recall while acute dissociative symptoms could have the opposite effect [123,124,126].

Discussion

Principal Expectations

In this review, we were interested in the state-of-the-art application of new digital measurements to assess symptoms in patients with several psychiatric conditions. In particular, we explored digital markers through speech and behavioral analysis, as well as physiological measures such as HRV and EDA. We focused our research on depression and its different clinical characteristics.

Limitations

Although several databases have been included, relevant papers may have been missed due to the choice of keywords. This paper is not intended to be a systematic review and a broader review could explore other interesting areas. We included a wide range of study designs due to the exploratory nature of our review but the design of the included studies also limited the review. Other reviewers have attempted to reduce the study selection bias but they have acknowledged the possibility of subjectivity in their analyses of the findings.

In addition, we noticed that overall, most studies only investigated the use of 1 digital measure, except for a few that tried to combine, for instance, the measure of EDA with HRV or speech with behavioral analysis.

Apart from a few ethological analyses, most studies analyzed the patient alone, without being in interaction. However, social interaction, especially during clinical interviews, seems to be the most important source of information for clinical assessment.

Moreover, we frequently found that each study used different measures that raise a problem of harmonization. For example, in speech analysis, some studies assessed the type of word used, prosodic or spectral features, or even analyzed a "toolkit" with hundreds of features. Further, certain studies investigated features extracted from free speech, whereas others focused on more constrained speech tasks.

Video analysis can be heterogeneous as well, in that some studies are interested in eyes, head, or torso movements at rest or during an evaluation. In HRV assessment, several studies used electrocardiogram but others used a finger clip sensor. In the same way, EDA can be measured with or without a concurrent task and with different types of technological tools, with variable degrees of precision.

Finally, we did not take into consideration articles dealing with depression with comorbid anxiety. Future research on this topic would be relevant to complete our findings.

Implications for Practice

Assuming that digital markers would help in the differential diagnosis for an MDE, clinicians will be able to make an early diagnosis. Thus, psychotherapy and drug treatment could be introduced early, and diagnostic and therapeutic mistakes could be limited.

Despite the lack of comparative studies, promising digital markers appear to emerge from this review, particularly within physiological measures such as HRV. In fact, patients with bipolar depression had more severe decrease than those with unipolar depression (ie, HF and LF-HRV) [99,100], whereas the presence of comorbid PTSD in MDD seems to induce a greater reactivity of physiological measures, especially when the clinician touched upon psychological trauma [126]. This reactivity with PTSD also seems to be reflected in speech and nonverbal behavioral analysis, which suggests other candidates for digital markers [108,109,114,115].

Conclusion and Future Directions

Despite many limitations, the studies detailed in our review showed promising results on the usefulness of digital phenotyping for the differential diagnosis of MDE. The authors believe that additional research is needed to better understand the potential value of digital phenotyping in clinical practice. Indeed, what should be relevant is a prospective and observational study in which different digital tools will be combined during recordings of social interaction. At least three distinct groups would be required: MDD, bipolar depression, and depression comorbid with PTSD. For each group, it would be interesting to merge speech, behavior, and physiological measures extracted from recordings of clinical interaction between patients and clinicians. By designing larger and comparative studies, a distinct digital phenotype could be defined for each of the aforesaid 3 groups. Thus, digital markers to help differential diagnosis could emerge.

Moreover, if large prospective studies are conducted, validated digital markers could be used for follow-up and for predicting treatment response or risk of relapse. Digital phenotyping could therefore contribute to personalized and precision medicine.

Acknowledgments

This research is funded by the MEPHESTO project (BMBF grant number 01IS20075; DRI-0120105/291). We thank the "Innovation Alzheimer" Association for its financial support to the publication.

Authors' Contributions

EE, AK, PM, JH, BG, DP, HL, MB, AL, and PR conceptualized the study. EE and AK wrote and prepared the original draft. AK, PM, JH, and PR provided visualization and supervision support. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strings.
[DOCX File , 72 KB - mental_v10i1e37225_app1.docx]

Multimedia Appendix 2
Summary of speech analysis, nonverbal behavioral analysis, HRV, and EDA studies in individuals with BD, PTSD or psychological trauma. BD: Bipolar Disorder; EDA: electrodermal activity; HRV: heart rate variability; MDD: major depressive disorder; PTSD: Post Traumatic Stress Disorder.
[DOCX File , 34 KB - mental_v10i1e37225_app2.docx]

References

1. GBD 2017 DiseaseInjury Incidence Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis

for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Diagnostic and statistical manual of mental disorders (DSM), 5th Edition. Washington, DC: American Psychiatric Association; 2013:XIV-947.
3. Ostergaard SD, Jensen SOW, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. Acta Psychiatr Scand 2011 Dec;124(6):495-496. [doi: 10.1111/j.1600-0447.2011.01744.x] [Medline: 21838736]
4. Gaynes BN, Warden D, Trivedi MH, Wisniewski SR, Fava M, Rush AJ. What did STAR*D teach us? Results from a large-scale, practical, clinical trial for patients with depression. Psychiatr Serv 2009 Nov;60(11):1439-1445. [doi: 10.1176/ps.2009.60.11.1439] [Medline: 19880458]
5. Hirschfeld R. Differential diagnosis of bipolar disorder and major depressive disorder. J Affect Disord 2014 Dec;169:S12-S16. [doi: 10.1016/s0165-0327(14)70004-7]
6. Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. Arch Gen Psychiatry 2002 Jun 01;59(6):530-537. [doi: 10.1001/archpsyc.59.6.530] [Medline: 12044195]
7. Angst J, Merikangas KR, Cui L, Van Meter A, Ajdacic-Gross V, Rössler W. Bipolar spectrum in major depressive disorders. Eur Arch Psychiatry Clin Neurosci 2018 Dec 21;268(8):741-748. [doi: 10.1007/s00406-018-0927-x] [Medline: 30032467]
8. Hirschfeld RMA, Lewis L, Vornik LA. Perceptions and Impact of Bipolar Disorder. J Clin Psychiatry 2003 Feb 15;64(2):161-174. [doi: 10.4088/jcp.v64n0209]
9. Galvão F, Sportiche S, Lambert J, Amiez M, Musa C, Nieto I, et al. Clinical differences between unipolar and bipolar depression: interest of BDRS (Bipolar Depression Rating Scale). Compr Psychiatry 2013 Aug;54(6):605-610. [doi: 10.1016/j.comppsych.2012.12.023] [Medline: 23375261]
10. Rytwinski NK, Scur MD, Feeny NC, Youngstrom EA. The co-occurrence of major depressive disorder among individuals with posttraumatic stress disorder: a meta-analysis. J Trauma Stress 2013 Jun 20;26(3):299-309. [doi: 10.1002/jts.21814] [Medline: 23696449]
11. Gootzeit J, Markon K. Factors of PTSD: differential specificity and external correlates. Clin Psychol Rev 2011 Aug;31(6):993-1003. [doi: 10.1016/j.cpr.2011.06.005] [Medline: 21741913]
12. Gros DF, Price M, Magruder KM, Frueh BC. Symptom overlap in posttraumatic stress disorder and major depression. Psychiatry Res 2012 Apr 30;196(2-3):267-270 [FREE Full text] [doi: 10.1016/j.psychres.2011.10.022] [Medline: 22386220]
13. McFarlane AC, Papay P. Multiple diagnoses in posttraumatic stress disorder in the victims of a natural disaster. J Nerv Ment Dis 1992 Aug;180(8):498-504. [doi: 10.1097/00005053-199208000-00004] [Medline: 1500931]
14. Alarcon RD, Glover SG, Deering CG. The cascade model: an alternative to comorbidity in the pathogenesis of posttraumatic stress disorder. Psychiatry 1999 Sep 26;62(2):114-124. [doi: 10.1080/00332747.1999.11024860] [Medline: 10420426]
15. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
16. Waring OM, Majumder MS. Introduction to Digital Phenotyping for Global Health. In: Celi LA, Majumder MS, Ordóñez P, Osorio JS, Paik KE, Somai M, editors. Leveraging Data Science for Global Health. Cham, Switzerland: Springer International Publishing; 2020:251-261.
17. Andrea A, Agulia A, Serafini G, Amore M. Digital biomarkers and digital phenotyping in mental health care and prevention. Eur J Public Health 2020;30(Supplement_5):161-166. [doi: 10.1093/eurpub/ckaa165.1080]
18. Brietzke E, Hawken ER, Idzikowski M, Pong J, Kennedy SH, Soares CN. Integrating digital phenotyping in clinical characterization of individuals with mood disorders. Neurosci Biobehav Rev 2019 Sep;104:223-230. [doi: 10.1016/j.neubiorev.2019.07.009] [Medline: 31330197]
19. Schilbach L. Using interaction-based phenotyping to assess the behavioral and neural mechanisms of transdiagnostic social impairments in psychiatry. Eur Arch Psychiatry Clin Neurosci 2019 Apr 13;269(3):273-274. [doi: 10.1007/s00406-019-00998-y] [Medline: 30868216]
20. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019 Sep 6;2(1):88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
21. Alghowinem S. Detecting depression: A comparison between spontaneous and read speech. New York, NY: IEEE; 2013 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing; May 26-31, 2013; Vancouver, Canada p. 7547-7551. [doi: 10.1109/ICASSP.2013.6639130]
22. Arevian AC, Bone D, Malandrakis N, Martinez VR, Wells KB, Miklowitz DJ, et al. Clinical state tracking in serious mental illness through computational analysis of speech. PLoS ONE 2020 Jan 15;15(1):e0225695 [FREE Full text] [doi: 10.1371/journal.pone.0225695] [Medline: 31940347]
23. Hönig F. Automatic modelling of depressed speech: relevant features and relevance of gender. 2014 Presented at: Interspeech 2014; September 14-18, 2014; Singapore City, Singapore p. 1248-1252 URL: https://tinyurl.com/bdem6mj7 [doi: 10.21437/interspeech.2014-313]

24. Horwitz R. On the relative importance of vocal source, system,prosody in human depression. New York, NY: IEEE; 2013 Presented at: IEEE International Conference on Body Sensor Networks; May 6-9, 2013; Cambridge, MA p. 1-6. [doi: 10.1109/bsn.2013.6575522]
25. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. Biol Psychiatry 2012 Oct 01;72(7):580-587 [FREE Full text] [doi: 10.1016/j.biopsych.2012.03.015] [Medline: 22541039]
26. Quatieri T. Vocal-Source Biomarkers for Depression: A Link to Psychomotor Activity. 2012 Presented at: Interspeech 2012; September 9-13, 2012; Portland, OR p. 1059-1062 URL: https://www.researchgate.net/publication/283629170_Vocal-source_biomarkers_for_depression_A_link_to_psychomotor_activity [doi: 10.21437/interspeech.2012-311]
27. Shinohara S, Nakamura M, Omiya Y, Higuchi M, Hagiwara N, Mitsuyoshi S, et al. Depressive Mood Assessment Method Based on Emotion Level Derived from Voice: Comparison of Voice Features of Individuals with Major Depressive Disorders and Healthy Controls. Int J Environ Res Public Health 2021 May 19;18(10):5435 [FREE Full text] [doi: 10.3390/ijerph18105435] [Medline: 34069609]
28. Taguchi T, Tachikawa H, Nemoto K, Suzuki M, Nagano T, Tachibana R, et al. Major depressive disorder discrimination using vocal acoustic features. J Affect Disord 2018 Jan 01;225:214-220. [doi: 10.1016/j.jad.2017.08.038] [Medline: 28841483]
29. Wang J, Zhang L, Liu T, Pan W, Hu B, Zhu T. Acoustic differences between healthy and depressed people: a cross-situation study. BMC Psychiatry 2019 Oct 15;19(1):300 [FREE Full text] [doi: 10.1186/s12888-019-2300-7] [Medline: 31615470]
30. Xu S, Yang Z, Chakraborty D, Chua YHV, Dauwels J, Thalmann D, et al. Automated Verbal and Non-verbal Speech Analysis of Interviews of Individuals with Schizophrenia and Depression. New York, NY: IEEE; 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 23-27, 2019; Berlin, Germany p. 225. [doi: 10.1109/embc.2019.8857071]
31. Yamamoto M, Takamiya A, Sawada K, Yoshimura M, Kitazawa M, Liang K, et al. Using speech recognition technology to investigate the association between timing-related speech features and depression severity. PLoS One 2020 Sep 11;15(9):e0238726 [FREE Full text] [doi: 10.1371/journal.pone.0238726] [Medline: 32915846]
32. Alghowinem S, Goecke R, Cohn JF, Wagner G. Cross-cultural detection of depression from nonverbal behaviour. 2015 Presented at: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition; May 4-8; Ljubljana, Slovenia p. 1-8. [doi: 10.1109/fg.2015.7163113]
33. Bhatia S, Goecke R, Hammal Z, Cohn JF. Automated Measurement of Head Movement Synchrony during Dyadic Depression Severity Interviews. 2019 Presented at: 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019); May 14-18; Lille, France p. 1-8. [doi: 10.1109/fg.2019.8756509]
34. Dibeklioglu H, Hammal Z, Cohn JF. Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding. IEEE J Biomed Health Inform 2018 Mar;22(2):525-536. [doi: 10.1109/jbhi.2017.2676878]
35. Fiquer JT, Boggio PS, Gorenstein C. Talking bodies: nonverbal behavior in the assessment of depression severity. J Affect Disord 2013 Sep 25;150(3):1114-1119. [doi: 10.1016/j.jad.2013.05.002] [Medline: 23706840]
36. Fiquer JT, Moreno RA, Canales JZ, Cavalcanti A, Gorenstein C. Is nonverbal behavior in patients and interviewers relevant to the assessment of depression and its recovery? A study with Dutch and Brazilian patients. Psychiatry Res 2017 Apr;250:59-64. [doi: 10.1016/j.psychres.2017.01.053] [Medline: 28142067]
37. Fiquer JT, Moreno RA, Brunoni AR, Barros VB, Fernandes F, Gorenstein C. What is the nonverbal communication of depression? Assessing expressive differences between depressive patients and healthy volunteers during clinical interviews. J Affect Disord 2018 Oct 01;238:636-644. [doi: 10.1016/j.jad.2018.05.071] [Medline: 29957481]
38. Girard JM, Cohn JF, Mahoor MH, Mavadati S, Rosenwald DP. Social risk and depression: Evidence from manual and automatic facial expression analysis. 2013 Presented at: 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition; April 22-26; Shanghai, China p. 1-8. [doi: 10.1109/fg.2013.6553748]
39. Girard JM, Cohn JF, Mahoor MH, Mavadati SM, Hammal Z, Rosenwald DP. Nonverbal Social Withdrawal in Depression: Evidence from manual and automatic analysis. Image Vis Comput 2014 Oct;32(10):641-647 [FREE Full text] [doi: 10.1016/j.imavis.2013.12.007] [Medline: 25378765]
40. Guo W, Yang H, Liu Z, Xu Y, Hu B. Deep Neural Networks for Depression Recognition Based on 2D and 3D Facial Expressions Under Emotional Stimulus Tasks. Front Neurosci 2021 Apr 23;15:609760 [FREE Full text] [doi: 10.3389/fnins.2021.609760] [Medline: 33967675]
41. Jiang Z, Harati S, Crowell A, Mayberg HS, Nemati S, Clifford GD. Classifying Major Depressive Disorder and Response to Deep Brain Stimulation Over Time by Analyzing Facial Expressions. IEEE Trans Biomed Eng 2021 Feb;68(2):664-672. [doi: 10.1109/tbme.2020.3010472]
42. Kacem A, Hammal Z, Daoudi M, Cohn J. Detecting Depression Severity by Interpretable Representations of Motion Dynamics. 2018 Presented at: 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018); May 15-19; Xi'an, China p. 739-745. [doi: 10.1109/fg.2018.00116]
43. Adolph D, Teismann T, Forkmann T, Wannemüller A, Margraf J. High frequency heart rate variability: Evidence for a transdiagnostic association with suicide ideation. Biol Psychol 2018 Oct;138:165-171. [doi: 10.1016/j.biopsycho.2018.09.006] [Medline: 30253232]

44. Fernandes RA, Fiquer JT, Gorenstein C, Razza LB, Fraguas R, Borrione L, et al. Nonverbal behaviors are associated with increased vagal activity in major depressive disorder: Implications for the polyvagal theory. J Affect Disord 2017 Feb;209:18-22. [doi: 10.1016/j.jad.2016.11.010] [Medline: 27870941]
45. Giurgi-Oncu C, Tudoran C, Enatescu VR, Tudoran M, Pop GN, Bredicean C. Evolution of Heart Rate Variability and Heart Rate Turbulence in Patients with Depressive Illness Treated with Selective Serotonin Reuptake Inhibitors. Medicina (Kaunas) 2020 Nov 05;56(11):590 [FREE Full text] [doi: 10.3390/medicina56110590] [Medline: 33167574]
46. Hartmann R, Schmidt FM, Sander C, Hegerl U. Heart Rate Variability as Indicator of Clinical State in Depression. Front Psychiatry 2018 Jan 17;9:735 [FREE Full text] [doi: 10.3389/fpsyt.2018.00735] [Medline: 30705641]
47. Kircanski K, Williams LM, Gotlib IH. Heart rate variability as a biomarker of anxious depression response to antidepressant medication. Depress Anxiety 2019 Jan 12;36(1):63-71 [FREE Full text] [doi: 10.1002/da.22843] [Medline: 30311742]
48. Lee M, Seo YE, Mok YE, Lee SH. Heart Rate Variability after Treatment for Depression in North Korean Defectors. Appl Psychophysiol Biofeedback 2021 Mar 19;46(1):11-18. [doi: 10.1007/s10484-020-09491-y] [Medline: 33074464]
49. Neyer S, Witthöft M, Cropley M, Pawelzik M, Lugo RG, Sütterlin S. Reduction of depressive symptoms during inpatient treatment is not associated with changes in heart rate variability. PLoS One 2021 Mar 23;16(3):e0248686 [FREE Full text] [doi: 10.1371/journal.pone.0248686] [Medline: 33755668]
50. Sarlon J, Staniloiu A, Kordon A. Heart Rate Variability Changes in Patients With Major Depressive Disorder: Related to Confounding Factors, Not to Symptom Severity? Front Neurosci 2021 Jul 5;15:675624 [FREE Full text] [doi: 10.3389/fnins.2021.675624] [Medline: 34326716]
51. Kim AY, Jang EH, Kim S, Choi KW, Jeon HJ, Yu HY, et al. Automatic detection of major depressive disorder using electrodermal activity. Sci Rep 2018;8(1):1. [doi: 10.1038/s41598-018-35147-3]
52. Kim AY, Jang EH, Choi KW, Jeon HJ, Byun S, Sim JY, et al. Skin conductance responses in Major Depressive Disorder (MDD) under mental arithmetic stress. PLoS One 2019 Apr 3;14(4):e0213140 [FREE Full text] [doi: 10.1371/journal.pone.0213140] [Medline: 30943195]
53. Litwińska-Bołtuć M, Święcicki ?, Spreco A, Timpka T. Clinical effectiveness of the electrodermal orienting reactivity test for evaluating relapse and recurrence risk in patients hospitalized for depression. BMC Psychiatry 2021 Feb 10;21(1):88 [FREE Full text] [doi: 10.1186/s12888-021-03088-3] [Medline: 33568134]
54. Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhathena D, et al. Monitoring Changes in Depression Severity Using Wearable and Mobile Sensors. Front Psychiatry 2020 Dec 18;11:584711 [FREE Full text] [doi: 10.3389/fpsyt.2020.584711] [Medline: 33391050]
55. Smith LT, Levita L, Amico F, Fagan J, Yek JH, Brophy J, et al. Using Resting State Heart Rate Variability and Skin Conductance Response to Detect Depression in Adults. Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) 2020:5004-5007. [doi: 10.1109/embc44109.2020.9176304]
56. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Communication 2015 Jul;71:10-49. [doi: 10.1016/j.specom.2015.03.004]
57. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: A systematic review. Laryngoscope Investig Otolaryngol 2020 Feb 31;5(1):96-116 [FREE Full text] [doi: 10.1002/lio2.354] [Medline: 32128436]
58. Kiss G, Vicsi K. Mono- and multi-lingual depression prediction based on speech processing. Int J Speech Technol 2017 Sep 20;20(4):919-935. [doi: 10.1007/s10772-017-9455-8]
59. Geerts E, Brune M. Ethological approaches to psychiatric disorders: focus on depression and schizophrenia. Aus & New Zealand J of Psychiatry 2009 Nov;43(11):1007-1015. [doi: 10.1080/00048670903270498]
60. Alghowinem SM, Gedeon T, Goecke R, Cohn J, Parker G. Interpretation of Depression Detection Models via Feature Selection Methods. IEEE Trans Affective Comput 2020:1-1. [doi: 10.1109/taffc.2020.3035535]
61. Servant D, Logier R, Mouster Y, Goudemand M. [Heart rate variability. Applications in psychiatry]. Encephale 2009 Oct;35(5):423-428. [doi: 10.1016/j.encep.2008.06.016] [Medline: 19853714]
62. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, et al. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal 1996 Mar 01;17(3):354-381. [doi: 10.1093/oxfordjournals.eurheartj.a014868]
63. Appelhans BM, Luecken LJ. Heart rate variability and pain: associations of two interrelated homeostatic processes. Biol Psychol 2008 Feb;77(2):174-182. [doi: 10.1016/j.biopsycho.2007.10.004] [Medline: 18023960]
64. Schwerdtfeger AR, Schwarz G, Pfurtscheller K, Thayer JF, Jarczok MN, Pfurtscheller G. Heart rate variability (HRV): From brain death to resonance breathing at 6 breaths per minute. Clin Neurophysiol 2020 Mar;131(3):676-693. [doi: 10.1016/j.clinph.2019.11.013] [Medline: 31978852]
65. Hillebrand S, Gast KB, de Mutsert R, Swenne CA, Jukema JW, Middeldorp S, et al. Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: meta-analysis and dose-response meta-regression. Europace 2013 May;15(5):742-749. [doi: 10.1093/europace/eus341] [Medline: 23370966]
66. Koch C, Wilhelm M, Salzmann S, Rief W, Euteneuer F. A meta-analysis of heart rate variability in major depression. Psychol Med 2019 Jun 26;49(12):1948-1957. [doi: 10.1017/s0033291719001351]

67. Sarchiapone M, Gramaglia C, Iosue M, Carli V, Mandelli L, Serretti A, et al. The association between electrodermal activity (EDA), depression and suicidal behaviour: A systematic review and narrative synthesis. BMC Psychiatry 2018 Jan 25;18(1):22 [FREE Full text] [doi: 10.1186/s12888-017-1551-4] [Medline: 29370787]
68. Bloch V. New aspects of the psychogalvanic or electrodermographic method as a criterion of emotional tensions. L'annee Psychologique 1952:329-362.
69. Sequeira H, Hot P, Silvert L, Delplanque S. Electrical autonomic correlates of emotion. Int J Psychophysiol 2009 Jan;71(1):50-56. [doi: 10.1016/j.ijpsycho.2008.07.009] [Medline: 18723054]
70. Balconi M, Brambilla E, Falbo L. Appetitive vs. defensive responses to emotional cues. Autonomic measures and brain oscillation modulation. Brain Res 2009 Nov 03;1296:72-84. [doi: 10.1016/j.brainres.2009.08.056] [Medline: 19703433]
71. Crider A. Personality and electrodermal response lability: an interpretation. Appl Psychophysiol Biofeedback 2008 Sep 29;33(3):141-148. [doi: 10.1007/s10484-008-9057-y] [Medline: 18509756]
72. Hamm A. Affective blindsight: intact fear conditioning to a visual cue in a cortically blind patient. Brain 2003:267-275. [doi: 10.1093/brain/awg037]
73. Bauer RM, Verfaellie M. Electrodermal discrimination of familiar but not unfamiliar faces in prosopagnosia. Brain and Cognition 1988 Oct;8(2):240-252. [doi: 10.1016/0278-2626(88)90052-8]
74. Silvert L, Delplanque S, Bouwalerh H, Verpoort C, Sequeira H. Autonomic responding to aversive words without conscious valence discrimination. Int J Psychophysiol 2004 Jul;53(2):135-145. [doi: 10.1016/j.ijpsycho.2004.03.005] [Medline: 15210291]
75. Bernstein AS, Schnur DB, Bernstein P, Yeager A, Wrable J, Smith S. Differing patterns of electrodermal and finger pulse responsivity in schizophrenia and depression. Psychol Med 1995;25(1):51-62. [doi: 10.1017/s0033291700028087]
76. Lader M, Wing L. Physiological measures in agitated and retarded depressed patients. J Psychiatr Res 1969 Dec;7(2):89-100. [doi: 10.1016/0022-3956(69)90014-4]
77. Thorell L. Electrodermal activity in suicidal and nonsuicidal depressive patients and in matched healthy subjects. Acta Psychiatr Scand 1987 Oct;76(4):420-430. [doi: 10.1111/j.1600-0447.1987.tb05627.x]
78. Byrne DG. A Psychophysiological Distinction between Types of Depressive States. Aust N Z J Psychiatry 1975;9(3):181-185. [doi: 10.3109/00048677509159846]
79. Mirkin AM, Coppen A. Electrodermal Activity in Depression: Clinical and Biochemical Correlates. Br J Psychiatry 1980;137(1):93-97. [doi: 10.1192/bjp.137.1.93]
80. Bouhuys A, Jansen C, van den Hoofdakker R. Analysis of observed behaviors displayed by depressed patients during a clinical interview: relationships between behavioral factors and clinical concepts of activation. J Affect Disord 1991 Feb;21(2):79-88. [doi: 10.1016/0165-0327(91)90053-u]
81. Bouhuys A, Van den Hoofdakker R. A longitudinal study of interaction patterns of a psychiatrist and severely depressed patients based on observed behaviour: an ethological approach of interpersonal theories of depression. J Affect Disord 1993 Feb;27(2):87-99. [doi: 10.1016/0165-0327(93)90081-t]
82. Geerts E, Bouhuys AL, Bloem GM. Nonverbal support giving induces nonverbal support seeking in depressed patients. J Clin Psychol 1997 Jan;53(1):35-39. [doi: 10.1002/(sici)1097-4679(199701)53:1<35::aid-jclp5>3.0.co;2-u]
83. Bouhuys A, Albersnagel F. Do interactional capacities based on observed behaviour interfere with improvement in severely depressed patients? J Affect Disord 1992 Jun;25(2):107-116. [doi: 10.1016/0165-0327(92)90073-f]
84. Maxhuni A, Muñoz-Meléndez A, Osmani V, Perez H, Mayora O, Morales EF. Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. Pervasive and Mobile Computing 2016 Sep;31:50-66. [doi: 10.1016/j.pmcj.2016.01.008]
85. Guidi A, Schoentgen J, Bertschy G, Gentili C, Scilingo E, Vanello N. Features of vocal frequency contour and speech rhythm in bipolar disorder. Biomedical Signal Processing and Control 2017 Aug;37:23-31. [doi: 10.1016/j.bspc.2017.01.017]
86. Guidi A. Voice quality in patients suffering from bipolar disease. New York, NY: IEEE; 2015 Presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 25-29, 2015; Milan, Italy p. 6106-6109. [doi: 10.1109/embc.2015.7319785]
87. Guidi A. Analysis of running speech for the characterization of mood state in bipolar patients. New York, NY: IEEE; 2015 Presented at: 2015 AEIT International Annual Conference (AEIT); October 14-16, 2015; Naples, Italy p. 1-6. [doi: 10.1109/aeit.2015.7415275]
88. Vanello N. Speech analysis for mood state characterization in bipolar patients. New York, NY: IEEE; 2012 Presented at: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 28, 2012, to September 1, 2012; San Diego, CA, USA p. 2104-2107. [doi: 10.1109/embc.2012.6346375]
89. Zhang J, Pan Z, Gui C, Xue T, Lin Y, Zhu J, et al. Analysis on speech signal features of manic patients. J Psychiatr Res 2018 Mar;98:59-63. [doi: 10.1016/j.jpsychires.2017.12.012] [Medline: 29291581]
90. Gideon J, Provost EM, McInnis M. Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. 2016 Mar Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2016; Shanghai, China p. 2359-2363 URL: http://europepmc.org/abstract/MED/27570493 [doi: 10.1109/ICASSP.2016.7472099]

91. Karam ZN. Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. New York, NY: IEEE; 2014 Presented at: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 4-9, 2014; Florence, italy p. 4858-4862. [doi: 10.1109/icassp.2014.6854525]
92. Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, et al. Voice analysis as an objective state marker in bipolar disorder. Transl Psychiatry 2016 Jul 19;6(7):e856-e856 [FREE Full text] [doi: 10.1038/tp.2016.123] [Medline: 27434490]
93. Faurholt-Jepsen M, Kessing LV, Munkholm K. Heart rate variability in bipolar disorder: A systematic review and meta-analysis. Neurosci Biobehav Rev 2017 Feb;73:68-80. [doi: 10.1016/j.neubiorev.2016.12.007] [Medline: 27986468]
94. Wazen GLL, Gregório ML, Kemp AH, Godoy MFD. Heart rate variability in patients with bipolar disorder: From mania to euthymia. J Psychiatr Res 2018 Apr;99:33-38. [doi: 10.1016/j.jpsychires.2018.01.008] [Medline: 29407285]
95. Hage B, Britton B, Daniels D, Heilman K, Porges SW, Halaris A. Diminution of Heart Rate Variability in Bipolar Depression. Front Public Health 2017 Dec 06;5:312 [FREE Full text] [doi: 10.3389/fpubh.2017.00312] [Medline: 29270399]
96. Benjamin BR, Valstad M, Elvsåshagen T, Jönsson EG, Moberget T, Winterton A, et al. Heart rate variability is associated with disease severity in psychosis spectrum disorders. Prog Neuropsychopharmacol Biol Psychiatry 2021 Dec 20;111:110108 [FREE Full text] [doi: 10.1016/j.pnpbp.2020.110108] [Medline: 32946948]
97. Ortiz A, Bradler K, Moorti P, MacLean S, Husain MI, Sanches M, et al. Reduced heart rate variability is associated with higher illness burden in bipolar disorder. J Psychosom Res 2021 Jun;145:110478. [doi: 10.1016/j.jpsychores.2021.110478] [Medline: 33820643]
98. Freyberg J, Brage S, Kessing LV, Faurholt-Jepsen M. Differences in psychomotor activity and heart rate variability in patients with newly diagnosed bipolar disorder, unaffected relatives, and healthy individuals. J Affect Disord 2020 Apr 01;266:30-36 [FREE Full text] [doi: 10.1016/j.jad.2020.01.110] [Medline: 32056891]
99. Chang H, Chang C, Kuo TBJ, Huang S. Distinguishing bipolar II depression from unipolar major depressive disorder: Differences in heart rate variability. World J Biol Psychiatry 2015 Mar 24;16(5):351-360. [doi: 10.3109/15622975.2015.1017606] [Medline: 25800950]
100. Hage B, Britton B, Daniels D, Heilman K, Porges SW, Halaris A. Low cardiac vagal tone index by heart rate variability differentiates bipolar from major depression. World J Biol Psychiatry 2019 Jun 05;20(5):359-367. [doi: 10.1080/15622975.2017.1376113] [Medline: 28895492]
101. Greco A, Valenza G, Lanata A, Rota G, Scilingo EP. Electrodermal activity in bipolar patients during affective elicitation. IEEE J Biomed Health Inform 2014 Nov;18(6):1865-1873. [doi: 10.1109/JBHI.2014.2300940] [Medline: 25375684]
102. Lemaire M, El-Hage W, Frangou S. Increased affective reactivity to neutral stimuli and decreased maintenance of affective responses in bipolar disorder. Eur Psychiatry 2015 Oct;30(7):852-860. [doi: 10.1016/j.eurpsy.2015.07.008] [Medline: 26443053]
103. Faurholt-Jepsen M, Brage S, Kessing LV, Munkholm K. State-related differences in heart rate variability in bipolar disorder. J Psychiatr Res 2017 Jan;84:169-173 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.10.005] [Medline: 27743529]
104. Monti E, D'Andrea W, Freed S, Kidd DC, Feuer S, Carroll LM, et al. Does Self-Reported Childhood Trauma Relate to Vocal Acoustic Measures? Preliminary Findings at Trauma Recall. J Nonverbal Behav 2021 Jan 22;45(3):389-408. [doi: 10.1007/s10919-020-00355-x]
105. Monti E, Kidd DC, Carroll LM, Castano E. What's in a singer's voice: The effect of attachment, emotions and trauma. Logoped Phoniatr Vocol 2017 Jul;42(2):62-72. [doi: 10.3109/14015439.2016.1166394] [Medline: 27075743]
106. Schultebraucks K, Yadav V, Shalev AY, Bonanno GA, Galatzer-Levy IR. Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood. Psychol Med 2020 Aug 03;52(5):957-967. [doi: 10.1017/s0033291720002718]
107. Kleim B, Horn AB, Kraehenmann R, Mehl MR, Ehlers A. Early Linguistic Markers of Trauma-Specific Processing Predict Post-trauma Adjustment. Front Psychiatry 2018;9:645 [FREE Full text] [doi: 10.3389/fpsyt.2018.00645] [Medline: 30568607]
108. He Q, Veldkamp BP, Glas CAW, de Vries T. Automated Assessment of Patients' Self-Narratives for Posttraumatic Stress Disorder Screening Using Natural Language Processing and Text Mining. Assessment 2017 Mar;24(2):157-172. [doi: 10.1177/1073191115602551] [Medline: 26358713]
109. Huon P, Stutz N. Les marqueurs linguistiques du temps et de la subjectivité dans le récit de psychotraumatisme. L'Évolution Psychiatrique 2020 Oct;85(4):479-508. [doi: 10.1016/j.evopsy.2020.06.008]
110. Scherer S. Investigating voice quality as a speaker-independent indicator of depression and PTSD. Interspeech 2013:851. [doi: 10.21437/interspeech.2013-240]
111. Xu R, Mei G, Zhang G, Gao P, Judkins T, Cannizzaro M, et al. A voice-based automated system for PTSD screening and monitoring. Stud Health Technol Inform 2012;173:552-558. [Medline: 22357057]
112. Scherer S. Self-Reported Symptoms of Depression and PTSD Are Associated with Reduced Vowel Space in Screening Interviews. IEEE Transactions on Affective Computing 2016 Jan 1;7(1):59-73. [doi: 10.1109/TAFFC.2015.2440264]
113. Marmar CR, Brown AD, Qian M, Laska E, Siegel C, Li M, et al. Speech-based markers for posttraumatic stress disorder in US veterans. Depress Anxiety 2019 Jul 22;36(7):607-616 [FREE Full text] [doi: 10.1002/da.22890] [Medline: 31006959]

114. McTeague LM, Lang PJ, Laplante M, Cuthbert BN, Shumen JR, Bradley MM. Aversive imagery in posttraumatic stress disorder: trauma recurrence, comorbidity, and physiological reactivity. Biol Psychiatry 2010 Feb 15;67(4):346-356 [FREE Full text] [doi: 10.1016/j.biopsych.2009.08.023] [Medline: 19875104]
115. Blechert J, Michael T, Wilhelm FH. Video-Based Analysis of Bodily Startle and Subsequent Emotional Facial Expression in Posttraumatic Stress Disorder. J Exp Psychopathol 2019 May 21;4(4):435-447 [FREE Full text] [doi: 10.5127/jep.030712] [Medline: 31156801]
116. Lazarov A, Suarez-Jimenez B, Tamman A, Falzon L, Zhu X, Edmondson DE, et al. Attention to threat in posttraumatic stress disorder as indexed by eye-tracking indices: a systematic review. Psychol Med 2018 Sep 04;49(5):705-726. [doi: 10.1017/s0033291718002313]
117. Mekawi Y, Murphy L, Munoz A, Briscione M, Tone EB, Norrholm SD, et al. The role of negative affect in the association between attention bias to threat and posttraumatic stress: An eye-tracking study. Psychiatry Res 2020 Feb;284:112674 [FREE Full text] [doi: 10.1016/j.psychres.2019.112674] [Medline: 31831200]
118. Katz C, Hershkowitz I, Malloy LC, Lamb ME, Atabaki A, Spindler S. Non-verbal behavior of children who disclose or do not disclose child abuse in investigative interviews. Child Abuse Negl 2012 Jan;36(1):12-20. [doi: 10.1016/j.chiabu.2011.08.006] [Medline: 22265935]
119. Schneider M, Schwerdtfeger A. Autonomic dysfunction in posttraumatic stress disorder indexed by heart rate variability: a meta-analysis. Psychol Med 2020 Aug 28;50(12):1937-1948. [doi: 10.1017/s003329172000207x]
120. Ge F, Yuan M, Li Y, Zhang W. Posttraumatic Stress Disorder and Alterations in Resting Heart Rate Variability: A Systematic Review and Meta-Analysis. Psychiatry Investig 2020 Jan;17(1):9-20 [FREE Full text] [doi: 10.30773/pi.2019.0112] [Medline: 31995968]
121. Sigrist C, Mürner-Lavanchy I, Peschel SK, Schmidt SJ, Kaess M, Koenig J. Early life maltreatment and resting-state heart rate variability: A systematic review and meta-analysis. Neurosci Biobehav Rev 2021 Jan;120:307-334. [doi: 10.1016/j.neubiorev.2020.10.026] [Medline: 33171141]
122. Stone LB, Amole MC, Cyranowski JM, Swartz HA. History of childhood emotional abuse predicts lower resting-state high-frequency heart rate variability in depressed women. Psychiatry Res 2018 Nov;269:681-687 [FREE Full text] [doi: 10.1016/j.psychres.2018.08.106] [Medline: 30273892]
123. Hopper JW, Frewen PA, Sack M, Lanius RA, van der Kolk BA. The Responses to Script-Driven Imagery Scale (RSDI): Assessment of State Posttraumatic Symptoms for Psychobiological and Treatment Research. J Psychopathol Behav Assess 2007 May 18;29(4):249-268. [doi: 10.1007/s10862-007-9046-0]
124. Sack M, Cillien M, Hopper JW. Acute dissociation and cardiac reactivity to script-driven imagery in trauma-related disorders. Eur J Psychotraumatol 2012;3(1):17419 [FREE Full text] [doi: 10.3402/ejpt.v3i0.17419] [Medline: 23198029]
125. Nemeroff CB, Bremner JD, Foa EB, Mayberg HS, North CS, Stein MB. Posttraumatic stress disorder: a state-of-the-science review. J Psychiatr Res 2006 Feb;40(1):1-21. [doi: 10.1016/j.jpsychires.2005.07.005] [Medline: 16242154]
126. O'Kearney R, Parry L. Comparative physiological reactivity during script-driven recall in depression and posttraumatic stress disorder. J Abnorm Psychol 2014 Aug;123(3):523-532. [doi: 10.1037/a0037326] [Medline: 25000153]
127. D'Andrea W, Pole N, DePierro J, Freed S, Wallace DB. Heterogeneity of defensive responses after exposure to trauma: blunted autonomic reactivity in response to startling sounds. Int J Psychophysiol 2013 Oct;90(1):80-89. [doi: 10.1016/j.ijpsycho.2013.07.008] [Medline: 23896169]

Abbreviations

AUC: area under the curve
BD: bipolar disorder
CEA: childhood emotional abuse
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EDA: electrodermal activity
ELM: early life maltreatment
F0: fundamental frequency
GAF: Global Assessment of Functioning Scale
HAM-D: Hamilton Rating Scale for Depression
HC: healthy control
HDRS: Hamilton Depression Rating Scale
HF: high frequency
HNR: harmonic-to-noise ratio
HR: heart rate
HRV: heart rate variability
ICD-11: International Classification of Diseases 11th Revision
LF: low frequency
MDD: major depressive disorder

MDE: major depressive episode
MFCC: mel-frequency cepstral coefficient
PANSS: Positive and Negative Syndrome Scale
PNN50: proportion of NN50 divided by the total number of NN (R-R) intervals
PNS: parasympathetic nervous system
PTSD: posttraumatic stress disorder
RMSSD: root-mean-square surface distance
RR: beat-to-beat interval
RSA: respiratory sinus arrhythmia
SC: skin conductance
SCL: skin conductance level
SCR: skin conductance response
SD1: SD of points perpendicular to the major axis of the Poincaré plot
SD2: SD of points along the major axis of the Poincaré plot
SDNN: SD of the NN (R-R) intervals
SNS: sympathetic nervous system
TAT: Thematic Appreciation Test
VLF: very-low frequency

Edited by J Torous; submitted 11.02.22; peer-reviewed by S Mangelsdorf, G Mayer, K Reuter; comments to author 30.05.22; revised version received 02.09.22; accepted 30.09.22; published 23.01.23.

Please cite as:
Ettore E, Müller P, Hinze J, Riemenschneider M, Benoit M, Giordana B, Postin D, Hurlemann R, Lecomte A, Musiol M, Lindsay H, Robert P, König A
Digital Phenotyping for Differential Diagnosis of Major Depressive Episode: Narrative Review
JMIR Ment Health 2023;10:e37225
URL: https://mental.jmir.org/2023/1/e37225
doi:10.2196/37225
PMID:36689265

Original Paper

The Efficacy of Be a Mom, a Web-Based Intervention to Prevent Postpartum Depression: Examining Mechanisms of Change in a Randomized Controlled Trial

Carlos Carona[1], PhD; Marco Pereira[1], PhD; Anabela Araújo-Pedrosa[1,2], PhD; Maria Cristina Canavarro[1], PhD; Ana Fonseca[1], PhD

[1]Center for Research in Neuropsychology and Cognitive-Behavioral Intervention, University of Coimbra, Coimbra, Portugal

[2]Obstetrics Unit A, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal

Corresponding Author:
Carlos Carona, PhD
Center for Research in Neuropsychology and Cognitive-Behavioral Intervention
University of Coimbra
Colégio Novo St.
Coimbra, 3000-115
Portugal
Phone: 351 239851450
Email: ccarona@fpce.uc.pt

Abstract

Background: Postpartum depression (PPD) is treatable and preventable, but most women do not seek professional help for their perinatal depressive symptoms. One increasingly popular approach of improving access to care is the use of web-based intervention programs.

Objective: The objective of this study was 2-fold: first, to assess the efficacy of *Be a Mom*, a brief web-based selective or indicated preventive intervention, in reducing depressive and anxiety symptoms of women at high risk for PPD; and second, to examine mechanisms of change linking modifiable self-regulatory skills (ie, emotion regulation, self-compassion, and psychological flexibility) to improved perinatal mental health outcomes.

Methods: This 2-arm, open-label randomized controlled trial involved a sample of 1053 perinatal women presenting high risk for PPD who were allocated to the *Be a Mom* intervention group or a waitlist control group and completed self-report measures at baseline and postintervention assessments. Univariate latent change score models were computed to determine changes over time in adjustment processes and outcomes, with a multigroup-model approach to detect differences between the intervention and control groups and a 2-wave latent change score model to examine whether changes in processes were related to changes in outcomes.

Results: *Be a Mom* was found to be effective in reducing depressive (intervention group: μΔ=–3.35; *P*<.001 vs control group: μΔ=–1.48; *P*<.001) and anxiety symptoms (intervention group: μΔ=–2.24; *P*<.001 vs control group: μΔ=–0.43; *P*=.04) in comparison with the control group, where such changes were inexistent or much smaller. All 3 psychological processes under study improved statistically significantly in posttreatment assessments: emotion regulation ability ($\Delta\chi^2_3$=12.3; *P*=.007) and psychological flexibility ($\Delta\chi^2_3$=34.9; *P*<.001) improved only in the intervention group, and although self-compassion increased in both groups ($\Delta\chi^2_3$=65.6; *P*<.001), these improvements were considerably greater in the intervention group.

Conclusions: These results suggest that *Be a Mom*, a low-intensity cognitive behavioral therapy program, is a promising first-line intervention for helping perinatal women, particularly those with early-onset PPD symptoms.

Trial Registration: ClinicalTrials.gov NCT03024645; https://clinicaltrials.gov/ct2/show/NCT03024645

(JMIR Ment Health 2023;10:e39253) doi:10.2196/39253

KEYWORDS

Be a Mom; randomized controlled trial; postpartum depression; web-based interventions; cognitive behavioral therapy; prevention; mobile phone

Introduction

Background

The perinatal period is a time of physical, psychological, and interpersonal changes, representing a particularly vulnerable moment in a woman's life for the development of mental health problems [1]. During this challenging period, marked by positive and negative emotions that vary in frequency and intensity over time, women are at increased risk of developing mental health conditions of differing levels of severity, ranging from mild and transient emotional disturbance to more severe forms of psychopathology [2]. Postpartum depression (PPD), often concurrent with heightened anxiety, is a common and serious perinatal disorder with devastating and enduring effects on mothers' well-being, infants' health development, marital relationship, and overall family functioning [3].

Although PPD is treatable and amenable to preventive efforts, most women do not seek professional help for their perinatal depression symptoms [4], with recent research showing that attitudinal (eg, thinking that no one will be able to help), knowledge (eg, not knowing whether one's problems are a reason to ask for help) and structural barriers (eg, not having time or not being able to afford treatment) are the most common help-seeking barriers of perinatal women [5]. One increasingly popular approach of improving access to treatment is the use of web-based intervention programs, particularly low-intensity interventions. These interventions offer several advantages over traditional formats of assistance, including greater temporal and local independency, anonymity, accessibility, and flexible delivery [6], making them particularly recommended for perinatal women.

Other web-based interventions have been developed for PPD [7], but most of them were not designed or tested for preventive purposes (with the notable exceptions of *Mothers and Babies Course* [8] and *Sunnyside* [9], which nevertheless failed to provide compelling evidence on their efficacy), and they relied exclusively on classic cognitive behavioral therapy (CBT) principles [10,11], thus neglecting various concepts emphasized by *third wave* CBT (eg, acceptance and cognitive defusion), which were recently added to traditional CBT packages to foster intervention processes and outcomes [12]. *Third wave* CBT represents an evolution and extension of traditional CBT approaches by highlighting the role of changeable transdiagnostic psychological processes (ie, functionally important pathways of change) over the reduction or elimination of psychological and emotional symptoms. *Third wave* methods emphasized issues such as emotion regulation, acceptance, values, goals, and metacognition, and rather than focusing on the content of a person's thoughts and internal experiences, they focused on the context, processes, and functions of how a person relates to their internal experiences [13].

Be a Mom, a Web-Based Intervention to Prevent PPD

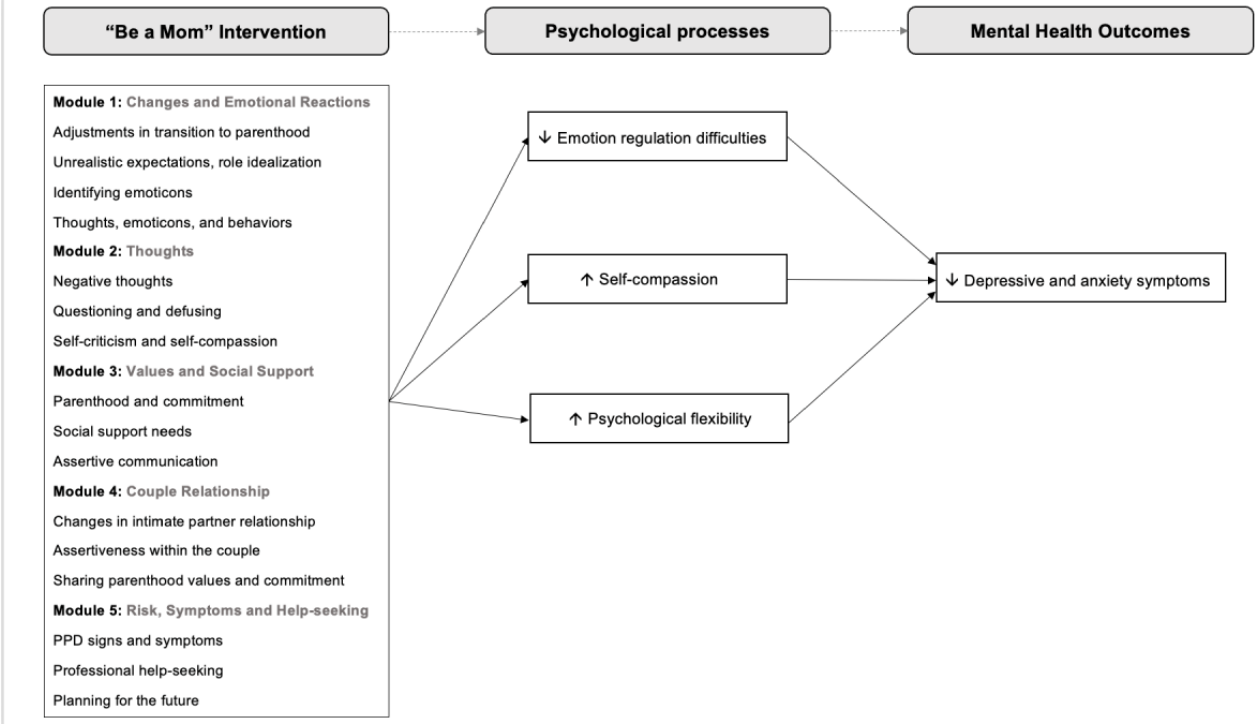
Be a Mom is a brief, self-guided, web-based selective or indicated preventive intervention that is primarily targeted at women who exhibit risk for PPD or present early-onset PPD symptoms. Grounded in CBT principles, it combines the classic CBT approach (eg, cognitive restructuring and activity scheduling) with more recent third wave CBT features (eg, values clarification and self-compassion). *Be a Mom* is a structured program with a modular approach that encompasses five sequential modules on the following topics: (1) changes and emotional reactions (changes and reorganizations during transition to parenthood; unrealistic expectations, role idealization, and the myths of perfect motherhood; and the cognitive-emotional-behavioral link); (2) cognitions (negative thoughts; reducing the power of thoughts: questioning and defusing; and self-criticism and self-compassion); (3) values and social support (values and commitment; social network: how to identify support needs and ask for help; and assertive communication: dealing with family and friends); (4) couple relationship (changes in the couple relationship during the postpartum period; assertive communication within the couple: negotiation and conflict resolution skills; and sharing parenthood values and commitments); and (5) PPD alert signs and professional help seeking (identify PPD signs and symptoms; professional help seeking: treatment options and how to seek help; and a continuing journey: planning for the future). The content of each module includes psychoeducational information combined with practical exercises and endorses the structured and goal-oriented nature of CBT sessions [14].

Preliminary evidence from a pilot randomized controlled trial (RCT) attested the feasibility of the program and its effectiveness in reducing depressive and anxiety symptoms of women at high risk for PPD [14]. Furthermore, a greater decrease in depressive symptoms was found to be associated with a greater decrease in emotion regulation difficulties and a greater increase in self-compassion, whereas psychological flexibility was found to be unrelated to changes in depression levels [15].

This Study

Although the mechanisms of change are the core of evidence-based interventions [12], an examination of these mechanisms within the context of web-based interventions for PPD had not been performed until recently [15]. Given the fact that interventions to prevent PPD are the most effective when they target women considered to be at risk [16], this RCT focused on a large sample of women at high risk for PPD to ascertain the mechanisms of change associated with the efficacy of *Be a Mom* in reducing PPD and postpartum anxiety symptoms because they often occur in comorbidity. Specifically, bearing in mind that depression and anxiety are distinct but highly concurrent clinical phenomena [3], the study sought to examine whether changes in psychological processes (ie, emotion regulation, self-compassion, and psychological flexibility) were associated with changes in depressive and anxiety symptoms among women who participated in the *Be a Mom* program (Figure 1).

Figure 1. Conceptual diagram of the study.



Methods

Procedure

This study was a 2-arm, open-label RCT to assess the efficacy of the *Be a Mom* intervention, in comparison with a waitlist control group receiving usual care, among women presenting high risk for PPD. The trial has been registered at ClinicalTrials.gov (NCT03024645). The eligibility criteria to participate in the study were as follows: (1) being an adult woman (aged ≥18 years) in the early postpartum period (up to 3 months post partum), (2) presenting high risk for PPD (a score of ≥5.5 on the Postpartum Depression Predictors Inventory-Revised [PDPI-R; refer to the *Measures* subsection]), (3) having a computer or tablet device or smartphone and internet access at home, (4) having the ability to read and speak Portuguese, and (5) being a resident of Portugal. The exclusion criterion was the presence of a severe medically diagnosed health condition in the infant or in the mother (eg, cerebral palsy in the infant or schizophrenia spectrum or other psychotic disorders in the mother), as self-reported by mothers. The women who did not meet the eligibility criteria were informed through email and were advised to seek professional help, if needed.

Recruitment occurred on the web between January 25, 2019, and January 30, 2021, both through unpaid cross-posting and paid advertisements on social media networks (Facebook and Instagram). Paid advertisements and campaigns targeted women aged 18 to 45 years with interests in maternity topics with the following tagline: "Did you have a baby in the last three months? We want to know if 'Be a Mom' is effective in promoting postpartum women's mental health, and you can help us! To know if you are eligible to participate in the study fill out the following form and we will contact you." Before receiving access to the eligibility form (including a set of questions to assess eligibility criteria and contact information), women were given information about the study goals and procedures (including voluntary participation and data protection issues), the participants' and researchers' roles were clarified, and women were asked to provide their informed consent to participate in the study (by clicking the option "I understand and accept the conditions of the study").

Eligible women who consented to participate in the study were administered baseline assessment (T0) using the web-based survey platform LimeSurvey (LimeSurvey GmbH). After completing the baseline assessment, eligible participants were randomly assigned using a computerized random number generator (allocation rate: 1:1) to the intervention group (*Be a Mom* intervention) or to the control group (waitlist group receiving usual care). According to the guidelines established by the pregnancy surveillance program in Portugal [17], usual postpartum care encompasses at least 1 medical appointment with a designated obstetrician or the woman's family physician, as well as weekly visits to nursing services in primary care to monitor the general health status of the mothers and their infants during the first 2 months post partum. One of the researchers was responsible for randomization, whereas the other 2 researchers were responsible for the participants' enrollment and group assignment. Participants were informed about their assigned group through email (no blinding to the assigned group) and instructed to seek no other intervention besides *Be a Mom* or usual care while participating in the RCT. Postintervention (T1) assessment was also performed on the web (through the LimeSurvey platform) in both intervention and control groups (after the intervention or 8 weeks after randomization). To reduce attrition, weekly reminders (email and SMS text messages on an alternate basis) were sent during 1 month to women who failed to complete the baseline and postintervention assessments.

The CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [18,19] (See Multimedia Appendix 1) was used for study reporting.

Intervention and Control Arms

Women assigned to the intervention arm were invited by email to register on a password-protected website that contained the *Be a Mom* intervention [20] (access to the program was restricted to the invited women). Approximately 2 weeks after registration, participants were contacted via telephone by the research team to assess whether there were any difficulties in accessing the website and to clarify any questions concerning the program's flow and functioning. After registration, the women had access to the 5 modules (the *Couple Relationship* module was only presented to women in a relationship), and they were instructed to complete 1 module per week in sequence, although a slower pace was allowed. Each module is approximately 45 minutes long. The women were given the option of pausing the module and resuming a session from the last page visited during subsequent access. Although it was generally recommended that they complete 1 module per week, given the usual competing demands of the postpartum period, they were allowed to complete the program within a maximum period of 8 weeks. Email reminders were sent automatically to the participants if they went 3, 7, and 13 days without accessing the program. The periodicity of the reminders was defined according to the guidelines provided by the software company that developed the web-based program and based on previous feedback obtained from postpartum women during the pilot RCT [14]. Of note, the sequence of periodical reminders restarted every time a woman returned to proceed with the program, that is, the reminders were unrelated to the amount of content that had already been completed. Asynchronous communication channels were provided for program-related support only. Access to the program was free of cost, and no compensation was offered to participants. Participants assigned to the control group were offered no intervention but were free to access usual care (as were all participants). At the end of the RCT (including 3-month and 12-month follow-up assessments), they were offered access to the *Be a Mom* intervention. Given that the first 3 to 4 months post partum seem to be a high-risk period for depression [21], the women in the waitlist control group were informed at the outset about their risk for PPD and instructed to interrupt their participation in the RCT and seek professional help within usual care if they felt that their health condition was worsening.

Measures

Sociodemographic and Clinical Information

A self-report form was used to gather information concerning the women's sociodemographic (eg, age, marital status, number of children, employment status, educational level, household monthly income, and residence), clinical (eg, risk score), and infant (eg, age and sex) data.

PPD Risk

The Portuguese version of the PDPI-R [22] was used to identify women presenting high risk for PPD. This self-report questionnaire includes 39 items assessing well-identified risk factors for PPD (eg, prenatal depression and low social support) answered on a *yes* versus *no* dichotomous scale (except for items assessing marital and socioeconomic status). The PDPI-R total score ranges from 0 to 39, with higher scores indicating an increased risk for PPD. In Portuguese validation studies, a score of ≥5.5 was indicative of higher risk for PPD [22].

Adjustment Outcomes

Depressive Symptoms

The Portuguese version of the Edinburgh Postpartum Depression Scale (EPDS) [23] was used to assess PPD symptoms. The EPDS comprises 10 items that assess symptoms of depression in terms of their presence and severity in the previous 7 days. The items were rated on a 4-point frequency scale (ranging from 0 to 3), with higher scores being indicative of more severe depressive symptoms. In this study, the Cronbach α values for the EPDS were ≥.86.

Anxiety Symptoms

The anxiety subscale of the Hospital Anxiety and Depression Scale (Portuguese version) was used to assess anxiety symptoms [24]. The anxiety subscale comprises 7 items, answered on a 4-point response scale (ranging from 0 to 3) to evaluate the presence of anxiety symptoms in the week before completion of the program. Higher scores were indicative of more severe anxiety symptoms. The Cronbach α values for the anxiety subscale were ≥.82.

Psychological Processes

Emotion Regulation

The Portuguese version of the Difficulties in Emotion Regulation Scale-Short Form [25] was used to assess women's emotion regulation abilities. This self-report questionnaire comprises 18 items developed to assess difficulties in using adaptive emotional regulation strategies, which were answered on a 5-point scale (ranging from 1=*almost never* to 5=*almost always*). The total score is used as a measure of emotion dysregulation [17]. Higher scores were suggestive of lower emotion regulation ability. The Cronbach α values for this scale were ≥.91.

Self-compassion

The Portuguese version of the Self-Compassion Scale-Short Form (SCS-SF) [26] was used to assess the women's levels of self-compassion. The SCS-SF is a self-report questionnaire with 12 items, answered on a 5-point scale (ranging from 1=*almost never* to 5=*almost always*). A total score may be computed, with higher scores being indicative of higher levels of self-compassion. In this study, the Cronbach α values for the SCS-SF were ≥.88.

Psychological Flexibility

The Portuguese version of the Comprehensive Assessment of Acceptance and Commitment Therapy Processes (CompACT) [27] was used in this study to measure psychological flexibility. The CompACT is a self-report questionnaire with 18 items organized into 3 subscales: openness to experience, behavioral awareness, and valued action. The items were answered on a

7-point response scale (ranging from 0=*strongly disagree* to 6=*strongly disagree*). It is possible to compute a total score, with higher scores indicating higher psychological flexibility. In this study, the Cronbach α values for the CompACT were ≥.86.

Data Analyses

SPSS software (version 22.0; IBM Corp) was used for descriptive and comparative analyses, and the Mplus program (version 7.0; Muthén & Muthén) was used for the examination of univariate latent change score (LCS) models. Descriptive statistics and comparison tests were computed for sample characterization and for comparing study completers and dropouts. Missed end points ranged from 32.4% (EPDS) to 40.2% (SCS-SF), and the pattern-of-missingness analysis suggested that data are missing completely at random (Little missing completely at random test: χ^2_{14}=9.7; *P*=.78). Missing data were handled using the full information maximum likelihood estimation [28].

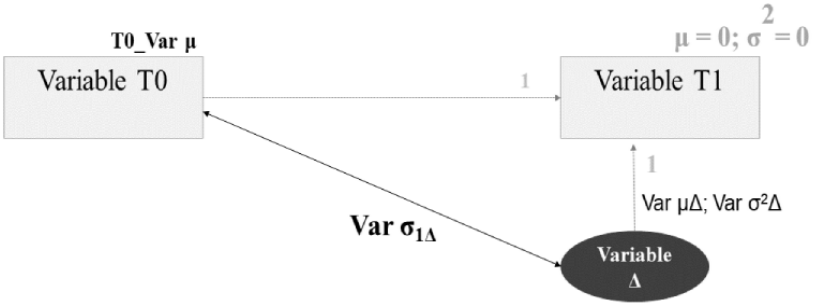
To examine changes over time both in the adjustment outcomes (depressive and anxiety symptoms) and in the psychological processes (emotion regulation, self-compassion, and psychological flexibility), univariate LCS models were computed [29]. LCS is a structural equation modeling approach in which a within-participants approach is considered, with change between 2 time points in 1 variable being modeled as a latent factor (defined as the part of the score of the variable at T1 that is not identical to the score of the variable at T0). LCS models allow the estimation of (1) the mean intercept of the change between T0 and T1 (μΔ, latent factor), (2) the variance or residual variance of the change between T0 and T1 ($\alpha^2\Delta$), (3) the covariance between the individual's score at T0 and the latent change factor (α1Δ), and (4) the mean scores at T0 (Figure 2). A statistically significant positive mean intercept of the change (μΔ) suggests that, on average, an individual's scores increased over time, whereas a statistically significant negative mean intercept suggests a decrease in the individual's scores from T0 to T1. Moreover, a statistically significant variance or residual variance in the LCS factor suggests heterogeneity across individuals regarding the averaged trajectory [29,30].

A multigroup-model approach was used to check for differences between the intervention and control groups in LCS models: the full constrained model (in which the LCS estimates were constrained to be equal across the groups) and the unconstrained model (in which the LCS estimates were free to vary across the groups) were compared concerning model fit indices and chi-square differences ($\Delta\chi^2$). Statistically significant chi-square changes ($\Delta\chi^2$) indicate that the LCS model is different across the groups. The overall model fit was ascertained using the chi-square goodness-of-fit statistic (significance at *P*>.05), the comparative fit index (CFI; ≥0.95), the root mean square of approximation (RMSEA; ≤0.08), and the standardized root mean square residual (SRMR; ≤0.08) [31,32]. In addition to the main analyses, multigroup-model comparison analyses (in which the full constrained model and the unconstrained models were compared) were repeated in 2 subgroups of women: evaluative respondents (ie, intention-to-treat analysis, by including women who completed both baseline and postintervention assessments in the statistical analyses) and women presenting clinically relevant depressive symptoms (ie, baseline EPDS scores of >9). Although *Be a Mom* was originally developed as a web-based CBT intervention to prevent PPD, the multigroup comparison analysis was conducted in this subgroup of women to ascertain the efficacy of *Be a Mom* in improving the mental health of those women who were already presenting clinical levels of depressive symptoms.

Finally, to examine whether changes in psychological processes (emotion regulation, self-compassion, and psychological flexibility) were associated with changes in women's adjustment outcomes (depressive and anxiety symptoms), a 2-wave LCS (2W-LCS) model was tested [30,33,34]. The 2W-LCS allows the estimation of the univariate LCS for each variable (ie, psychological processes: emotion regulation, self-compassion, and psychological flexibility; and adjustment outcomes: depressive and anxiety symptoms) and of the change-to-change effects (ie, the estimate of the effect of the change in one variable on the change of the other variable) of psychological processes (emotion regulation, self-compassion, and psychological flexibility) on adjustment outcomes. A positive and significative regression coefficient indicates that higher change scores in the antecedent variable are associated with higher change scores in the outcome variable, whereas a negative and significative regression coefficient indicates that higher change scores in the antecedent variable are associated with lower change scores in the outcome variable. As this study adopted a 2–time point longitudinal design, 2W-LCS was preferred over mediational analyses that typically require a 3–time point longitudinal design to be accurately performed. Moreover, this study sought to examine whether the proposed mechanisms of change differed in the intervention and control groups across time before analyzing their associations with changes in the levels of depressive and anxiety symptoms.

The 2W-LCS model allows the estimation of the correlations among individuals' scores at baseline (ie, correlations among emotion regulation, self-compassion, psychological flexibility, depressive symptoms, and anxiety symptoms) and the correlations among LCSs (ie, correlations among changes in emotion regulation, self-compassion, and psychological flexibility, as well as correlations between depressive and anxiety symptoms). In addition, the 2W-LCS model allows the estimation of the cross-lagged paths between the scores of psychological processes (emotion regulation, self-compassion, and psychological flexibility) at baseline and the change score of anxiety and depression symptoms, as well as that of the cross-lagged paths between the scores of depressive and anxiety symptoms at baseline and the change score of psychological processes. A multigroup-model approach was used to check for differences between the intervention and control groups in 2W-LCS models: the full constrained model (in which the 2W-LCS estimates were constrained to be equal across the groups) was compared with the unconstrained model concerning model fit indices and chi-square differences ($\Delta\chi^2$).

Figure 2. Univariate latent change score–model representation. Var: variable.



Ethics Approval

The study was approved by the ethics committee of the Faculty of Psychology and Educational Sciences, University of Coimbra (FPCE-CEDI-17.01.19).

Results

Participants

Of the 2367 women enrolled in the study between January 25, 2019, and January 30, 2021, a total of 1980 (83.65%) women were assessed concerning risk for PPD. Of these 1980 women, 1376 (69.49%) who presented high risk for PPD were administered the baseline assessment. Of the 1376 eligible women, 1053 (76.52%) completed baseline assessment and were randomized to the intervention (n=542, 51.47%) or control (n=511, 48.52%) conditions. The participant flowchart is presented in Figure 3. The baseline sociodemographic and clinical characteristics of the participants are presented in Table 1. No baseline differences were found between the groups, with the exception of parity: a higher number of women in the intervention group reported that this is their first child (347/542, 63.7%, vs 296/511, 57.7%, in the control group; *P*=.046). Moreover, of the 1053 participants, 341 (32.38%) did not complete the postintervention assessments. Of the 267 intervention group participants who completed the postintervention assessments, 129 (48.3%) completed the *Be a Mom* program, whereas 62 (23.2%) only partially completed the program (ie, completed 2 or 3 modules), and 76 (28.5%) did not complete the program (ie, completed only 1 module). Differences between study completers and dropouts in terms of baseline characteristics were found concerning marital status (χ^2_3=14.2; *P*=.003; a higher proportion of the dropouts were not married or cohabiting: completers, 61/712, 8.6%, vs dropouts, 73/341, 15.6%), educational level (χ^2_5=19.7; *P*=.001; a higher proportion of the completers had a higher education degree: completers, 315/483, 65.2%, vs dropouts, 124/250, 49.5%), infant age (t_{1051}=2.29; *P*=.02; women who dropped out from the study had older infants [mean age: 2.13, SD 0.98 years] than those who completed the study [mean age: 1.98, SD 0.94 years]), and level of risk (t_{1051}=3.88; *P*<.001; women who dropped out from the study presented higher risk for PPD [mean 12.14, SD 5.00] than those who completed the study [mean 10.99, SD 4.26]). Of note, despite the higher proportion of dropouts in the intervention group than in the control group, the sociodemographic and clinical characteristics of women who completed both assessments (*evaluative respondents*) are similar in both groups, with the exception of parity (*P*<.001). Variables that differ between the intervention and control groups and between the completers and dropouts were introduced as covariates in the analyses.

Figure 3. Participant flow diagram. PPD: postpartum depression.

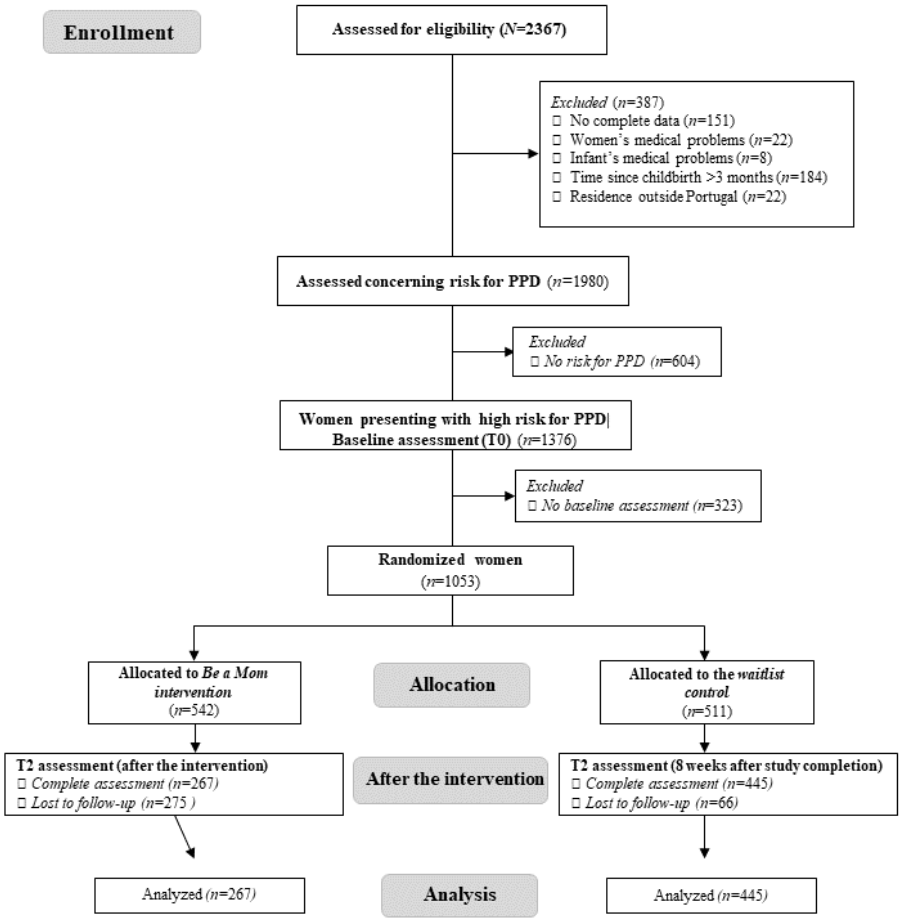


Table 1. Baseline sociodemographic and clinical characteristics: comparison between the control and intervention groups (N=1053).

	Control group (n=511)	Intervention group (n=542)	Chi-square (*df*)	*t* test (*df*)	*P* value
Women's sociodemographic characteristics					
Age (years), mean (SD)	32.81 (4.66)	32.91 (4.40)	N/A[a]	–0.35 (1051)	.73
Marital status (married or living together), n (%)	457 (89.4)	482 (88.9)	1.5 (3)	N/A	.69
Number of children (primiparous), n (%)	296 (57.9)	347 (64.0)	4.0 (1)	N/A	.046
Educational level (higher education), n (%)	214 (60.2)	225 (59.5)	3.5 (5)	N/A	.63
Professional status (employed), n (%)	427 (85.3)	465 (88.1)	6.3 (4)	N/A	.18
Monthly income (<€1000 [<US $1074]), n (%)	325 (63.6)	344 (63.4)	3.6 (4)	N/A	.46
Residence (urban), n (%)	404 (79.1)	434 (80.1)	0.2 (1)	N/A	.68
Women's baseline risk score (PDPI-R[b]), mean (SD)	11.44 (4.57)	11.28 (4.53)	N/A	0.57 (1051)	.57
Infant's characteristics					
Age (months), mean (SD)	1.99 (0.93)	2.06 (0.98)	N/A	–1.08 (1051)	.28
Sex (male), n (%)	263 (51.5)	260 (48)	2.8 (1)	N/A	.25

[a]N/A: not applicable.

[b]PDPI-R: Postpartum Depression Predictors Inventory-Revised.

Change From Baseline to Posttreatment Assessment in Adjustment Outcomes and Psychological Processes: Comparing the Intervention and Control Groups

Table 2 presents the descriptive statistics at the baseline and postintervention assessments for the study variables, as well as the LCS estimates for the intervention and control groups.

Concerning adjustment outcomes (depressive and anxiety symptoms), the full constrained model presented globally a poor fit to data (depressive symptoms: χ^2_4=41.4; *P*<.001; CFI=0.894; RMSEA=0.133, 90% CI 0.098-0.171; *P*<.001; SRMR=0.111; anxiety symptoms: χ^2_4=36.0; *P*<.001; CFI=0.906; RMSEA=0.123, 90% CI 0.088-0.162; *P*<.001; SRMR=0.097), whereas the unconstrained model presented a good fit (depressive symptoms: χ^2_1=5.8; *P*=.02; CFI=0.986; RMSEA=0.095, 90% CI 0.033-0.176; *P*=.11; SRMR=0.039; $\Delta\chi^2_3$=35.7; *P*<.001; anxiety symptoms: χ^2_1=6.0; *P*=.01; CFI=0.985; RMSEA=0.097, 90% CI 0.035-0.178; *P*=.10; SRMR=0.040; $\Delta\chi^2_3$=30.0; *P*<.001), supporting differences in LCSs among the groups. The LCS estimates (Table 2) suggest a statistically significant decrease in depressive symptoms from baseline to after the intervention in both groups. However, the decrease is 2.9-fold higher in the intervention group. Moreover, a statistically significant decrease in anxiety symptoms from T0 to T1 was found only in the intervention group. Changes were found to be heterogeneous across individuals in both the groups (Table 2). Furthermore, the covariance between the levels of anxiety and depressive symptoms at T0 and the amount of change over time were statistically significant (*P*<.001; Table 2), suggesting that the higher the depressive and anxiety symptoms at baseline, the greater the decrease in change scores. Subgroup analyses were performed considering 2 subgroups: evaluative respondents (women who completed both assessments) and women presenting an EPDS score of >9 at baseline. Concerning the evaluative respondents, the unconstrained model presented a better fit to the data than the constrained model for both depressive ($\Delta\chi^2_3$=179.0; *P*<.001) and anxiety symptoms ($\Delta\chi^2_3$=35.3; *P*<.001). The women in the intervention group reported a statistically significant higher decrease in depressive (intervention group: μΔ=–2.41; *P*<.001 vs control group: μΔ=–0.71; *P*<.001) and anxiety symptoms (intervention group: μΔ=–2.41; *P*<.001 vs control group: μΔ=–0.71; *P*=.74). Concerning the women presenting an EPDS score of >9 at baseline, the unconstrained model also presented a better fit to the data than the constrained model for depressive ($\Delta\chi^2_3$=88.9; *P*<.001) and anxiety symptoms ($\Delta\chi^2_3$=27.8; *P*<.001). Once again, the women in the intervention group reported a significantly higher decrease in depressive (intervention group: μΔ=–3.35; *P*<.001 vs control group: μΔ=–1.48; *P*<.001) and anxiety symptoms (intervention group: μΔ=–2.24; *P*<.001 vs control group: μΔ=–0.43; *P*=.04).

Concerning psychological processes (emotion regulation, self-compassion, and psychological flexibility) too, the unconstrained models presented a better model fit than the full constrained models, suggesting differences in LCS estimates among the groups. Specifically, for emotion regulation (full constrained model: χ^2_4=18.0; *P*<.001; CFI=0.965; RMSEA=0.082, 90% CI 0.046-0.122; *P*<.001; SRMR=0.072; unconstrained model: χ^2_1=5.7; *P*=.02; CFI=0.988; RMSEA=0.095, 90% CI 0.03-0.176; *P*=.11; SRMR=0.041; $\Delta\chi^2_3$=12.3; *P*=.007), the LCSs suggest a statistically significant decrease in emotion regulation difficulties from T0 to T1 in the intervention group but not in the control group. Concerning self-compassion (full constrained model: χ^2_4=68.1; *P*<.001; CFI=0.889; RMSEA=0.174, 90% CI 0.140-0.212; *P*<.001; SRMR=0.160; unconstrained model: χ^2_1=2.6; *P*=.11; CFI=0.997; RMSEA=0.054, 90% CI 0.000-0.141; *P*=.33; SRMR=0.027; $\Delta\chi^2_3$=65.6; *P*<.001), the LCS estimates suggest a statistically significant increase in the levels of self-compassion over time in both groups. However, the increase is 4.5-fold higher in the intervention group. Finally, concerning psychological flexibility (full constrained model: χ^2_4=41.1; *P*<.001; CFI=0.923; RMSEA=0.133, 90% CI 0.098-0.171; *P*<.001; SRMR=0.05; unconstrained model: χ^2_1=6.2; *P*=.01; CFI=0.989; RMSEA=0.099, 90% CI 0.360-0.180; *P*=.09; SRMR=0.041; $\Delta\chi^2_3$=34.9; *P*<.001), a statistically significant increase in the levels of psychological flexibility over time was found in the intervention group but not in the control group (Table 2). Changes were found to be heterogeneous across individuals in both groups. Furthermore, the covariance between the levels of emotion regulation, self-compassion, and psychological flexibility at T0 and amount of change over time were statistically significant (Table 2), suggesting that the higher the emotion regulation difficulties at baseline, the greater the decrease in emotion regulation difficulties over time, and the lower the self-compassion and psychological flexibility at baseline, the greater the increase in these variables over time. Subgroup analyses were also performed. Concerning the evaluative respondents, the unconstrained models presented a better fit to the data than the constrained models for all the variables (emotion regulation: $\Delta\chi^2_3$=8.5; *P*<.001; self-compassion: $\Delta\chi^2_3$=59.4; *P*<.001; and psychological flexibility: $\Delta\chi^2_3$=32.3; *P*<.001). The LCS estimates were different across the groups: in the intervention group, women reported a statistically significant decrease in emotion regulation difficulties (μΔ=–1.506; *P*=.047) and a statistically significant increase in self-compassion (μΔ=2.585; *P*<.001) and psychological flexibility (μΔ=2.490; *P*=.003), whereas no statistically significant changes were found in the control group (emotion regulation: μΔ=–0.689; *P*=.25; self-compassion: μΔ=0.537; *P*=.06; and psychological flexibility: μΔ=0.474; *P*=.32). Concerning the women presenting an EPDS score of >9 at baseline, a similar pattern was found with the unconstrained models presenting a better fit to the data for all the variables than the constrained models (emotion regulation: $\Delta\chi^2_3$=10.9; *P*<.001; self-compassion: $\Delta\chi^2_3$=44.4; *P*<.001; and psychological flexibility: $\Delta\chi^2_3$=25.4; *P*<.001). In addition, the women in the intervention group reported a statistically significant decrease in emotion regulation difficulties (μΔ=–3.748; *P*<.001) and a statistically significant increase in

self-compassion (μΔ=3.229; *P*<.001) and psychological flexibility (μΔ=4.880; *P*<.001), whereas the women in the control group reported no statistically significant change in emotion regulation (μΔ=–0.264; *P*=.70) or self-compassion (μΔ=0.499; *P*=.16) and a statistically significant but smaller change only in psychological flexibility (μΔ=1.158; *P*=.049). In fact, when we analyze the LCS estimates of the total sample and of the women presenting an EPDS score of >9 at baseline, we observe a higher change in psychological processes among the women presenting clinically relevant depressive symptoms who participate in the *Be a Mom* intervention. See Multimedia Appendix 2, Table S1, for exact *P* values.

Table 2. Changes in adjustment outcomes and psychological processes over time in the intervention and control groups: descriptive statistics and univariate latent change score (LCS) estimates.

	Descriptive statistics				Univariate LCS estimates[a]					
	Intervention group, mean (SD)		Control group, mean (SD)		Intervention group, B (SE)			Control group, B (SE)		
	T0[b] (n=542)	T1[c] (n=267)	T0 (n=511)	T1 (n=445)	μ_Δ[d]	σ^2_Δ[e]	σ1	μ_Δ	σ^2_Δ	σ1
EPDS[f]	10.99 (5.14)	8.75 (4.53)	11.73 (4.81)	10.78 (5.06)	–2.27 (0.24)***	20.11 (1.71)***	–12.97 (1.43)***	–0.79 (0.19)***	16.67 (1.12)***	–6.92 (0.99)***
HADS-A[g]	7.77 (4.13)	6.57 (3.93)	8.40 (4.21)	8.24 (4.21)	–1.36 (0.20)***	13.53 (1.16)***	–7.53 (0.95)***	0.00 (0.16)	12.61 (0.85)***	–.62 (0.77)***
DERS[h]	42.28 (13.29)	39.47 (13.07)	44.26 (13.50)	42.77 (13.87)	–2.87 (0.71)***	135.62 (12.59)***	–69.88 (10.26)***	–0.82 (0.48)	103.71 (7.22)***	–44.13 (7.15)***
SCS-SF[i]	35.80 (8.95)	34.93 (8.64)	38.17 (8.68)	35.65 (9.41)	2.90 (0.46)***	57.25 (5.21)***	–32.77 (4.25)***	0.65 (0.27)*	29.35 (2.06)***	–8.96 (2.31)***
CompACT[j]	58.56 (14.33)	61.30 (15.53)	56.36 (14.47)	57.25 (15.11)	2.82 (0.83)**	185.18 (16.58)***	–79.15 (12.04)***	0.37 (0.48)	103.41 (7.07)***	–45.51 (7.26)***

[a]The latent change score estimates presented correspond to the unconstrained model (where the latent change score parameters were free to vary across the groups, except for means at baseline, which were constrained to be equal across the groups). Unstandardized estimates are presented.

[b]T0: baseline assessment.

[c]T1: postintervention assessment.

[d]μ_Δ: mean intercept of the latent change factor.

[e]σ^2_Δ: variance of the latent change factor.

[f]EPDS: Edinburgh Postpartum Depression Scale (depressive symptoms).

[g]HADS-A: Hospital Anxiety and Depression Scale, anxiety subscale (anxiety symptoms).

[h]DERS: Difficulties in Emotion Regulation Scale (emotion regulation difficulties).

[i]SCS-SF: Self-Compassion Scale-Short Form (self-compassion).

[j]CompACT: Comprehensive Assessment of Acceptance and Commitment Therapy Processes (psychological flexibility).

*P<.05.

**P<.01.

***P<.001.

Change-to-Change Effects of Psychological Processes on Adjustment Outcomes

A 2W-LCS model was tested, in which we computed not only the univariate LCS for each variable but also the change-to-change effects of the psychological processes (emotion regulation, self-compassion, and psychological flexibility) on the adjustment outcomes (depressive and anxiety symptoms). A multigroup-model approach was used to test whether the path models were similar or different across the intervention and control groups. The first model to be tested was the fully constrained model (model A, in which the univariate LCSs, the cross-lagged effects, and the change-to-change effects were constrained to be equal across the groups). This model presented an acceptable fit to the data (χ^2_{103}=326.3; *P*<.001; CFI=0.946; RMSEA=0.064, 90% CI 0.056-0.072; *P*=.002; SRMR=0.078). The second model to be tested (model B) was the model in which the LCSs and the cross-lagged effects were constrained to be equal across the groups, but the change-to-change effects were free to vary across the groups. This model presented a slightly better fit to the data than the fully constrained model ($\Delta\chi^2_4$=3.7; *P*<.001; χ^2_{99}=322.7; *P*<.001; CFI=0.946; RMSEA=0.066, 90% CI 0.058-0.074; *P*=.001; SRMR=0.078), suggesting that change-to-change effects differ across the groups. The third model to be tested (model C) was the model in which the LCSs were constrained to be equal across the groups, but the cross-lagged effects and the change-to-change effects were free to vary across the groups. Compared with model B, model C presented a better fit to the data ($\Delta\chi^2_{12}$=21.2; *P*<.001; χ^2_{87}=301.5; *P*<.001; CFI=0.948; RMSEA=0.068, 90% CI 0.060-0.077; *P*=.001; SRMR=0.076). Finally, the unconstrained model (model D, in which the LCSs, cross-lagged effects, and change-to-change effects were free to vary across the groups) was tested and compared with model

C, presenting a statistically significantly better fit to the data ($\Delta\chi^2_5$=48.6; *P*<.001; χ^2_{82}=252.8; *P*<.001; CFI=0.96; RMSEA=0.063, 90% CI 0.054-0.072; *P*=.001; SRMR=0.072), which suggests that the univariate LCSs, the cross-lagged effects, and the change-to-change effects were different across the groups. Table 3 presents the standardized estimates of the different parameters for both the intervention and control groups.

Concerning correlations among the variables at baseline, statistically significant and large associations were found in both groups (Table 3). Of note, higher levels of emotion regulation difficulties and lower levels of self-compassion and psychological flexibility were statistically significantly associated with higher levels of depressive and anxiety symptoms at baseline in both groups.

Concerning cross-lagged effects, different patterns were found in the intervention and control groups. In the intervention group, higher levels of emotion regulation difficulties at baseline were associated with a higher change in depressive and anxiety symptoms, and higher levels of self-compassion and psychological flexibility at baseline were associated with a lower change in anxiety symptoms. Moreover, higher levels of depressive symptoms at baseline were associated with a significantly higher change in the levels of self-compassion and psychological flexibility. Conversely, in the control group, higher levels of emotion regulation difficulties at baseline were associated with a higher change in anxiety symptoms, and higher levels of psychological flexibility at baseline were associated with a higher change in depressive and anxiety symptoms. In addition, in the control group, higher levels of anxiety and depressive symptoms were associated with a lower change in the levels of emotion regulation over time (Table 3).

Concerning correlations among the change effects (univariate LCSs), statistically significant and moderate associations were found between changes in depressive symptoms and changes in anxiety symptoms in both groups (intervention group: B=0.377, SE 0.053; *P*<.001; control group: B=0.356, SE 0.042; *P*<.001), suggesting that the higher the decrease in depressive symptoms, the higher the decrease in anxiety symptoms. Moreover, statistically significant associations were found among changes in psychological processes in both groups, with a higher decrease in emotion regulation difficulties being associated with a higher increase in self-compassion and psychological flexibility and a higher increase in self-compassion being associated with a higher increase in psychological flexibility. However, the strength of the associations seems to be higher in the intervention group than in the control group (Table 3).

Finally, concerning change-to-change effects, a higher increase in self-compassion and psychological flexibility was statistically significantly associated with a higher reduction in depressive and anxiety symptoms in both the intervention and control groups. Moreover, a higher decrease in emotion regulation difficulties was associated with a higher reduction in anxiety symptoms in the intervention group and with a higher reduction in both depressive and anxiety symptoms in the control group (Table 3).

Table 3. Two-wave latent change score–model estimates: correlations at baseline, cross-lagged effects, and change-to-change effects in the intervention group and control group.

	Intervention group		Control group	
	B (SE)	*P* value	B (SE)	*P* value
Correlations between variables at baseline				
EPDS[a]$_{T0}$[b] ↔HADS-A[c]$_{T0}$	0.721 (0.021)	<.001	0.711 (0.022)	<.001
EPDS$_{T0}$ ↔ DERS[d]$_{T0}$	0.609 (0.026)	<.001	0.610 (0.027)	<.001
EPDS$_{T0}$ ↔ SCS-SF[e]$_{T0}$	–0.566 (0.028)	<.001	–0.565 (0.030)	<.001
EPDS$_{T0}$ ↔ CompACT[f]$_{T0}$	–0.594 (0.027)	<.001	–0.580 (0.029)	<.001
HADS-A$_{T0}$ ↔ DERS$_{T0}$	0.604 (0.027)	<.001	0.508 (0.033)	<.001
HADS-A$_{T0}$ ↔ SCS-SF$_{T0}$	–0.562 (0.030)	<.001	–0.506 (0.033)	<.001
HADS-A$_{T0}$ ↔ CompACT$_{T0}$	–0.574 (0.029)	<.001	–0.535 (0.031)	<.001
DERS$_{T0}$ ↔ SCS-SF$_{T0}$	–0.720 (0.021)	<.001	–0.714 (0.022)	<.001
DERS$_{T0}$ ↔ CompACT$_{T0}$	–0.651 (0.025)	<.001	–0.661 (0.025)	<.001
SCS-SF$_{T0}$ ↔ CompACT$_{T0}$	0.589 (0.028)	<.001	0.638 (0.027)	<.001
Cross-lagged effects between psychological processes at T0 and change in adjustment outcomes				
DERS$_{T0}$ → EPDS change[g]	0.087 (0.043)	.04	0.038 (0.040)	.34
SCS-SF$_{T0}$ → EPDS change	–0.042 (0.043)	.33	–0.061 (0.040)	.13
CompACT$_{T0}$ → EPDS change	–0.059 (0.044)	.18	–0.147 (0.038)	<.001
DERS$_{T0}$ → HADS-A change[h]	0.213 (0.037)	<.001	0.121 (0.034)	<.001
SCS-SF$_{T0}$ → HADS-A change	–0.162 (0.040)	<.001	–0.058 (0.037)	.12
CompACT$_{T0}$ → HADS-A change	–0.089 (0.041)	.03	–0.115 (0.035)	.001
Cross-lagged effects between adjustment outcomes at T0 and change in psychological processes				
EPDS$_{T0}$ → DERS change[i]	–0.064 (0.051)	.21	–0.115 (0.041)	.005
HADS-A$_{T0}$ → DERS change	–0.098 (0.054)	.07	–0.105 (0.045)	.02
EPDS$_{T0}$ → SCS-SF change[j]	0.101 (0.051)	.046	0.009 (0.043)	.84
HADS-A$_{T0}$ → SCS-SF change	0.057 (0.054)	.29	–0.002 (0.047)	.96
EPDS$_{T0}$ → CompACT change[k]	0.114 (0.051)	.02	0.056 (0.041)	.17
HADS-A$_{T0}$ → CompACT change	0.073 (0.054)	.17	0.061 (0.044)	.17
Correlations between changes in psychological processes				
DERS change ↔ SCS-SF change	–0.556 (0.044)	<001	–0.510 (0.037)	<.001
DERS change ↔ CompACT change	–0.490 (0.049)	<.001	–0.362 (0.043)	<.001
SCS-SF change ↔ CompACT change	0.492 (0.049)	<.001	0.303 (0.045)	<.001
Change-to-change effects				
DERS change → EPDS change	0.099 (0.065)	.13	0.255 (0.050)	<.001
SCS-SF change → EPDS change	–0.245 (0.065)	<.001	–0.101 (0.050)	.045
CompACT change → EPDS change	–0.333 (0.060)	<.001	–0.276 (0.044)	<.001
DERS change → HADS-A change	0.219 (0.062)	.001	0.198 (0.053)	<.001
SCS-SF change → HADS-A change	–0.172 (0.063)	.006	–0.107 (0.053)	.04
CompACT change → HADS-A change	–0.290 (–0.059)	<.001	–0.262 (0.046)	<.001

[a]EPDS: Edinburgh Postpartum Depression Scale (depressive symptoms).

[b]T0: baseline.

[c]HADS-A: Hospital Anxiety and Depression Scale, anxiety subscale (anxiety symptoms).

[d]DERS: Difficulties in Emotion Regulation Scale (emotion regulation difficulties).

[e]SCS-SF: Self-Compassion Scale-Short Form (self-compassion).

[f]CompACT: Comprehensive Assessment of Acceptance and Commitment Therapy Processes (psychological flexibility).

[g]EPDS change: change in depressive symptoms over time.

[h]HADS-A change: change in anxiety symptoms over time.

[i]DERS change: change in emotion regulation difficulties over time.

[j]SCS-SF change: change in self-compassion over time.

[k]CompACT change: change in psychological flexibility over time.

Discussion

Principal Findings

This trial investigated the mechanisms of change associated with the efficacy of *Be a Mom*, a web-based preventive intervention, in postpartum women at high risk for PPD. *Be a Mom* was found to be uniquely effective in reducing depressive and anxiety symptoms from baseline to after the intervention in comparison with the control group where such changes were inexistent or much smaller. This pattern of symptom reduction was similar when considering only the participants who completed both assessments and when considering women presenting clinically relevant symptoms of depression (ie, EPDS score of >9) at baseline. This finding consolidates and extends previous evidence for the efficacy of *Be a Mom* in targeting both clusters of symptoms [14]. Moreover, the fact that higher levels of psychopathological symptoms at baseline were related to greater posttreatment improvements suggests that this low-intensity program may be a promising first-line intervention for helping women considered to be at risk, particularly those with early-onset PPD symptoms.

In addition to the positive effects that *Be a Mom* revealed at the level of symptom reduction, the examination of concomitant psychological processes was conducted as a crucial step in producing a reliable account of the mechanisms involved in the change processes [12]. All the 3 psychological processes under study improved statistically significantly (with greater enhancements over time related to greater impairments at the baseline) in posttreatment assessments: emotion regulation ability and psychological flexibility improved only in the intervention group, and although self-compassion increased in both groups, these improvements were considerably greater in the intervention group. These psychological processes were nuclear processes that were expected to be promoted through *third wave* CBT approaches [13] and therefore in line with the therapeutic framework that supported the development of the *Be a Mom* program. Moreover, a higher increase in self-compassion and psychological flexibility was associated with greater overall symptom reduction, although a higher decrease in emotion regulation difficulties was related to a reduction in anxiety symptoms only. This pattern of associations is consistent with previous reports of difficulties in emotion regulation (as assessed herein) being distinctively related to anxiety symptoms, even when depressive symptoms are controlled [35].

In a systematic review of CBT interventions to prevent PPD, Werner et al [16] noted that 4 out of 6 studies reporting success in the alleviation of PPD targeted only women considered to be at risk (with sample sizes ranging from 27 to 241 participants). Nevertheless, none of these face-to-face studies examined the mechanisms of change underlying the observed reductions in depressive symptoms. The results from our study were based on a large sample of women at high risk for PPD and make important contributions to the field by ascertaining the links between enhanced self-regulatory skills and improved perinatal mental health outcomes in a web-based CBT individual intervention that was specifically devised to prevent PPD. It bears recalling that most web-based interventions for PPD were not specifically designed or tested for preventive purposes and that they relied exclusively on classic CBT [7,10,11]; moreover, 2 of the few internet-based CBT interventions for the prevention of PPD—*Mothers and Babies Course* and *Sunnyside*—also relied exclusively on traditional CBT, with the former resulting from an adaptation of a group-based face-to-face intervention [8] and the latter resulting from an adaptation of a group-based internet intervention [9]. In both cases, the programs encompassed 8 sessions, and the examination of the effects of the preventive interventions revealed that they failed to reach statistical significance in the respective RCTs [8,9].

Currently, the available internet-based interventions for PPD are mostly based on CBT, and even if the effect of CBT alone cannot be fully ascertained regarding the prevention of PPD, the overall trend gives cause for optimism, provided that the development of such programs is tailored to the needs and expectations of end users [36]. Taken together, this study contributes to the existing literature on CBT interventions to prevent PPD in 3 ways: first, it gathered compelling evidence for the efficacy of *Be a Mom* in reducing depressive and anxiety symptoms in a fairly large sample of women identified as being at high risk for PPD; second, it determined the role of self-regulatory psychological processes in the putative mechanisms of change linking participation in the CBT-based *Be a Mom* intervention with improved mental health outcomes; and third, it lends support to a preventive web-based intervention that was originally developed with the integration of classic and *third wave* CBT elements and methodically based on end users' appraisals [14,37].

Limitations

This RCT provided good evidence for the efficacy of *Be a Mom* and related mechanisms of change, from which causal and

functional accounts can be further developed. Nevertheless, a few limitations should be considered in this study. First, web-based recruitment was the preferred method for enrolling participants at the cost of excluding women with the lowest levels of digital literacy as well as those who, through lack of interest or opportunity, do not use social media at all. The recruitment strategy targeted women through specific advertisements, thus being prone to volunteer bias. Moreover, although the women were instructed to seek no other intervention besides *Be a Mom* or usual care while participating in the RCT, the access to other web-based resources was not monitored. Second, eligible women were assessed on the web regarding their risk for developing PPD, which may also increase the risk for self-selection bias (ie, women with greater interest in the study or experiencing higher distress may have been prone to participate, thus not ensuring the representativity of all women at risk for PPD in the resultant sample). Third, the dropout rate after the treatment (341/1053, 32.23%, in the total sample) was higher in the intervention group than in the control group. This could be because no compensation was offered in the study, and most women in the intervention group who dropped out from the *Be a Mom* program reported lack of time as the main reason for dropping out (suggesting that they may not have had the time to complete the assessment protocol). Despite this imbalance, both groups presented similar sociodemographic characteristics even when considering only women who completed both assessments, and the characteristics that distinguish completers and dropouts (in the total sample) were introduced as covariates in the analyses. Fourth and last, data were exclusively collected through self-report measures, which may be easily affected by social desirability factors (eg, the myths of *perfect motherhood*).

Conclusions and Future Directions

This study demonstrated the efficacy of *Be a Mom* in reducing the psychological maladjustment of postpartum women considered to be at risk for PPD through the enhancement of core psychological processes such as emotion regulation, psychological flexibility, and self-compassion. To provide further evidence of the efficacy of *Be a Mom* in preventing PPD, more follow-up assessments are needed (up to 12 months post partum), while controlling for the time between assessments in the respective analyses. According to the criteria suggested by the American Psychological Association (Division 12 Task Force on Psychological Interventions) for empirically validated treatments, the conduction of additional RCTs would be desirable to eventually ascertain *Be a Mom* as a well-established intervention. Moreover, *Be a Mom* was recently investigated as a universal preventive intervention, with, as in our study, promising results in the promotion of positive mental health among postpartum women with low risk for PPD [38].

Acknowledgments

The authors are particularly grateful to Maria Inês Caçador for her assistance in formatting the final version of the original manuscript. The authors also thank all women in this trial for their participation. This project was cofunded by the Center for Research in Neuropsychology and Cognitive and Behavioral Intervention–University of Coimbra (UIDB/PSI/00730/2020), the European Regional Development Fund through the Portugal 2020 program (CENTRO-01-0145-FEDER-028699), and the Portuguese Foundation for Science and Technology under the tutelage and superintendence of the Ministry of Science, Technology, and Higher Education through national funds (Central Government Development Expenditure and Investment Program).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors have made an intellectual contribution to this research trial. AF and MCC conceived the design and wrote the grant application. AF is the principal investigator of the trial, with overall responsibility for the project. MP and AAP assisted AF with data collection, curation, and formal analysis. AF and CC drafted the manuscript (original draft), and CC reviewed and edited the final draft. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1378 KB - mental_v10i1e39253_app1.pdf]

Multimedia Appendix 2
Table S1. Changes in adjustment outcomes and psychological processes over time in the intervention and control groups: descriptive statistics and univariate latent change score (LCS) estimates.
[DOCX File , 25 KB - mental_v10i1e39253_app2.docx]

References

1. Dagher RK, Bruckheim HE, Colpe LJ, Edwards E, White DB. Perinatal depression: challenges and opportunities. J Womens Health (Larchmt) 2021 Feb;30(2):154-159 [FREE Full text] [doi: 10.1089/jwh.2020.8862] [Medline: 33156730]
2. Wouk K, Gottfredson NC, Tucker C, Pence BW, Meltzer-Brody S, Zvara B, et al. Positive emotions during infant feeding and postpartum mental health. J Womens Health (Larchmt) 2019 Feb;28(2):194-202 [FREE Full text] [doi: 10.1089/jwh.2017.6889] [Medline: 30307779]
3. Smorti M, Ponti L, Pancetti F. A comprehensive analysis of post-partum depression risk factors: the role of socio-demographic, individual, relational, and delivery characteristics. Front Public Health 2019;7:295 [FREE Full text] [doi: 10.3389/fpubh.2019.00295] [Medline: 31709213]
4. Patel SR, Wisner KL. Decision making for depression treatment during pregnancy and the postpartum period. Depress Anxiety 2011 Jul 16;28(7):589-595 [FREE Full text] [doi: 10.1002/da.20844] [Medline: 21681871]
5. Fonseca A, Gorayeb R, Canavarro MC. Women s help-seeking behaviours for depressive symptoms during the perinatal period: socio-demographic and clinical correlates and perceived barriers to seeking professional help. Midwifery 2015 Dec;31(12):1177-1185. [doi: 10.1016/j.midw.2015.09.002] [Medline: 26433622]
6. Spanhel K, Balci S, Feldhahn F, Bengel J, Baumeister H, Sander LB. Cultural adaptation of internet- and mobile-based interventions for mental disorders: a systematic review. NPJ Digit Med 2021 Aug 25;4(1):128 [FREE Full text] [doi: 10.1038/s41746-021-00498-1] [Medline: 34433875]
7. Ashford MT, Olander EK, Ayers S. Computer- or web-based interventions for perinatal mental health: a systematic review. J Affect Disord 2016 Jun;197:134-146 [FREE Full text] [doi: 10.1016/j.jad.2016.02.057] [Medline: 26991368]
8. Barrera AZ, Wickham RE, Muñoz RF. Online prevention of postpartum depression for Spanish- and English-speaking pregnant women: a pilot randomized controlled trial. Internet Interv 2015 Sep 01;2(3):257-265 [FREE Full text] [doi: 10.1016/j.invent.2015.06.002] [Medline: 26273567]
9. Duffecy J, Grekin R, Long JD, Mills JA, O'Hara M. Randomized controlled trial of Sunnyside: individual versus group-based online interventions to prevent postpartum depression. J Affect Disord 2022 Aug 15;311:538-547. [doi: 10.1016/j.jad.2022.05.123] [Medline: 35654284]
10. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. MomMoodBooster web-based intervention for postpartum depression: feasibility trial results. J Med Internet Res 2013 Nov 04;15(11):e242 [FREE Full text] [doi: 10.2196/jmir.2876] [Medline: 24191345]
11. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol Med 2014 Jun;44(8):1675-1689 [FREE Full text] [doi: 10.1017/S0033291713002092] [Medline: 24148703]
12. Hofmann SG, Hayes SC. The future of intervention science: process-based therapy. Clin Psychol Sci 2019 Jan 29;7(1):37-50 [FREE Full text] [doi: 10.1177/2167702618772296] [Medline: 30713811]
13. Hayes SC, Hofmann SG. The third wave of cognitive behavioral therapy and the rise of process-based care. World Psychiatry 2017 Oct;16(3):245-246 [FREE Full text] [doi: 10.1002/wps.20442] [Medline: 28941087]
14. Fonseca A, Alves S, Monteiro F, Gorayeb R, Canavarro MC. Be a mom, a web-based intervention to prevent postpartum depression: results from a pilot randomized controlled trial. Behav Ther 2020 Jul;51(4):616-633. [doi: 10.1016/j.beth.2019.09.007] [Medline: 32586434]
15. Fonseca A, Monteiro F, Alves S, Gorayeb R, Canavarro MC. Be a mom, a web-based intervention to prevent postpartum depression: the enhancement of self-regulatory skills and its association with postpartum depressive symptoms. Front Psychol 2019;10:265 [FREE Full text] [doi: 10.3389/fpsyg.2019.00265] [Medline: 30873060]
16. Werner E, Miller M, Osborne LM, Kuzava S, Monk C. Preventing postpartum depression: review and recommendations. Arch Womens Ment Health 2015 Feb;18(1):41-60 [FREE Full text] [doi: 10.1007/s00737-014-0475-y] [Medline: 25422150]
17. Programa Nacional para a Vigilância da Gravidez de Baixo Risco. Lisboa: Gráfica Maiadouro; 2015.
18. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
19. Eysenbach G, CONSORT-eHEALTH Group. CONSORT-eHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
20. What is Be a Mom? Be a Mom. URL: https://beamom.pt/ [accessed 2023-03-06]
21. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol 2005 Nov;106(5 Pt 1):1071-1083. [doi: 10.1097/01.AOG.0000183597.31630.db] [Medline: 16260528]
22. Alves S, Fonseca A, Canavarro MC, Pereira M. Preliminary psychometric testing of the postpartum depression predictors inventory-revised (PDPI-R) in portuguese women. Matern Child Health J 2018 Apr 11;22(4):571-578. [doi: 10.1007/s10995-017-2426-5] [Medline: 29327321]
23. Areias ME, Kumar R, Barros H, Figueiredo E. Comparative incidence of depression in women and men, during pregnancy and after childbirth. Validation of the Edinburgh Postnatal Depression Scale in Portuguese mothers. Br J Psychiatry 1996 Jul;169(1):30-35. [doi: 10.1192/bjp.169.1.30] [Medline: 8818365]

24. Pais-Ribeiro J, Silva I, Ferreira T, Martins A, Meneses R, Baltar M. Validation study of a Portuguese version of the hospital anxiety and depression scale. Psychol Health Med 2007 Mar;12(2):225-35; quiz 235. [doi: 10.1080/13548500500524088] [Medline: 17365902]
25. Moreira H, Gouveia MJ, Canavarro MC. A bifactor analysis of the difficulties in emotion regulation scale - short form (DERS-SF) in a sample of adolescents and adults. Curr Psychol 2020 Jan 12;41(2):757-782. [doi: 10.1007/s12144-019-00602-5]
26. Castilho P, Pinto-Gouveia J, Duarte J. Evaluating the multifactor structure of the long and short versions of the self-compassion scale in a clinical sample. J Clin Psychol 2015 Sep 22;71(9):856-870. [doi: 10.1002/jclp.22187] [Medline: 25907562]
27. Trindade IA, Ferreira NB, Mendes AL, Ferreira C, Dawson D, Golijani-Moghaddam N. Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT): measure refinement and study of measurement invariance across Portuguese and UK samples. J Contextual Behavioral Sci 2021 Jul;21:30-36. [doi: 10.1016/j.jcbs.2021.05.002]
28. Enders CK. Applied Missing Data Analysis. New York: Guilford Publications; 2022.
29. McArdle JJ. Latent variable modeling of differences and changes with longitudinal data. Annu Rev Psychol 2009 Jan 01;60(1):577-605. [doi: 10.1146/annurev.psych.60.110707.163612] [Medline: 18817479]
30. Henk CM, Castro-Schilo L. Preliminary detection of relations among dynamic processes with two-occasion data. Structural Equation Model Multidisciplin J 2015 Jul 28;23(2):180-193. [doi: 10.1080/10705511.2015.1030022]
31. Hu L, Bentler PM. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. Psychological Method 1998 Dec;3(4):424-453. [doi: 10.1037/1082-989X.3.4.424]
32. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Model Multidisciplin J 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
33. Kievit RA, Brandmaier AM, Ziegler G, van Harmelen AL, de Mooij SM, Moutoussis M, NSPN Consortium, et al. Developmental cognitive neuroscience using latent change score models: a tutorial and applications. Dev Cogn Neurosci 2018 Oct;33:99-117 [FREE Full text] [doi: 10.1016/j.dcn.2017.11.007] [Medline: 29325701]
34. Valente NJ, MacKinnon DP. Comparing models of change to estimate the mediated effect in the pretest-posttest control group design. Struct Equ Modeling 2017;24(3):428-450 [FREE Full text] [doi: 10.1080/10705511.2016.1274657] [Medline: 28845097]
35. Cisler JM, Olatunji BO, Feldner MT, Forsyth JP. Emotion regulation and the anxiety disorders: an integrative review. J Psychopathol Behav Assess 2010 Mar;32(1):68-82 [FREE Full text] [doi: 10.1007/s10862-009-9161-1] [Medline: 20622981]
36. Mu TY, Li YH, Xu RX, Chen J, Wang YY, Shen CZ. Internet-based interventions for postpartum depression: a systematic review and meta-analysis. Nurs Open 2021 May;8(3):1125-1134 [FREE Full text] [doi: 10.1002/nop2.724] [Medline: 33373101]
37. Fonseca A, Pereira M, Araújo-Pedrosa A, Gorayeb R, Ramos MM, Canavarro MC. Be a mom: formative evaluation of a web-based psychological intervention to prevent postpartum depression. Cognit Behavioral Pract 2018 Nov;25(4):473-495. [doi: 10.1016/j.cbpra.2018.02.002]
38. Monteiro F, Pereira M, Canavarro MC, Fonseca A. Be a Mom's efficacy in enhancing positive mental health among postpartum women presenting low risk for postpartum depression: results from a pilot randomized trial. Int J Environ Res Public Health 2020 Jun 29;17(13):4679 [FREE Full text] [doi: 10.3390/ijerph17134679] [Medline: 32610640]

Abbreviations

2W-LCS: 2-wave latent change score
CBT: cognitive behavioral therapy
CFI: comparative fit index
CompACT: Comprehensive Assessment of Acceptance and Commitment Therapy Processes
CONSORT-eHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
EPDS: Edinburgh Postpartum Depression Scale
LCS: latent change score
PDPI-R: Postpartum Depression Predictors Inventory-Revised
PPD: postpartum depression
RCT: randomized controlled trial
RMSEA: root mean square of approximation
SCS-SF: Self-Compassion Scale-Short Form
SRMR: standardized root mean square residual

Edited by J Torous; submitted 04.05.22; peer-reviewed by K Uludag, S Chan, D Tandon; comments to author 25.11.22; revised version received 16.12.22; accepted 03.01.23; published 17.03.23.

Please cite as:
Carona C, Pereira M, Araújo-Pedrosa A, Canavarro MC, Fonseca A
The Efficacy of Be a Mom, a Web-Based Intervention to Prevent Postpartum Depression: Examining Mechanisms of Change in a Randomized Controlled Trial
JMIR Ment Health 2023;10:e39253
URL: https://mental.jmir.org/2023/1/e39253
doi:10.2196/39253
PMID:36930182

Original Paper

Prediction of Mental Health Problem Using Annual Student Health Survey: Machine Learning Approach

Ayako Baba[1*], PhD; Kyosuke Bunji[2*], PhD

[1]Health Service Center, Kanazawa University, Ishikawa, Japan

[2]Graduate School of Business Administration, Kobe University, Hyogo, Japan

[*]all authors contributed equally

Corresponding Author:
Ayako Baba, PhD
Health Service Center
Kanazawa University
Kakuma-machi, Kanazawa-city
Ishikawa, 9201192
Japan
Phone: 81 762645254
Email: a_baba@staff.kanazawa-u.ac.jp

Abstract

Background: One of the reasons why students go to counseling is being called on based on self-reported health survey results. However, there is no concordant standard for such calls.

Objective: This study aims to develop a machine learning (ML) model to predict students' mental health problems in 1 year and the following year using the health survey's content and answering time (response time, response time stamp, and answer date).

Methods: Data were obtained from the responses of 3561 (62.58%) of 5690 undergraduate students from University A in Japan (a national university) who completed the health survey in 2020 and 2021. We performed 2 analyses; in analysis 1, a mental health problem in 2020 was predicted from demographics, answers for the health survey, and answering time in the same year, and in analysis 2, a mental health problem in 2021 was predicted from the same input variables as in analysis 1. We compared the results from different ML models, such as logistic regression, elastic net, random forest, XGBoost, and LightGBM. The results with and without *answering time* conditions were compared using the adopted model.

Results: On the basis of the comparison of the models, we adopted the LightGBM model. In this model, both analyses and conditions achieved adequate performance (eg, Matthews correlation coefficient [MCC] of with *answering time* condition in analysis 1 was 0.970 and MCC of without *answering time* condition in analysis 1 was 0.976; MCC of with *answering time* condition in analysis 2 was 0.986 and that of without *answering time* condition in analysis 2 was 0.971). In both analyses and in both conditions, the response to the questions about campus life (eg, anxiety and future) had the highest impact (Gain 0.131-0.216; Shapley additive explanations 0.018-0.028). Shapley additive explanations of 5 to 6 input variables from questions about campus life were included in the top 10. In contrast to our expectation, the inclusion of answering time–related variables did not exhibit substantial improvement in the prediction of students' mental health problems. However, certain variables generated based on the answering time are apparently helpful in improving the prediction and affecting the prediction probability.

Conclusions: These results demonstrate the possibility of predicting mental health across years using health survey data. Demographic and behavioral data, including answering time, were effective as well as self-rating items. This model demonstrates the possibility of synergistically using the characteristics of health surveys and advantages of ML. These findings can improve health survey items and calling criteria.

(JMIR Ment Health 2023;10:e42420) doi:10.2196/42420

KEYWORDS

student counseling; health survey; machine learning; mental health problem; response time

Introduction

Background

In Japan, 57% of female and 48% of male students seek counseling voluntarily [1]. Others seek counseling after being referred by teachers, families, friends, or health service centers. Health service centers use health surveys as clues for counseling calls. A health survey is mainly conducted as a self-reported screening test or an interview along with health checkups [2].

The style of mental health surveys differs among universities. Participants in the survey range from first-year students to undergraduate students to all students, including graduate students. The content is also diverse, with the University Personality Inventory (UPI) [3] and General Health Questionnaire (GHQ) [4,5] being frequently used [6]. Similarly, the criteria for calling using a mental health survey are inconsistent [7]. It is challenging to devise uniform criteria that will aid in efficient calling. To address this issue, we propose that machine learning (ML) would aid in developing a universal method for detecting and predicting mental health problems among students.

ML has several advantages in modeling highly individual phenomena with many variables, such as allowing the simultaneous testing of numerous input variables and their complex interactions [8], permitting nonlinearity in producing predictive algorithms [9], and having the ability to test all possible relationships to identify the superlative algorithm and model without a priori hypothesis by researchers [8]. These advantages of ML can help detect and predict mental health problems.

Related Work

Many trials detect mental health problems using biological, behavioral, and subjective data [10,11]. Some studies have diagnosed suicidal ideation or behavior using various predictors, such as psychological problems and self-injuries [12,13]. To predict mental health problems, the following broader variables were tested: birth information; physical illness; environmental factors; behavioral data (eg, number of naps, study duration, and use of a cell phone); biological data (eg, sleep onset time, skin conductance, and temperature) collected with wearable sensors, and text data posted on social media [14-17].

There are also some examples of ML with on-campus data, known as learning analytics and educational data mining [18]. In some studies, repeaters and dropouts have been predicted using log data from a learning management system or e-learning, information about entrance examinations, registered courses, attendance, grades, and submission status of assignments [19-24].

On-campus mental health surveys are insufficiently used as ML predictors. We consider them suitable for statistical analysis for 3 reasons. First, the mental health survey is open to all students, and the response rate is high. They must complete it to receive health checkups. Second, mental health surveys are conducted annually at many universities. Thus, it is possible to observe a change over time. Third, web-based mental health surveys, which have become a standard under COVID-19, can collect well-organized self-reported data and *answering time* data.

We use the word *answering time* as a broad concept, including response time (RT), which is the amount of time in which a student responds to the question item after it is shown on the screen; RT stamp (the time of day when a student accesses the survey form); and answer date (the number of days elapsed in which a student completes the survey after it was announced). In particular, RT has recently attracted attention mainly in psychological assessment and educational testing; however, little is known about how to use RT in mental health measurement [25,26]. A few studies have shown that RT is related to depression estimation and the prediction of suicide ideation and attempt status [25,27]. RT would be worth investigating for predicting mental health based on the response behavior.

Goal of This Study

This study aimed to predict students' mental health problems even without a specific mental health score. In this study, we conducted 2 analyses. In analysis 1, we attempted to predict mental health problems from other items when responding to the survey. Analysis 2 attempted to predict mental health problems in 2021 from the responses in the survey conducted in 2020 (1 year later). We also tested the effect of answering time on the prediction of mental health problems.

The main objective of this study was to build an exploratory predictive model. We also attempted to evaluate the impact of each input variable on the outcome based on ML and examine the interpretable relationship between input variables and mental health problems.

This study demonstrates the possibility of using data collected from health surveys conducted at most universities to precisely detect students experiencing mental health problems and reduce call costs. This will also pave the way to using existing health survey data even without mental health scores.

Methods

Data Collection

The data set was obtained from a closed health survey conducted at the beginning of 2020 and 2021 school years at University A, a national university in Japan. According to the email announcement of the annual health checkup, all students were asked to answer the survey on the learning management system. Although incentives were not offered and penalties were not levied, the students were required to answer this survey to obtain a health certification. However, the survey did not comprise adaptive questions. There were 28 nonrandomized questions (health survey in Table 1) that were presented one per page without completeness check. Students could review their answers before submission and resubmit them before the deadline. When there were duplicate submissions from identical IDs, the last submission was recorded.

Table 1. List of items in the health survey.

Item	Options
Individual data	
Faculty and department	24 categories
Sex	Male and female
International student	Yes and no
Age	(Free description)
Years of university	1, 2, 3, 4, 5, and 6
Health survey	
Case history	Yes, no, and do not know
Circulatory problem	
Digestive problem	
Cranial nerve and mental disorders	
Respiratory problem	
Allergic disease	
Measles	
Rubella	
Mumps	
Varicella	
Other diseases	
Currently under medical treatment	Yes and no
Health concerns or worries	Yes and no
Physical or mental disability	Yes and no
Physical or mental disability certificate	Yes and no
Needs for consultation or support from the university	Yes, no, and do not know
Blood type	A, B, O, AB, and do not know
Tuberculin skin test result	Negative, positive, and do not know
Vaccination	1 time, ≥2 times, never, and do not know
Measles	
Rubella	
Mumps	
Varicella	
Have meals	Every day, sometimes, and never
Breakfast	
Lunch	
Dinner	
Sleeping hours	15 categories (1-hour increments; from <3 to >17 hours)
Exercise	Every day, sometimes, and never
Drinking	Every day, sometimes, and never
Smoking	Yes and no
Thoughts on quitting smoking	Want, do not want, and nonsmoker
Residence	Home, apartment, student dormitory, and others
Commuting method	Walking, bus, train and bus, bicycle, motorbike, car, and others
Club activities	Yes, used to be, and never

Item	Options
Hours of internet use	20 categories (1-hour increments; from <0.5 to >19 hours)
Nine questions about campus life	Yes and no
This is not the university or major that I wanted.	
I have a strong anxiety about my campus life.	
It takes me a long time to get used to new surroundings.	
I am worried about my relationships with other people.	
I am worried about my future.	
I had some trouble with someone close to me.	
I feel like my life rhythm has been upset recently.	
I am worried that I may not be able to take credit as I think.	
I always feel busy with my academic work.	
Kessler 6 (mental health scale) [28]	4=all of the time, 3=most of the time, 2=some of the time, 1=a little of the time, and 0=none of the time
Have lost >3 kg in a month	Yes and no
Presently	
In the past (>18 years old)	
In the past (<18 years old)	
What you want to tell us and questions	Free description
Menstrual trouble (for women only)	Yes and no
Coping with menstrual trouble (for women only)	Seeing a physician at a hospital, want to consult Health Service Center, and would like to wait and see how it goes
Menstrual cycle (for women only)	Regular every 26-32 days, not every 26-32 days but comes regularly, not regular but more than once every 3 months, and not more than once every 3 months
Menstrual pain (for women only)	No pain, no need to take painkillers, need to take painkillers, and painkillers do not alleviate my pain
Birth-control pills or oral contraceptives (for women only)	Yes and no
Taking	
To improve menstrual cycle	
To improve menstrual pain	
For other reasons	

In this study, data that met the following conditions were excluded: during the 2020 or 2021 survey, the participant was, not an undergraduate, did not answer by the deadline (27 days in 2020 and 39 days in 2021), did not respond to any items, did not respond to all items on the Kessler 6 mental health scale [29] (it was used as an outcome and detailed in Outcome and Input Variables section), was a nonregular student (eg, credited auditors), and did not give consent for data use. Incomplete submissions or atypical response duration were not excluded.

In 2020 and 2021, a total of 5690 students may have been affiliated with the university based on student enrollment conditions in 2020. In both the 2020 and 2021 surveys, 3680 students answered the survey. In total, 119 (3.23%) students' data were excluded: 10 (8.4%) students did not consent, 105 (88.2%) students did not complete Kessler 6 scale in 2020 or 2021 survey, and 4 (3.3%) students belonged to old affiliations. The resulting data set used for the analysis contains 3561 students (Table 2).

The mean age was 20.54 (SD 1.75) years. The distribution of students by sex was 1941 (54.51%) male and 1620 (45.49%) female students. The data set included 1234 (34.65%) first-year students, 1005 (28.22%) second-year students, 1018 (28.59%) third-year students, and 304 (8.54%) fourth- or higher-year students. The distribution of students by faculty was 1361 (38.22%) in humanities and social sciences (6 departments); 943 (26.48%) in science and engineering (8 departments); 1175 (33%) in medical, pharmaceutical, and health (4 departments); and 82 (2.3%) in others (2 departments). Furthermore, 10 (0.28%) students were international students.

Table 2. Demographic characteristics of participants in the 2020 school year (N=3561).

Variables	Values
Age (years), mean (SD)	20.54 (1.75)
Sex, n (%)	
Male	1941 (54.51)
Female	1620 (45.49)
Years of university, n (%)	
First-year students	1234 (34.65)
Second-year students	1005 (28.22)
Third-year students	1018 (28.59)
Fourth-year or higher-year students	304 (8.54)
Faculty, n (%)	
Humanities and social sciences (6 departments)	1361 (38.22)
Science and engineering (8 departments)	943 (26.48)
Medical, pharmaceutical, and health (4 departments)	1175 (33)
Others (2 departments)	82 (2.3)
International students, n (%)	10 (0.28)

Ethics Approval and Informed Consent

This study received ethics approval from the Medical Ethics Committee of Kanazawa University (test number:2011-113 [017]). All students were asked to complete the survey; however, they had the option to not consent for its use for research purposes. We obtained written informed consent from all participants. Data were anonymized and saved offline. The usability and technical functionality were tested by staff members in advance. The research was performed according to relevant guidelines and regulations.

Outcome and Input Variables

Our model did not use the diagnosis but used the Japanese version of the Kessler 6 [28] as an indicator of mental health problems. Diagnoses of the students were not necessarily shared with the university, whereas Kessler 6 could be obtained from the health survey conducted in the university. The Japanese version of the Kessler 6 demonstrates high performance in detecting mood or anxiety disorders according to the Diagnostic and Statistical Manual of Mental Disorders Fourth Edition [28,30]. Although diagnoses are not made based only on Kessler 6 scores and need consultations with physicians, they have been used widely for screening and surveys [31]. Analyses 1 and 2 used the Kessler 6 scores for 2020 and 2021, respectively, as outcomes. Both analyses divided scores into a dichotomous variable (1: problem and 0: no problem) according to a ≥13 cutoff point [29,30].

In this study, we examined several candidate ML models to determine which model to use. All candidate models are probabilistic prediction models: logistic regression, elastic net, random forest, XGBoost, and LightGBM. Elastic net is a logistic regression model with L1 and L2 regularization. Random forest combines independent decision trees and outputs the most popular class as a prediction [32]. XGBoost and LightGBM are implementations of the gradient boosting decision tree (GBDT) [33,34]. The GBDT is known to achieve high performance in many ML tasks, such as predicting diseases [35] or identifying psychological health risk factors [36].

We used the same input variables for analyses 1 and 2. All the responses to the 2020 questionnaire listed in Table 1 were coded differently according to the model. For tree-based models (random forest, XGBoost, and LightGBM), they were coded numerically, with missing values filled with −99. For example, the answer for *blood type* item, which was one among A, B, O, AB, do not know, and no response, was coded as 1, 2, 3, 4, 5, and −99, respectively. For the regression models (logistic regression and elastic net), different coding schemes were applied based on the type of input variables. Continuous or ordinal variables such as *hours of internet use* or *age* were numerically coded in the same manner as tree-based models, except that missing or nonresponse data were filled with the grand mean. In contrast, effect coding was applied to nominal variables such as *case history* or *blood type*. Missing data were treated as independent categories. As a result, the response to *blood type* (Q7), for instance, was transformed into 5 input variables: ANS7_B, ANS7_O, ANS7_AB, ANS7_DN (do not know), and ANS7_NA (missing). When the respondent answered “B,” ANS7_B was coded as 1 and the remaining 4 were coded as 0.

Because the aim of this study was to investigate which type of input variables were useful in the prediction of mental health problems, we used all variables except RT without dimensionality reduction. Notably, the tree-based models are less susceptible to the inclusion of irrelevant variables. In addition, an elastic net can reduce the number of input variables because of regularization.

In addition to the raw response variables in Table 1, we generated several input variables with the expectation that some

patterns would appear on each variable, as shown in Table 3. These conjectures are based on existing studies on students' mental health and suicide [37], shared knowledge among experts, or authors' clinical experiences at the university.

To deal with the content of answers to the free description question (*What you want to tell us and questions*), the answers were analyzed using KH Coder [38,39], a free text mining analysis software. First, words that appeared ≥10 times were clustered using the Ward method, and Jaccard distances between clusters were calculated. Six clusters were adopted for interpretation. Second, each free description was labeled as 0 or 1 for each cluster. If words in a cluster did not appear in a free description, the description was labeled as 0 for the cluster; if words contained in the cluster appeared even once in a free description, the description was labeled as 1 for the cluster. Cluster 1 contained, among other words, "counseling," "current," and "under treatment." Cluster 2 contained, among other words, "syndrome," "surgery," and "diagnosis." Cluster 3 contained, among other words, "menstruation" and "anemia." Cluster 4 contained, among other words, "left," "ear," and "right." Cluster 5 contained, among other words, "depression," "mental," "stress," and "anxiety." Cluster 6 contained, among other words, "hospital," "allergy," and "skin."

This study examined the effectiveness of answering time in predicting mental health problems. Table 4 lists the answering time–related input variables.

The 2 coordinates Xs and Ys represent the position of the time stamp in a 24-hour clock and are introduced to indicate that 0:00 and 24:00 are identical. They were calculated by

$$X_s = \cos([s / 43200]\pi),\ Y_s = \sin([s / 43200]\pi) \quad \textbf{(1)}$$

where *s* is the number of seconds elapsed in a day; for instance, when the time stamp is 13:34:42, *s* will be 13 × 3600 + 34 × 60 + 42 = 48882. RT-related variables were included as the first and second principal component scores because the RT of every single item is difficult to interpret, and these RTs can moderately affect the prediction.

Except for the text-mining analysis, the abovementioned feature engineering processes were executed using the R programming language [40] on a Windows 10 machine, with the missMDA package [41] for principal components analysis with missing data.

Table 3. Generated variables and possible conjecture.

Item	Conjecture
The academic year of admission (the values in the data were between 2012 and 2020)	There may be a cohort effect.
An indicator to show how many years after age 18 (standard in Japan) the student was enrolled (2750 students were never retained , while 811 were retained)	Some enrollments over the standard age may have reflected or caused some problems including mental health problems.
An indicator to show how many years the respondent had ever been retained (3527 students were never retained, while 34 were retained)	Some students who repeated school years may have had some mental health problems.
The number of unanswered items	Students with mental health problems may hesitate to answer or miss many items.
The number of characters in answer to the free description question, *what you want to tell us and questions*	Students with mental health problems may write a lot in free descriptions to make the university aware of their difficulty.
Dummy variables for including 6 specific word clusters[a]	Some students with mental health problems may have used specific words in common and expressions in the free description.
Number of "yes" answers in *case history* and *nine questions about campus life*	Students with mental health problems may have experienced some diseases or campus life problems.
The proportion of "yes" answers to each item in *nine questions about campus life* in the respondent's department	Some students who belonged to departments with many students with campus life problems may have felt stress.
The number of students and the proportion of same-sex students in the respondent's department	Some students who belonged to the departments which have few same-sex students may have had difficulty in communication and relationships.

[a]Cluster 1 contained, among other words, "counseling," "current," and "under treatment." Cluster 2 contained, among other words, "syndrome," "surgery," and "diagnosis." Cluster 3 contained, among other words, "menstruation" and "anemia." Cluster 4 contained, among other words, "left," "ear," and "right." Cluster 5 contained, among other words, "depression," "mental," "stress," and "anxiety." Cluster 6 contained, among other words, "hospital," "allergy," and "skin."

Table 4. Generated variables based on answering time and possible conjecture.

Item	Conjecture
Answer date (the number of days elapsed between the survey announcement and the completion of survey by a student)	Students with mental health problems (eg, procrastination traits and, lack of information gathering) may answer the survey later.
RT[a] per character in the free description question *what you want to tell us and questions*	Some students who wrote smooth free descriptions may have had some health problem to tell the university.
2D-dimensional coordinates of response time stamps[b]	Response time stamps may have reflected students' life rhythm and sleep quality, which may be related to mental health state.
Kernel density of *s*	Some students who answered at different times compared with others may have had problems with life rhythm and sleep quality, which may be related to mental health problems.
2 principal components scores of log-transformed standardized RTs of all items	Students with mental health problems may spend a longer or shorter time on specific items than other students.
Individual mean and SD of standardized log RTs calculated using all valid responses	Students with mental health problems may spend, on average, a longer or shorter time answering. Students with mental health problems may spend very different time on different items.
2 principal components scores of differences between the standardized log RTs of all items and the individual mean	Students with mental health problems may spend longer or shorter time on answering specific items than others.
2 principal components scores of kernel density of the standardized log RTs of all items	Students with mental health problems may show peculiar RT patterns.

[a]RT: response time.

[b]They were calculated with equation 1.

Procedure

In this procedure, we did not split the data into training and test data. This is because the values for the obtained metrics varied considerably depending on the data split owing to data imbalance. Thus, only 5.5% (196/3561) of the students in 2020 and 6.88% (245/3561) in 2021 had a higher risk of mental illness problems as assessed by Kessler 6, including 2.33% (83/3561) of the students who had problems both in 2020 and 2021. Typically, such imbalanced data are dealt with by applying sample weight to equal the total number of positive and negative observations. However, our preliminary comparison performed better when we did not impose sample weights on all candidate models. Therefore, we did not use any particular procedures for dealing with imbalanced data.

First, we compared the performance of candidate ML models to determine which model to use. The procedure was divided into 2 parts: parameter tuning and performance evaluation. Both parts were conducted based on the following procedure (*K* repeats of 5-fold cross-validation; Figure 1 shows the schema of the procedure):

1. Randomly shuffle the entire data set and split it into 5 folds so that the positive:negative ratio of the outcome variable is nearly the same among the data sets.
2. Choose 1 of the 5 data sets as the test data, and train the model using the remaining 4 data sets.
3. Evaluate the trained model on the test data.
4. Repeat steps 2 and 3 for each data set.
5. Calculate the means of 5 data sets on each evaluation metric.
6. Repeat steps 1 through 5 *K* times.
7. Calculate the means of *K* repeats on each evaluation metric.

In the parameter tuning part, the model's hyperparameters were optimized based on the tree-structured Parzen estimator [42].

The log-loss, as a model performance measure, in a specific parameter set was measured by *K*=50 repeats of 5-fold cross-validation. Log-loss at fold *f* of the *k*-th repeat is calculated by

$$\text{LogLoss}(k,f) = -(1/n)\Sigma([y_i]\log[p_i] + [1 - y_i]\log[1 - p_i]) \quad \textbf{(2)}$$

where y_i and p_i are the realized value and predicted probability of respondent *i*, respectively, and *n* is the number of observations in the fold *f*. As a result, the log-loss for a specific parameter set was obtained by

$$\text{LogLoss} = (1/K)\Sigma(1/5)\Sigma\text{LogLoss}(k,f) \quad \textbf{(3)}$$

All the hyperparameters tuned in this study are summarized in Table 5 [43-46]. These parameters can cope with overfitting if they are appropriately tuned. In addition, the learning rate parameters in the GBDT models (eta in XGBoost and learning rate in LightGBM) were fixed at 0.01. Generally, the smaller the learning rate, the more accurate is the prediction, although the computation time increases.

As a result, the performance in each parameter set was measured as the average of *K*=50 log-losses. Finally, the parameter values that minimize the objective performance measure were adopted, and we proceeded to the performance evaluation. By using the optimized parameter set, we obtained 8 model performance measures (log-loss, Brier loss, area under the curve of receiver operating characteristic curve [AUC-ROC], area under the curve of precision-recall curve [AUC-PR], specificity, precision, recall [also known as sensitivity], and Matthews correlation coefficient [MCC]). Log-loss, Brier loss, AUC-ROC, and AUC-PR are measures based on predictive probability (p_i). In contrast, the specificity, precision, recall, and MCC are obtained from the confusion matrix. For example, MCC is calculated as follows:

$$(TP \times TN - FP \times FN) / \sqrt{([TP + FP] \times [TP + FN] \times [TN + FP] \times [TN + FN])} \quad \textbf{(4)}$$

where *TP*, *TN*, *FP*, and *FN* are the numbers of true-positive, true-negative, false-positive, and false-negative respondents, respectively. As previously mentioned, this study used imbalanced data. Therefore, we included the AUC-PR and MCC, which are commonly used in such cases [47].

On the basis of the performance measures obtained, we selected the best model. To test the extent to which answering time–related variables contribute to prediction performance in the selected model, we also conducted parameter tuning and performance evaluation phases without all answering time–related variables (shown in Table 4). We call the analysis without answering time–related variables "*without* condition, whereas the analysis with all input variables is called *with* condition.

Furthermore, we calculated feature importance measures (Gain and Shapley additive explanations [SHAP]) on both *with* and *without* conditions. SHAP is based on the Shapley values, the average marginal contribution in game theory. With the firm theoretical basis of game theory, SHAP can explain the extent to which each input variable contributes to the model's prediction [48]. Although several approaches exist for interpreting the prediction, we used SHAP because it is easy to implement and understand. Using the SHAP dependence plot, we can understand the relationship between the actual input value and its effect on the predicted probability. The abovementioned procedures were executed using all the data in K=100 repeats of 5-fold cross-validation. After model evaluation, we obtained SHAP dependence plots on several variables to interpret the model prediction.

All the procedures described in this section were executed using the Julia programming language [49] with several libraries: MLJ.jl [50], TreeParzen.jl [51], ScikitLearn.jl [52], LightGBM.jl [53], and ShapML.jl [54]. Compared with several famous languages, such as R and Python, Julia is still a young language and is rapidly developing. However, it is already perceived as sufficiently stable to be used for research purposes [55,56].

Figure 1. Flowchart of 5-fold cross-validation in each part.

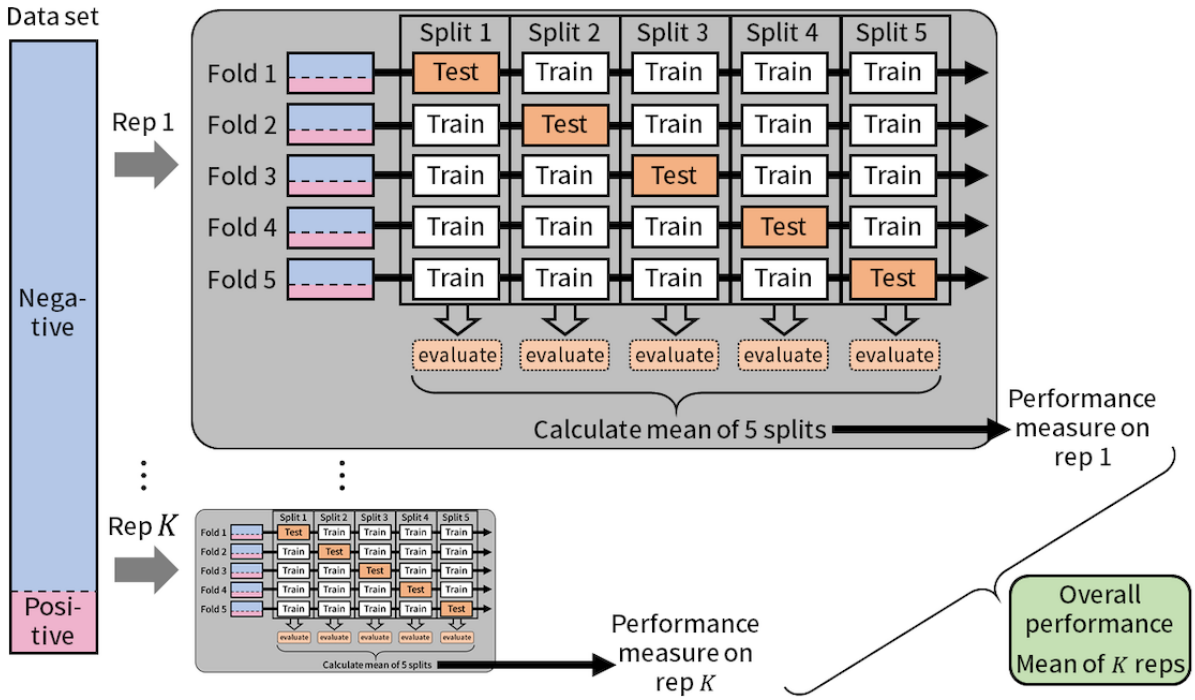


Table 5. Hyperparameters of each model tuned in the parameter tuning part.

Model and hyperparameter	Description
Elastic net	
gamma	Strength of the L1 regularization
lambda	Strength of the L2 regularization
Random forest	
n_estimators	The number of trees in the forest
max_depth	The maximum depth of the tree
min_samples_leaf	The minimum number of samples required to be at a leaf node.
max_features	The proportion of features used in each tree
max_samples	The proportion of samples used in each tree
XGBoost	
num_round	The number of boosting iterations
max_depth	The maximum depth of the tree
min_child_weight	Minimum sum of instance weight needed in a child
alpha	Strength of the L1 regularization
lambda	Strength of the L2 regularization
colsample_bytree	The proportion of features used in each tree
subsample	The proportion of samples used in each tree
LightGBM	
num_iterations	The number of boosting iterations
num_leaves	The maximum number of leaves in 1 tree
max_depth	The maximum depth of the tree
min_deta_in_leaf	The minimal amount of data in 1 leaf
lambda_l1	Strength of the L1 regularization
lambda_l2	Strength of the L2 regularization
feature_fraction	The proportion of features used in each tree
bagging_fraction	The proportion of samples used in each tree

Results

Tables 6 and 7 summarize mean and SD of each model's performance measures. As explained in *Procedure* section, the resultant values shown in Tables 6 and 7 are the means and SDs calculated from 100 trials. Italicized variables indicate the best model in terms of each measure. Evidently, the elastic net and LightGBM performed best. Regarding probabilistic measures (Table 6), elastic net afforded slightly better values than LightGBM. By contrast, binary measures, which were calculated from the confusion matrix (Table 7), were apparently better at LightGBM. Figure 2 shows the confusion matrix obtained from the elastic net and LightGBM in analyses 1 and 2. The actual values given below the percentages in each cell are the averages of K=100 repeats. In analysis 1, LightGBM misclassified only 11.27 (SD 1.75) observations (5.02, SD 1.58 false positives and 6.25, SD 1.18 false negatives) out of 3561 on average, whereas elastic net misclassified more than twice (average 25.94, SD 2.49) as many observations as LightGBM. Similar results were obtained in analysis 2. On the basis of these outcomes, we report the results for LightGBM.

Table 8 presents a comparison of the performances of LightGBM measures between *with* and *without* conditions. In contrast to our expectations, these results were mixed. This implies that the use of answering time–related variables does not necessarily improve the prediction.

Tables 9-12 show the top 10 input variables ranked by Gain and SHAP importance. SHAP importance was calculated as the average absolute value of all respondents. The number of "yes" answers to nine questions about campus life had the highest impact regardless of the analyses, conditions, and feature importance measures (Tables 9-12). In Table 9 (Gain in with condition), 5 of the top 10 input variables are answering time–related variables. In Table 10 (Gain in without condition), hours of internet use, years of university, and proportion of the same sex in the department were ranked common to both analyses 1 and 2, in addition to variables related to nine questions about campus life. In Table 11 (SHAP in with

condition), input variables from nine questions about campus life and answering time–related input variables mainly occupied the top 10 variables in analysis 1. In analysis 2, answering time–related input variables were out of rank, whereas years of university, the academic year of admission, diet-related variables, and proportion of the same sex in the department were ranked. In Table 12 (SHAP in without condition), years of university, academic year of admission, and diet-related variables were ranked common to both analyses 1 and 2 in addition to variables related to nine questions about campus life.

Table 6. Probabilistic performance measures of each model.

	Log-loss, mean (SD)	Brier loss, mean (SD)	AUC-ROC[a], mean (SD)	AUC-PR[b], mean (SD)
Analysis 1				
Logistic regression	0.252 (0.012)	0.109 (0.003)	0.760 (0.012)	0.215 (0.013)
Elastic net	*0.164 (0.001)* [c]	*0.090 (0.001)*	*0.862 (0.003)*	*0.292 (0.010)*
Random forest	0.210 (0.008)	0.113 (0.001)	0.833 (0.005)	0.218 (0.010)
XGBoost	0.165 (0.001)	0.091 (0.001)	0.855 (0.003)	0.278 (0.009)
LightGBM	0.165 (0.001)	*0.090 (0.001)*	0.857 (0.003)	0.281 (0.010)
Analysis 2				
Logistic regression	0.342 (0.015)	0.140 (0.003)	0.696 (0.009)	0.178 (0.011)
Elastic net	*0.211 (0.001)*	*0.115 (0.001)*	*0.796 (0.004)*	*0.264 (0.008)*
Random forest	0.250 (0.002)	0.140 (0.001)	0.768 (0.005)	0.200 (0.007)
XGBoost	0.213 (0.001)	0.116 (0.001)	0.791 (0.004)	0.249 (0.008)
LightGBM	0.213 (0.001)	0.116 (0.001)	0.789 (0.004)	0.246 (0.009)

[a]AUC-ROC: area under the curve of receiver operating characteristic curve.

[b]AUC-PR: area under the curve of precision-recall curve.

[c]Italicized variables indicate the best model in terms of each measure.

Table 7. Performance measures on each model from confusion matrix.

	Specificity, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	MCC[a], mean (SD)
Analysis 1				
Logistic regression	0.977 (0.002)	0.681 (0.020)	0.823 (0.015)	0.732 (0.016)
Elastic net	0.996 (0.001)	0.934 (0.009)	0.936 (0.009)	0.931 (0.007)
Random forest	0.976 (0.002)	0.681 (0.022)	0.859 (0.018)	0.749 (0.017)
XGBoost	0.998 (0.001)	0.963 (0.008)	0.963 (0.006)	0.961 (0.005)
LightGBM	*0.999 (0.000)* [b]	*0.975 (0.008)*	*0.968 (0.006)*	*0.970 (0.005)*
Analysis 2				
Logistic regression	0.976 (0.002)	0.730 (0.018)	0.884 (0.012)	0.787 (0.013)
Elastic net	0.998 (0.001)	0.970 (0.006)	0.963 (0.005)	0.964 (0.004)
Random forest	0.967 (0.002)	0.662 (0.017)	0.857 (0.015)	0.732 (0.014)
XGBoost	0.998 (0.000)	0.977 (0.005)	0.987 (0.004)	0.981 (0.004)
LightGBM	*0.999 (0.000)*	*0.984 (0.005)*	*0.990 (0.004)*	*0.986 (0.003)*

[a]MCC: Matthews correlation coefficient.

[b]Italicized variables indicate the best model in terms of each measure.

Figure 2. Confusion matrices on the elastic net (left half) and LightGBM (right half).

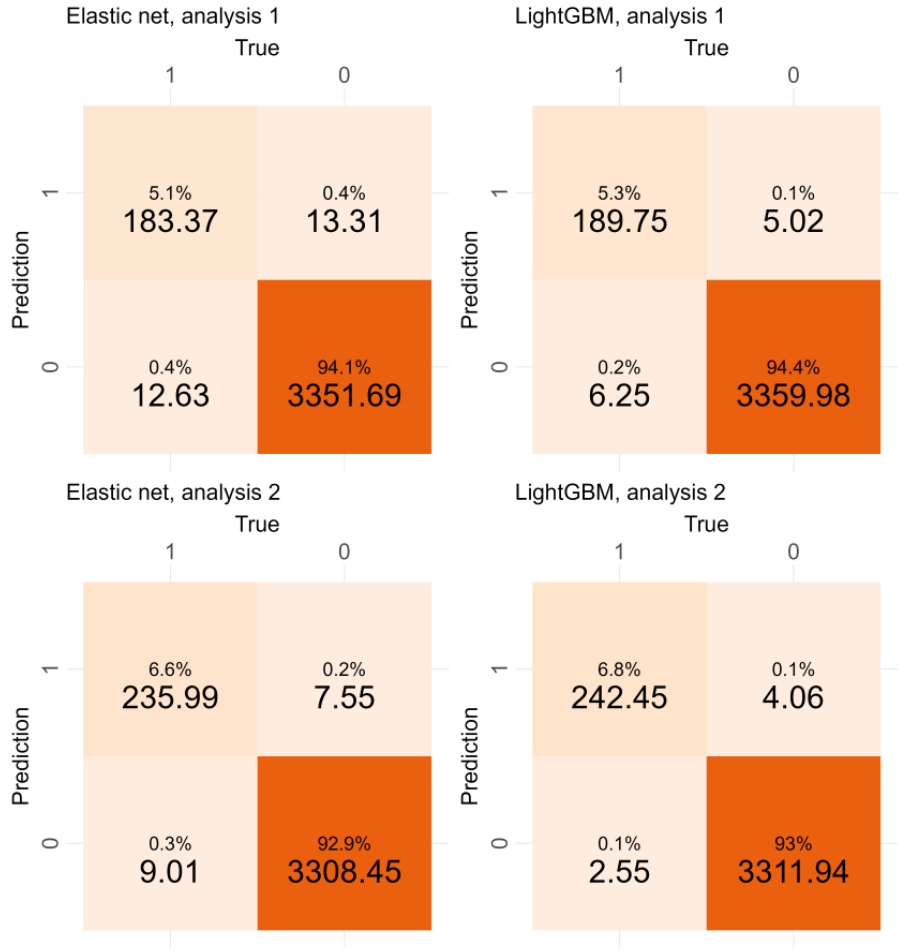


Table 8. Performance measures of with and without conditions on LightGBM.

Measure	Analysis 1		Analysis 2	
	With, mean (SD)	*Without*, mean (SD)	*With*, mean (SD)	*Without*, mean (SD)
Log-loss	*0.165 (0.001)* [a]	*0.165 (0.001)*	0.213 (0.001)	*0.211 (0.001)*
Brier loss	*0.090 (0.001)*	*0.090 (0.001)*	0.116 (0.001)	*0.115 (0.001)*
AUC-ROC[b]	0.857 (0.003)	*0.858 (0.003)*	0.789 (0.004)	*0.798 (0.004)*
AUC-PR[c]	*0.281 (0.010)*	0.276 (0.010)	0.246 (0.009)	*0.260 (0.010)*
Specificity	*0.999 (0.000)*	0.998 (0.000)	*0.999 (0.000)*	0.998 (0.001)
Precision	*0.975 (0.008)*	0.974 (0.008)	*0.984 (0.005)*	0.970 (0.007)
Recall	0.968 (0.006)	*0.981 (0.007)*	*0.990 (0.004)*	0.977 (0.005)
MCC[d]	0.970 (0.005)	*0.976 (0.005)*	*0.986 (0.003)*	0.971 (0.004)

[a]Italicized variables indicate the best model in terms of each measure.

[b]AUC-ROC: area under the curve of receiver operating characteristic curve.

[c]AUC-PR: area under the curve of precision-recall curve.

[d]MCC: Matthews correlation coefficient.

Table 9. Top 10 input variables ranked by Gain (with answering time condition).

Input variable	Gain
Analysis 1	
Nine questions about campus life (number of "yes")	0.216
"I have a strong anxiety about my campus life."	0.088
"I'm worried about my future."	0.033
2nd principal component score of $Diff_{RT}$[a]	0.031
2nd principal component score of $Dens_{RT}$[b]	0.030
x coordinate of response time stamp	0.026
Mean of "I had some trouble with someone close to me" in department	0.025
Individual mean of standardized log RT[c]	0.021
1st principal component score of $Diff_{RT}$	0.021
"I always feel busy with my academic work."	0.020
Analysis 2	
Nine questions about campus life (number of "yes")	0.131
"I have a strong anxiety about my campus life."	0.080
y coordinate of response time stamp	0.029
2nd principal component score of $Diff_{RT}$	0.028
Years of university	0.026
Individual SD of $Dens_{RT}$	0.024
1st principal component score of standardized log RT	0.024
"I'm worried about my future."	0.023
Individual SD of standardized log RT	0.023
Kernel density of *s*	0.023

[a]$Diff_{RT}$ is the difference between the standardized log RT of each item and individual mean.

[b]$Dens_{RT}$ is the kernel density of the standardized log RT of each item.

[c]RT: response time.

Table 10. Top 10 input variables ranked by Gain (without answering time condition).

Input variable	Gain
Analysis 1	
Nine questions about campus life (number of “yes”)	0.200
“I have a strong anxiety about my campus life.”	0.089
Mean of “I had some trouble with someone close to me.” in department	0.035
sleeping hours	0.032
“I'm worried about my future.”	0.029
Faculty and department	0.024
years of university	0.023
“I always feel busy with my academic work.”	0.023
Proportion of the same sex in department	0.023
Hours of internet use	0.022
Analysis 2	
Nine questions about campus life (number of “yes”)	0.143
“I have a strong anxiety about my campus life.”	0.119
“I'm worried about my future.”	0.047
“I always feel busy with my academic work.”	0.031
Proportion of the same sex in department	0.031
Years of university	0.030
“It takes me a long time to get used to new surroundings.”	0.030
The academic year of admission	0.028
Hours of internet use	0.027
Mean of “I have a strong anxiety about my campus life” in department	0.023

Table 11. Top 10 input variables ranked by Shapley additive explanations (SHAP; with answering time condition).

Input variable	SHAP
Analysis 1	
Nine questions about campus life (number of "yes")	0.027
"I have a strong anxiety about my campus life."	0.015
"I'm worried about my future."	0.006
Mean of "I had some trouble with someone close to me." in department	0.005
2nd principal component score of $Dens_{RT}$[a]	0.005
"I always feel busy with my academic work."	0.005
2nd principal component score of $Diff_{RT}$[b]	0.005
Years of university	0.004
"It takes me a long time to get used to new surroundings."	0.004
x coordinate of response time stamp	0.003
Analysis 2	
Nine questions about campus life (number of "yes")	0.019
"I have a strong anxiety about my campus life."	0.016
Years of university	0.006
"I'm worried about my future."	0.006
The academic year of admission	0.005
"It takes me a long time to get used to new surroundings."	0.005
Have meals (breakfast)	0.005
"I always feel busy with my academic work."	0.005
Have lost over 3 kg in a month (presently)	0.004
Proportion of the same sex in department	0.004

[a]$Dens_{RT}$ is the kernel density of the standardized log RT of each item.

[b]$Diff_{RT}$ is the difference between the standardized log RT of each item and individual mean.

Table 12. Top 10 input variables ranked by Shapley additive explanations (SHAP; without answering time condition).

Input variable	SHAP
Analysis 1	
Nine questions about campus life (number of "yes")	0.028
"I have a strong anxiety about my campus life."	0.015
Mean of "I had some trouble with someone close to me." in department	0.006
"I'm worried about my future."	0.006
Years of university	0.005
"I always feel busy with my academic work."	0.005
The academic year of admission	0.003
"It takes me a long time to get used to new surroundings."	0.003
Sleeping hours	0.003
Have lost over 3 kg in a month	0.003
Analysis 2	
"I have a strong anxiety about my campus life."	0.018
Nine questions about campus life (number of "yes")	0.016
"I'm worried about my future."	0.008
The academic year of admission	0.007
"I always feel busy with my academic work."	0.007
Years of university	0.007
"It takes me a long time to get used to new surroundings."	0.007
Have meals (breakfast)	0.006
Have lost over 3 kg in a month (presently)	0.005
Age	0.004

Figures 3-8 show SHAP dependence plots or violin plots of some interpretable input variables with high feature importance on average: the number of "yes" answers to *nine questions about campus life*, sleeping hours, years of university, proportion of the same sex in respondent's department, xy coordinates of RT stamp, and answer date. The blue lines indicate smoothed conditional mean based on the generalized additive model. Figure 3 shows that the averages of the SHAP effect were positive for ≥5 "yes" answers in 9 questions in common to both analyses 1 and 2, indicating that many campus life problems increased predictive probability. Figure 4 shows that the averages of the SHAP effect were positive for <6 hours or >8 hours of sleep, indicating that too short or too long sleep increased predictive probability. Figure 5 shows that the averages of the SHAP effect were positive for third- or higher-year students. Focusing on second-year students, the averages of the SHAP effect were positive in analysis 1 and negative in analysis 2, indicating that predictive probability decreased. Figure 6 shows that the averages of the SHAP effect were positive for fewer proportion of the same sex in the department, indicating high predictive probability. All the top 5 departments with a low proportion of the same sex were in the faculty of science and engineering; female students were 2.8% (4/141) to 14.4% (18/125). The proportion of men in the international department was only 17.1% (32/187).

Figure 7 shows the SHAP dependence plot on xy coordinates of the RT stamp. Both axes represent the realized values of the input. The correspondence between the RT stamp and the coordinates is shown in the graph. For example, if an answer was recorded at 00:00 (midnight), its RT stamp was transformed to (x,y)=(1,0). Red dots were plotted in conjunction with the sum of the SHAP values on the x and x coordinates. Their dots move outside the thick black circle ($x^2 + y^2 = 1$) if the sum is positive. As a result, if a respondent's answer was recorded at 00:00 (midnight) and the sum of SHAP values was 0.02, a red dot according to the respondent was plotted at (x,y)=(1,0)*0.02*10=(1.2,0). Note that 10 is the magnification ratio determined arbitrarily for visibility. The thin black lines are circles of x^2+y^2=0.8,0.9,1.1,1.2, indicating SHAP values of −0.02, −0.01, 0.01, and 0.02. The figure suggests that respondents whose answers were recorded at midnight tended to show large SHAP values, indicating a high predictive probability. By contrast, those whose answers were recorded in the morning (approximately 7:00-11:00) or afternoon (approximately 13:00-17:00) were plotted inside the thick black circle, indicating negative SHAP values. Figure 8 suggests that an ascending relationship exists between the answer date and SHAP. This indicates that students who responded to the survey more than 9 to 10 days after the announcement were more likely to be predicted as having mental health problems.

Figure 3. Shapley additive explanations (SHAP) dependence plot of the number of "yes" answers to 9 questions about campus life on "with" condition (left: analysis 1; right: analysis 2).

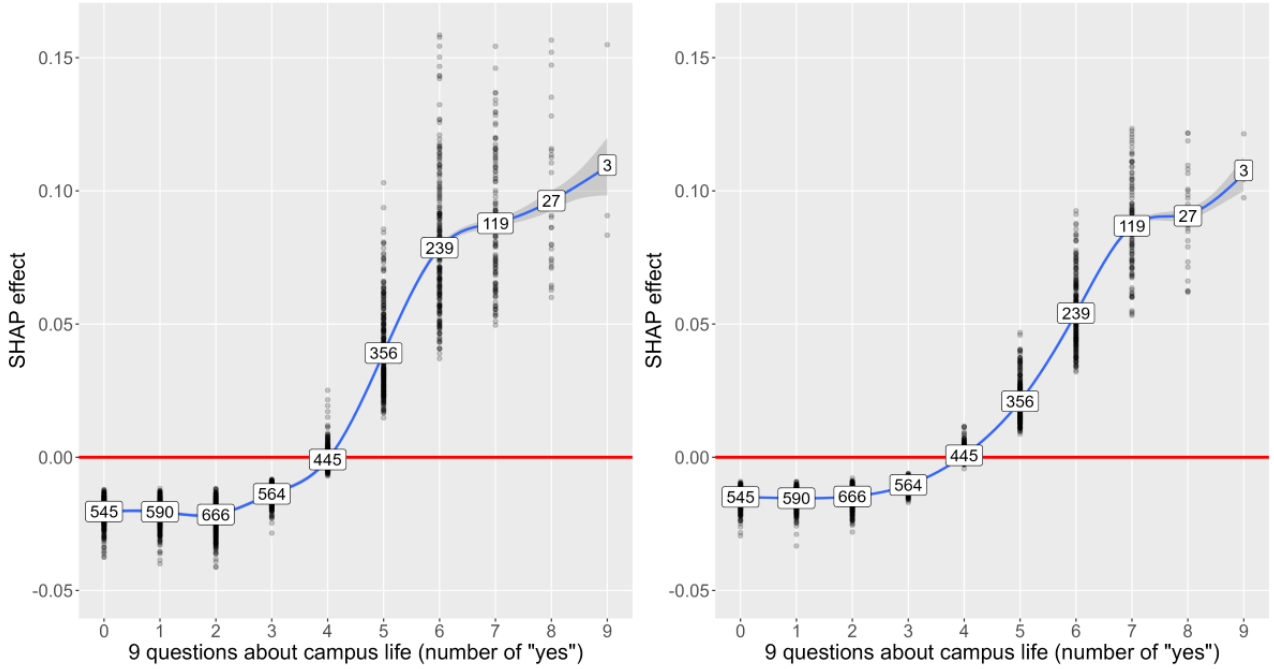


Figure 4. Shapley additive explanations (SHAP) dependence plot of sleeping hours (analysis 1, "without" condition).

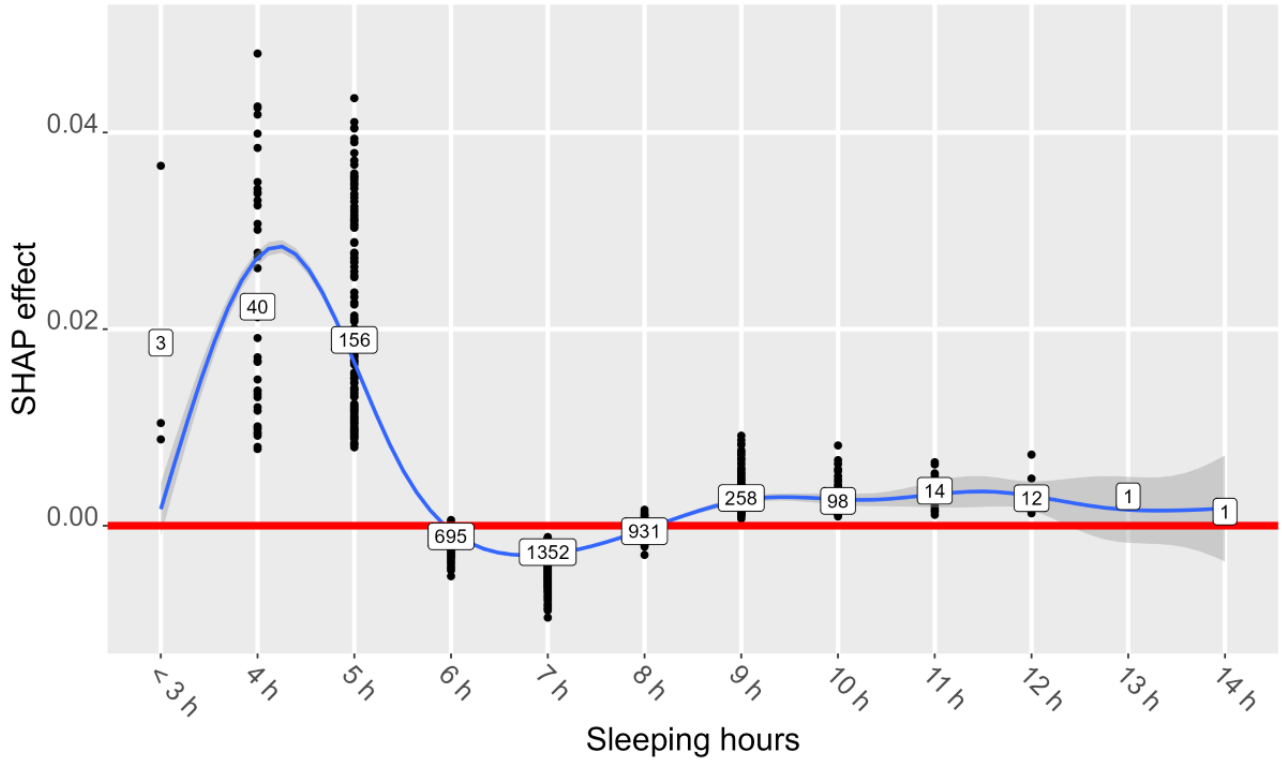


Figure 5. Shapley additive explanations (SHAP) violin plot of years of university on "with" condition (left: analysis 1; right: analysis 2).

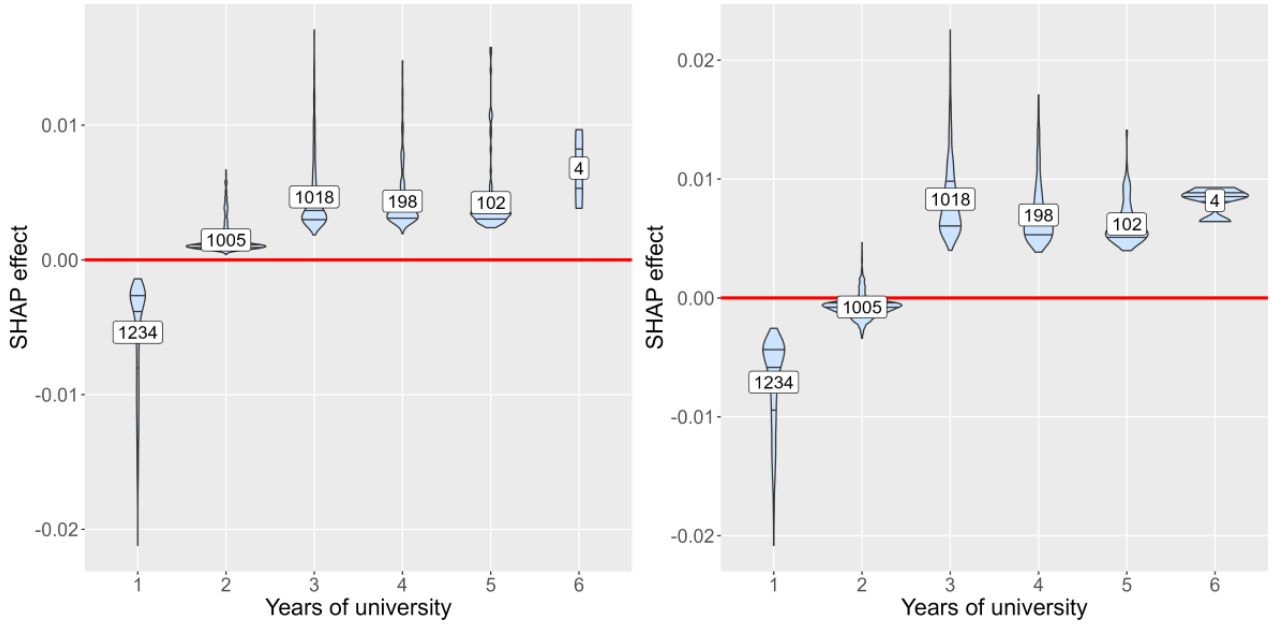


Figure 6. Shapley additive explanations (SHAP) dependence plot of the proportion of the same sex in respondent's department (analysis 2, "with" condition).

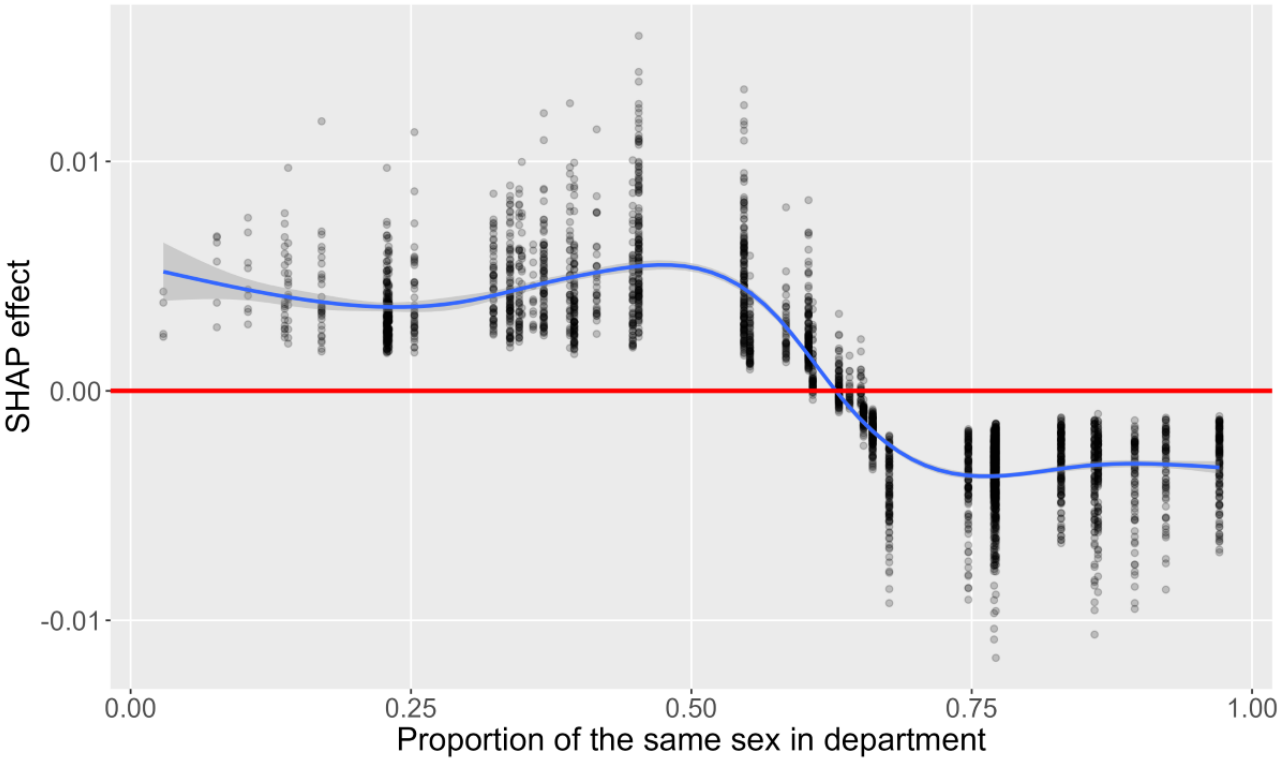


Figure 7. Shapley additive explanations (SHAP) dependence plot of xy coordinates of response time stamp (analysis 1, “with” condition).

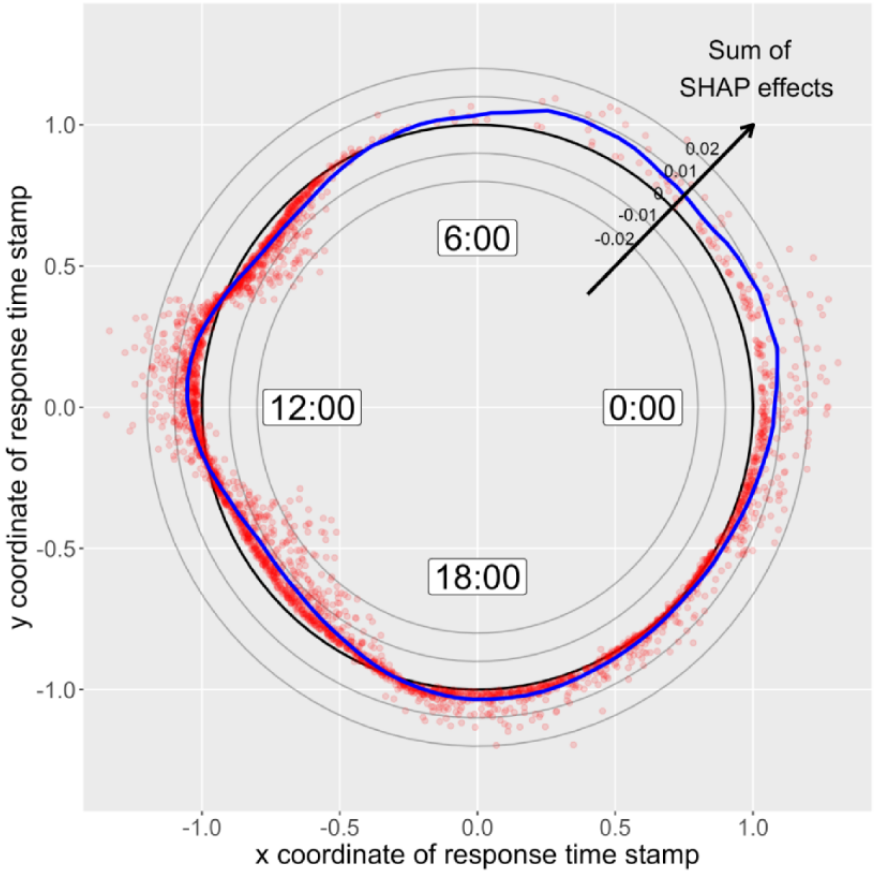
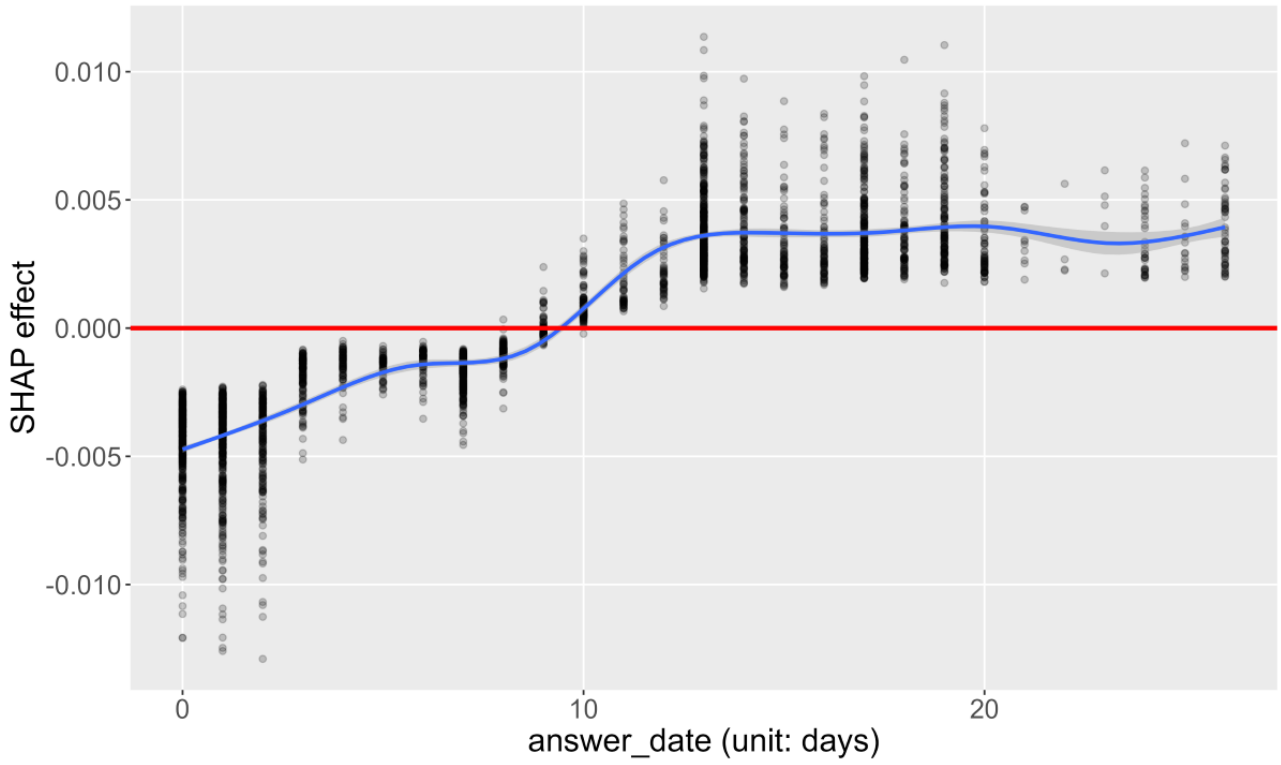


Figure 8. Shapley additive explanations (SHAP) dependence plot of answer date (analysis 2, “with” condition).



Discussion

Principal Findings

Among all the models, elastic net and LightGBM performed the best in analyses 1 and 2 (Tables 6 and 7). We adopted the LightGBM in this study based on the confusion matrix (Figure 2); however, the elastic net can also have an adequate ability to predict mental health problems not only in real time but also 1 year in the future.

Some of the answering time findings are listed in Table 8. Whether the *with* condition outperformed the *without* condition depended on the performance measure considered. This is possibly because the LightGBM can obtain excellent performance even without answering time–related variables on these data. However, as shown in Tables 9 and 11, certain variables are apparently helpful in improving the prediction and affecting the prediction probability. In summary, it is difficult to determine whether we should use answering time–related variables to predict mental health problems.

Focusing on input variables, those from *nine questions about campus life* improved performance (Tables 9 and 10) and were commonly effective for prediction (Tables 11 and 12). Comparing Gain and SHAP in *with* condition (Tables 9 and 11), it is suggested that some input variables converted from the answering time substantially impacted the learning. With respect to the *without* condition (Tables 10 and 12), it is suggested that the input variable, years of university, could improve the prediction to some extent.

Detailed Discussion on Input Variables

In Tables 9-12, the input variables from *nine questions about campus life* held a top-ranking position. In particular, the item "I have a strong anxiety about my campus life" had a stable and powerful influence on prediction. The questions were easy to use because there were only 9 yes or no questions. Figure 3 suggests that students were more likely to have mental health problems when they responded positively to ≥5 items of the 9 questions. Furthermore, the variable *nine questions about campus life* could be used by faculty members to advise student orally in addition to health staff members. Their response will be useful for whole-university support and for early detection.

Sleeping hours affected learning and prediction more in analysis 2 than in analysis 1 on the "without" condition (Tables 10 and 11). This implies that sleeping hours were less effective in learning and predicting mental health problems within 1 year. Sleeping hours may be a fluid measure that is affected by lifestyle and busyness. Figure 4 suggested that 6 to 8 hours of sleep were desirable in terms of mental health.

In Japan, 7 hours and 22 minutes are the average sleeping hours [57]. More than one-third of the students in this survey slept for 7 hours. Sleeping too little or too much is a representative symptom of mental health disorder (eg, depressive disorders) [58]. Even without any disorder, sleep diminution because of overwork and busyness could cause mental health problems, and sleeping too much may cause problems in academics. It can be useful for students sleeping <6 or >8 hours to be instructed on how to improve their life rhythm using flyers. Faculty members should keep in mind the need for students to sleep for >6 hours when handing assignments.

Years of university was an effective parameter in both analyses 1 and 2, and in both with and without conditions (Figure 5). The average SHAP effect of second-year students indicated that predictive probability decreased in the second year. It was reported that first- and last-year students tend to experience academic stress [59]. The results depicted in Figure 5 may imply that students adapt to the campus life from the beginning of the second year to their third year. Although some universities hold health surveys only for first-year students, this study indicated that third- or higher-year students should also be monitored.

Figure 6 is consistent with the hypothesis that some students who belonged to departments with few same-sex students may have had difficulties. By contrast, a low proportion of the same sex in a department equals a high proportion of the other sex in that department, indicating a low predictive probability of the other sex. Considering that even an approximately 0.50 proportion of the same sex in the department showed positive averages of the SHAP effect, it is unclear how much the male:female ratio is desirable for students' mental health. A low proportion of female students in the science and engineering departments reflected the gap in field of science, technology, engineering, and mathematics. The aspect of sex associated with these data was only male or female; the lesbian, gay, bisexual, transgender, queer community should be considered in future studies. Regardless of sex, being in a minority in a department should be a parameter to be examined; such examinations may afford opportunities for interaction.

There are certain findings regarding the input variables generated based on answering time. Concerning the RT stamp, the results depicted in Figure 7 were partly consistent with the hypothesis that RT stamps may reflect students' life rhythm and sleep quality, which may be related to their mental health state. Students whose answers were recorded at midnight may have stayed up late, which may have been related to mental health problems. In contrast, it is unclear why students whose answers were recorded at noon (approximately 12:00) or evening (approximately 18:00) had negative SHAP values. This might suggest that they had been alone when groups of other students enjoyed lunch or after school. RT stamp data have not been used before; such data can be informative for understanding students' lifestyles. Although the answer date was not ranked in Tables 9-12, the results depicted in Figure 8 were consistent with the hypothesis that late answer data may have reflected some problems. We believe that the high SHAP of late responses may be caused by the lack of ability to check their emails and obtain essential information from them. Students who answered the survey at a later date should be asked about difficulty in schoolwork (eg, submission of assignments) in health checkup interview. Demonstrating that answering time affects the prediction and learning of mental health problems is meaningful because this parameter can easily be collected through a web survey.

Limitations

The data were collected from only one university during COVID-19; therefore, the model cannot be generalized to other

universities and ages, whereas the approach can be. Campus life during COVID-19 differed from that without the virus. Moreover, this study could not reveal the condition of the 1994 students who did not provide answers in 2020 and 2021, even though they accounted for a major proportion of the population and may have had some mental health problems. The *nine questions about campus life*, which were the most powerful items in this study, are unique to University A and are not widely used. This model should be tested in other time periods or with other universities' data. Using this model to call for psychological counseling for students at risk in a specific field is entirely different. This model should be evaluated to determine whether it helps in precisely detecting students experiencing mental health problems and to estimate the impact on cost reduction.

Conclusions

Students' mental health problems were predicted in real time and for 1 year in the future. The *nine questions about campus life*, especially the question "I have a strong anxiety about my campus life" was an overwhelmingly powerful item. It was indicated that demographic data (eg, years of university, proportion of the same sex in department, etc) and behavioral data (sleeping hours and answering time), as well as self-rating items were effective.

The developed model itself should be adjusted for each university because it depends on the items used in the survey of the cooperating university. Nevertheless, this model demonstrates the possibility of synergistically using the characteristics of health surveys and advantages of ML. Consequently, it can be used for predicting mental health status from existing health data without a mental health scale.

Furthermore, the impact of some items on the prediction was discussed. These findings can improve health survey items and define the criteria for inviting to student counseling. Accurate calls for student counseling will lead to early detection and intervention and operational efficiency.

Acknowledgments

During the study, a portion of the interim results was presented at a briefing of the Research Support System for Female Researchers at Kanazawa University and the 85th Annual Convention of the Japanese Psychological Association. The staff at the Health Service Center at Kanazawa University performed data collection. Dr Satoshi Sugiyama (Keio University) and Dr Hiroshi Matsui (Hokkaido University) advised on our analysis. Prof Yumi Adachi (Kanazawa University) supported all processes. The authors would like to thank Editage for English editing.

The Research Support System supported this work for Female Researchers at Kanazawa University.

Authors' Contributions

AB contributed to the conceptualization, methodology, investigation, resources, data curation, writing—original draft, writing—review and editing, visualization, supervision, project administration, and funding acquisition. KB contributed to conceptualization, methodology, software, formal analysis, data curation, writing—original draft, writing—review and editing, and visualization.

Conflicts of Interest

None declared.

References

1. Furukawa T. Outcomes in psychological counseling. Shinri sodan ni okeru tenki ni tsuite. Campus Health 2020 Jun;57(2):56-63 [FREE Full text]
2. Hayakawa T. Implementation of the mental health survey: limitations and future challenges of the survey. Seishin kenko chosa no jisshi jyokyo: chosa no genkai to kongo no kadai. Campus Health 2020 Jun;57(2):51-55 [FREE Full text]
3. Hirayama K, Japanese Association for College Mental Health. User Guide for the UPI. UPI riyo no tebiki. Tokyo, Japan: Sozo-shuppan; 2011.
4. Goldberg DP. Manual of the General Health Questionnaire. Berkshire, UK: NFER-Nelson Publishers; 1978.
5. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychol Med 1979 Feb;9(1):139-145. [doi: 10.1017/s0033291700021644] [Medline: 424481]
6. Sakai W, Noguchi H. Comparison of tests of mental health for student counseling: formation of a common measure. Daigakusei wo taisyo toshita seishintekikenkodochosa no kyotsusyakudoka niyoru hikakukento. Jpn J Educ Psychol 2015;63(2):111-120 [FREE Full text] [doi: 10.5926/jjep.63.111]
7. Inaki K, Arai S, Kato Y, Takeuch H. Early detection and support of mental health issues through the university personality inventory (UPI) in student counseling. Bull Research Jin-Ai Univ Faculty of Human Studies 2020 Feb 20;18:1-9. [doi: 10.57426/00000868]
8. McArdle JJ, Ritschard G. Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences. London, UK: Routledge; 2013.

9. Burke TA, Jacobucci R, Ammerman BA, Alloy LB, Diamond G. Using machine learning to classify suicide attempt history among youth in medical care settings. J Affect Disord 2020 May 01;268:206-214. [doi: 10.1016/j.jad.2020.02.048] [Medline: 32174479]
10. Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
11. Thieme A, Belgrave D, Doherty G. Machine learning in mental health: a systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Trans Comput Hum Interact 2020 Oct 31;27(5):1-53 [FREE Full text] [doi: 10.1145/3398069]
12. Naghavi A, Teismann T, Asgari Z, Mohebbian MR, Mansourian M, Mañanas M. Accurate diagnosis of suicide ideation/behavior using robust ensemble machine learning: a university student population in the Middle East and North Africa (MENA) region. Diagnostics (Basel) 2020 Nov 16;10(11):956 [FREE Full text] [doi: 10.3390/diagnostics10110956] [Medline: 33207776]
13. Burke TA, Jacobucci R, Ammerman BA, Piccirillo M, McCloskey MS, Heimberg RG, et al. Identifying the relative importance of non-suicidal self-injury features in classifying suicidal ideation, plans, and behavior using exploratory data mining. Psychiatry Res 2018 Apr;262:175-183 [FREE Full text] [doi: 10.1016/j.psychres.2018.01.045] [Medline: 29453036]
14. Bagroy S, Kumaraguru P, De Choudhury M. A social media based index of mental well-being in college campuses. Proc SIGCHI Conf Hum Factor Comput Syst 2017 May;2017:1634-1646 [FREE Full text] [doi: 10.1145/3025453.3025909] [Medline: 28840202]
15. Sano A, Phillips AJ, Yu AZ, McHill AW, Taylor S, Jaques N, et al. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. Int Conf Wearable Implant Body Sens Netw 2015 Jun;2015:7299420 [FREE Full text] [doi: 10.1109/BSN.2015.7299420] [Medline: 28516162]
16. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
17. Tate AE, McCabe RC, Larsson H, Lundström S, Lichtenstein P, Kuja-Halkola R. Predicting mental health problems in adolescence using machine learning techniques. PLoS One 2020 Apr 06;15(4):e0230389 [FREE Full text] [doi: 10.1371/journal.pone.0230389] [Medline: 32251439]
18. Siemens G, Baker RS. Learning analytics and educational data mining: towards communication and collaboration. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. 2012 Presented at: LAK '12; April 29-May 2, 2012; Vancouver, Canada p. 252-254 URL: https://dl.acm.org/doi/10.1145/2330601.2330661 [doi: 10.1145/2330601.2330661]
19. Djulovic A, Li D. Towards freshman retention prediction: a comparative study. Int J Inf Educ 2013;3(5):494-500 [FREE Full text] [doi: 10.7763/IJIET.2013.V3.324]
20. Jayaprakash SM, Moody EW, Lauría EJ, Regan JR, Baron JD. Early alert of academically at-risk students: an open source analytics initiative. J Learn Anal 2014;1(1):6-47 [FREE Full text] [doi: 10.18608/jla.2014.11.3]
21. Kondo N, Hatanaka T. Modelling of students' learning states using big data of students through the baccalaureate degree program. Gakushi katei ni okeru daikibo data ni motoduku gakushu jotai no moderu ka. Jpn Soc Info Sys Edu 2016;33(2):94-103 [FREE Full text] [doi: 10.14926/jsise.33.94]
22. Otomo A, Iwayama Y, Mohri T. On-campus data utilization: Working on IR (institutional research) in universities. Gakunai data no katsuyo: Daigaku ni okeru IR (institutional research) heno torikumi. Fujitsu 2014 May;65(3):41-47 [FREE Full text]
23. Takaoka A, Nakai A, Sugiyama E, Nozue T, Shimizu R. The proposal for multi- faceted student support based on student data base: the collaboration of early detection of at-risk students and student counseling services. Gakusei no syugaku data wo katuyo shita takakuteki na gakuseishien no teian: youchui gakusei no sokihakken to gakuseisoudan tono kyodo. Meiji Gakuin University bulletin of psychology 2017;27(1):81-93 [FREE Full text]
24. Vihavainen A, Luukkainen M, Kurhila J. Using students' programming behavior to predict success in an introductory mathematics course. In: Proceedings of the 6th International Conference on Educational Data Mining. 2013 Presented at: EDM '13; July 6-9, 2013; Memphis, TN, USA p. 300-303 URL: https://www.educationaldatamining.org/EDM2013/papers/rn_paper_52.pdf
25. Iwata N, Kikuchi K, Fujihara Y. The usability of CAT system for assessing the depressive level of Japanese-a study on psychometric properties and response behavior. Int J Behav Med 2016 Aug;23(4):427-437. [doi: 10.1007/s12529-015-9503-1] [Medline: 26272358]
26. Wang C, Weiss DJ, Su S. Modeling response time and responses in multidimensional health measurement. Front Psychol 2019 Jan 29;10:51 [FREE Full text] [doi: 10.3389/fpsyg.2019.00051] [Medline: 30761036]
27. Nock MK, Banaji MR. Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. J Consult Clin Psychol 2007 Oct;75(5):707-715 [FREE Full text] [doi: 10.1037/0022-006X.75.5.707] [Medline: 17907852]
28. Furukawa TA, Kawakami N, Saitoh M, Ono Y, Nakane Y, Nakamura Y, et al. The performance of the Japanese version of the K6 and K10 in the World Mental Health Survey Japan. Int J Methods Psychiatr Res 2008;17(3):152-158 [FREE Full text] [doi: 10.1002/mpr.257] [Medline: 18763695]

29. Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, et al. Screening for serious mental illness in the general population. Arch Gen Psychiatry 2003 Feb;60(2):184-189. [doi: 10.1001/archpsyc.60.2.184] [Medline: 12578436]
30. Sakurai K, Nishi A, Kondo K, Yanagida K, Kawakami N. Screening performance of K6/K10 and other screening instruments for mood and anxiety disorders in Japan. Psychiatry Clin Neurosci 2011 Aug;65(5):434-441 [FREE Full text] [doi: 10.1111/j.1440-1819.2011.02236.x] [Medline: 21851452]
31. Comprehensive survey of living condition of the people on health and welfare 2019. Ministry of Health, Labour and Welfare. 2019. URL: https://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa19/dl/07.pdf [accessed 2023-02-01]
32. Breiman L. Random forests. Mach Lang 2001 Oct 01;45(1):5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
33. Ke G, Meng Q, Finley TW, Wang T, Chen WY, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17; Long Beach, CA, USA; December 4-9, 2017 p. 3149-3157 URL: https://dl.acm.org/doi/10.5555/3294996.3295074 [doi: 10.5555/3294996.3295074]
34. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16; August 13-17, 2016; San Francisco, CA, USA p. 785-794 URL: https://dl.acm.org/doi/10.1145/2939672.2939785 [doi: 10.1145/2939672.2939785]
35. Chen Y, Liu Y. Which risk factors matter more for psychological distress during the COVID-19 pandemic? An application approach of gradient boosting decision trees. Int J Environ Res Public Health 2021 May 30;18(11):5879 [FREE Full text] [doi: 10.3390/ijerph18115879] [Medline: 34070829]
36. Yuan X, Chen J, Zhang K, Wu Y, Yang T. A stable AI-based binary and multiple class heart disease prediction model for IoMT. IEEE Trans Industr Inform 2022 Mar;18(3):2032-2040 [FREE Full text] [doi: 10.1109/tii.2021.3098306]
37. Uchida C. Suicide among Japanese university students from the results of a 21 year survey: a clue to prevent suicide among university students. 21nenkan no chosa karamita daigakusei no jisatsu no tokucho to kikeninshi: yobo heno tegakari wo saguru. Article in Japanese. Psychiatria et neurologia Japonica 2010;112(6):543-560 [FREE Full text]
38. Higuchi K. A two-step approach to quantitative content analysis: KH coder tutorial using anne of green gables (part I). Retsumeikan Soc Sci Rev 2016;52(3):77-91 [FREE Full text] [doi: 10.34382/00003706]
39. Higuchi K. A two-step approach to quantitative content analysis: KH coder tutorial using anne of green gables (part II). Retsumeikan Soc Sci Rev 2017;53(1):137-147 [FREE Full text] [doi: 10.34382/00003731]
40. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2019. URL: https://www.R-project.org/ [accessed 2023-02-01]
41. Josse J, Husson F. missMDA: a package for handling missing values in multivariate data analysis. J Stat Softw 2016;70(1):1-31 [FREE Full text] [doi: 10.18637/jss.v070.i01]
42. Bergstra JA, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. 2011 Presented at: NIPS'11; December 12-15, 2011; Granada, Spain p. 2546-2554 URL: https://dl.acm.org/doi/10.5555/2986459.2986743
43. JuliaAI, The Alan Turing Institute. API: MLJLinearModels. GitHub. 2020. URL: https://juliaai.github.io/MLJLinearModels.jl/dev/api/ [accessed 2023-02-01]
44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text] [doi: 10.1145/2786984.2786995]
45. xgboost developers. XGBoost parameters. XGBoost. 2022. URL: https://xgboost.readthedocs.io/en/stable/parameter.html [accessed 2023-02-01]
46. Parameters. LightGBM 3.3.2.99. Redmond, WA, USA: Microsoft Corporation; 2022. URL: https://lightgbm.readthedocs.io/en/latest/Parameters.html [accessed 2023-01-13]
47. Fernández A, Galar M, García S, Herrera F, Krawczyk B, Prati RC. Learning from Imbalanced Data Sets. Basel, Switzerland: Springer; 2018.
48. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2nd edition. Victoria, Canada: Leanpub; 2022.
49. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: a fresh approach to numerical computing. SIAM Rev 2017 Jan;59(1):65-98 [FREE Full text] [doi: 10.1137/141000671]
50. Blaom A, Kiraly F, Lienart T, Simillides Y, Arenas D, Vollmer S. MLJ: a Julia package for composable machine learning. J Open Source Softw 2020 Nov 07;5(55):2704 [FREE Full text] [doi: 10.21105/joss.02704]
51. Bergstra J, IQVIA. TreeParzen.jl. GiHub. 2022. URL: https://github.com/IQVIA-ML/TreeParzen.jl [accessed 2023-02-01]
52. The scikit-learn developers, The ScikitLearn.jl developers. ScikitLearn.jl. GitHub. 2022. URL: https://github.com/cstjean/ScikitLearn.jl [accessed 2023-02-01]
53. van Mossel A. LightGBM.jl. GitHub. 2022. URL: https://github.com/IQVIA-ML/LightGBM.jl [accessed 2023-02-01]
54. Redell N. ShapML.jl. GitHub. 2022. URL: https://github.com/nredell/ShapML.jl [accessed 2023-02-01]
55. Gao K, Mei G, Piccialli F, Cuomo S, Tu J, Huo Z. Julia language in machine learning: algorithms, applications, and open issues. Comput Sci Rev 2020 Aug;37:100254 [FREE Full text] [doi: 10.1016/j.cosrev.2020.100254]
56. Perkel JM. Julia: come for the syntax, stay for the speed. Nature 2019 Aug;572(7767):141-142. [doi: 10.1038/d41586-019-02310-3] [Medline: 31363196]

57. Gender data portal time use across the world. Organisation for Economic Co-operation and Development. URL: https://www.oecd.org/gender/data/ [accessed 2023-02-12]
58. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th edition. Washington, DC, USA: American Psychiatric Association; 2013.
59. Bedewy D, Gabriel A. Examining perceptions of academic stress and its sources among university students: the perception of academic stress scale. Health Psychol Open 2015 Jul;2(2):2055102915596714 [FREE Full text] [doi: 10.1177/2055102915596714] [Medline: 28070363]

Abbreviations

AUC-PR: area under the curve of precision-recall curve
AUC-ROC: area under the curve of receiver operating characteristic curve
GBDT: gradient boosting decision tree
MCC: Matthews correlation coefficient
ML: machine learning
RT: response time
SHAP: Shapley additive explanations

Edited by J Torous, G Eysenbach; submitted 03.09.22; peer-reviewed by CH Li, A Teles; comments to author 29.12.22; revised version received 19.02.23; accepted 19.02.23; published 10.05.23.

Please cite as:
Baba A, Bunji K
Prediction of Mental Health Problem Using Annual Student Health Survey: Machine Learning Approach
JMIR Ment Health 2023;10:e42420
URL: https://mental.jmir.org/2023/1/e42420
doi:10.2196/42420
PMID:37163323

Original Paper

Association of Depressive Symptom Trajectory With Physical Activity Collected by mHealth Devices in the Electronic Framingham Heart Study: Cohort Study

Xuzhi Wang[1], MS; Chathurangi H Pathiravasan[1], PhD; Yuankai Zhang[1], MA; Ludovic Trinquart[1,2,3], PhD; Belinda Borrelli[4], PhD; Nicole L Spartano[5], PhD; Honghuang Lin[6], PhD; Christopher Nowak[7], PhD; Vik Kheterpal[7], MD; Emelia J Benjamin[8,9,10], ScM, MD; David D McManus[11,12], ScM, MD; Joanne M Murabito[9,13], ScM, MD; Chunyu Liu[1], PhD

[1]Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States

[2]Tufts Clinical and Translational Science Institute, Tufts University, Boston, MA, United States

[3]Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, United States

[4]Center for Behavioral Science Research, Boston University Henry M Goldman School of Dental Medicine, Boston, MA, United States

[5]Section of Endocrinology, Diabetes, Nutrition, and Weight Management, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States

[6]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[7]Care Evolution, Ann Arbor, MI, United States

[8]Section of Preventive Medicine and Epidemiology and Cardiovascular Medicine, Department of Medicine, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States

[9]Framingham Heart Study, Boston University and National Heart, Lung, and Blood Institute, Framingham, MA, United States

[10]Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

[11]Cardiology Division, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[12]Department of Quantitative Health Sciences, University of Massachusetts Chan Medical School, Worcester, MA, United States

[13]Section of General Internal Medicine, Department of Medicine, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, United States

Corresponding Author:
Chunyu Liu, PhD
Department of Biostatistics
Boston University School of Public Health
715 Albany Street
Boston, MA, 02118
United States
Phone: 1 6176385104
Email: liuc@bu.edu

Abstract

Background: Few studies have examined the association between depressive symptom trajectories and physical activity collected by mobile health (mHealth) devices.

Objective: We aimed to investigate if antecedent depressive symptom trajectories predict subsequent physical activity among participants in the electronic Framingham Heart Study (eFHS).

Methods: We performed group-based multi-trajectory modeling to construct depressive symptom trajectory groups using both depressive symptoms (Center for Epidemiological Studies-Depression [CES-D] scores) and antidepressant medication use in eFHS participants who attended 3 Framingham Heart Study research exams over 14 years. At the third exam, eFHS participants were instructed to use a smartphone app for submitting physical activity index (PAI) surveys. In addition, they were provided with a study smartwatch to track their daily step counts. We performed linear mixed models to examine the association between depressive symptom trajectories and physical activity including app-based PAI and smartwatch-collected step counts over a 1-year follow-up adjusting for age, sex, wear hour, BMI, smoking status, and other health variables.

Results: We identified 3 depressive symptom trajectory groups from 722 eFHS participants (mean age 53, SD 8.5 years; n=432, 60% women). The low symptom group (n=570; mean follow-up 287, SD 109 days) consisted of participants with consistently low CES-D scores, and a small proportion reported antidepressant use. The moderate symptom group (n=71; mean follow-up 280, SD 118 days) included participants with intermediate CES-D scores, who showed the highest and increasing likelihood of reporting antidepressant use across 3 exams. The high symptom group (n=81; mean follow-up 252, SD 116 days) comprised participants with the highest CES-D scores, and the proportion of antidepressant use fell between the other 2 groups. Compared to the low symptom group, the high symptom group had decreased PAI (mean difference –1.09, 95% CI –2.16 to –0.01) and the moderate symptom group walked fewer daily steps (823 fewer, 95% CI –1421 to –226) during the 1-year follow-up.

Conclusions: Antecedent depressive symptoms or antidepressant medication use was associated with lower subsequent physical activity collected by mHealth devices in eFHS. Future investigation of interventions to improve mood including via mHealth technologies to help promote people's daily physical activity is needed.

(JMIR Ment Health 2023;10:e44529) doi:10.2196/44529

KEYWORDS

depression; mobile health; risk factors; physical activity; eCohort; Framingham Heart Study

Introduction

Depression is described as having negative feelings such as sadness or a loss of interest in activities that a person once enjoyed. Depression is a common and serious medical illness that affects millions of Americans each year [1-3]. A recent study reported that the prevalence of elevated depressive symptoms is 32.8% in a national, population-representative, longitudinal study of US adults in 2021 [4]. Depression and physical activity have bidirectional relationships. Physical inactivity is a risk factor for depression [5-7]. Evidence suggests that the use of behavioral therapies such as exercise can help improve the treatment of depression [8]. Low doses of physical activity are protective against depression [5], and regular exercise can result in substantial reductions in depression in the general older adult population [9].

Most previous studies evaluated the relationships of depression with subjectively measured physical activity (ie, self-reported physical activity). Few studies have investigated how variation in mood is associated with physical activity [10-12]. An investigation of depressive symptoms and self-reported physical activity in the 1958 British Birth Cohort in adults from age 23 to 50 years suggests that depressive symptoms in early adulthood may be a barrier to physical activity [11]. The Whitehall Study reported that participants with depressive symptoms at baseline were more likely to fail to meet the recommended levels of physical activity at follow-up [12]. The Coronary Artery Risk Development in Young Adults study identified a longitudinal association between baseline depressive symptoms and self-reported moderate to vigorous intensity physical activity over 10-year follow-up but not with objectively measured moderate to vigorous intensity physical activity [13]. To our best knowledge, no study considered the association between the longitudinal trajectory of depressive symptoms and physical activity measured with mobile devices both self-reported and objectively.

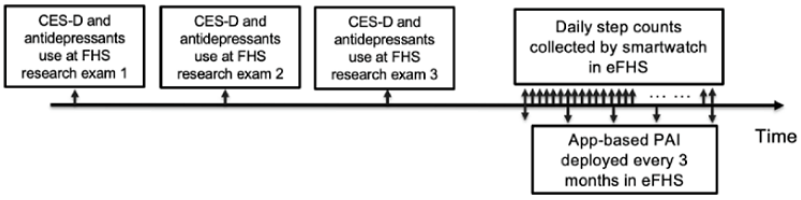
Mobile health (mHealth) devices, including smartphones and wearables, are increasingly used by many Americans to monitor their health and health-related behaviors [14]. The electronic Framingham Heart Study (eFHS) is an e-cohort nested within the larger multigenerational Framingham Heart Study (FHS). The participants in the eFHS were instructed to use a smartphone app to submit their health survey data. They were also invited to wear a study smartwatch to record daily step counts. The primary aim of this investigation was to examine the association of antecedent depressive symptoms and antidepressant medication use with subsequent physical activity collected during eFHS. We hypothesized that the long-term trend of depressive symptoms and antidepressant use prior to enrollment in the eFHS was associated with subsequent lower levels of physical activity during eFHS.

Methods

Overview

We captured the longitudinal trend of depressive symptoms and categorized them into distinct trajectory groups based on a composite depressive symptom score and the use of antidepressant medications assessed at FHS research examinations 1-3 across 14 years prior to the eFHS. We assessed whether the constructed depressive symptom patterns predicted the average level of physical activity collected by mHealth devices during short-term follow-up in the eFHS. Last, we examined whether the constructed depressive symptom patterns predicted habitual physical activity within an intermediate time span of 1-year follow-up in the eFHS (Figure 1). Study method and results are reported following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement for cross-sectional studies.

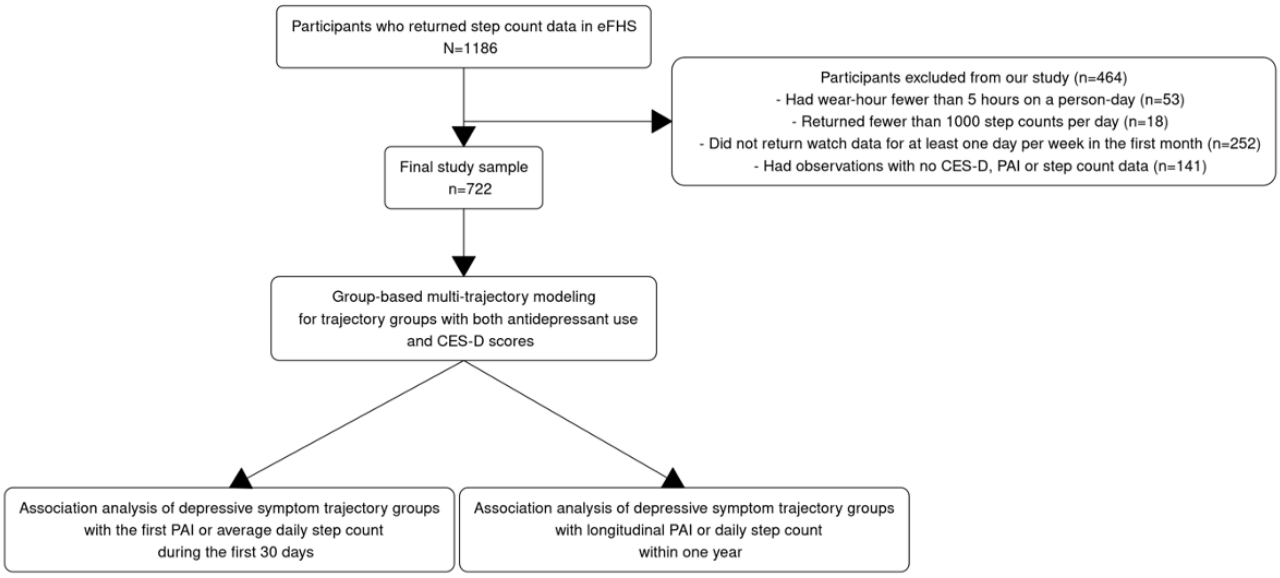
Figure 1. The timeline of data collection for CES-D scores and antidepressant use at Framingham Heart Study (FHS) research exams, as well as app-based PAI and smartwatch-collected step. CES-D: Center for Epidemiological Studies-Depression; eFHS: electronic Framingham Heart Study; PAI: physical activity index.



Study Participants

The FHS began in 1948 by enrolling 5209 residents (original cohort) from Framingham, Massachusetts. In 1971, a total of 5214 participants who were offspring of the original cohort, and the spouses of the offspring participants were enrolled in an offspring cohort. During 2002 and 2005, the FHS enrolled the third generation (Gen 3) cohort (n=4094), which included grandchildren of the original cohort. During the same time, the FHS enrolled 410 residents who were from underrepresented racial or ethnic groups in or near Framingham (Omni2), and New offspring spouse cohort (NOS, n=103). This study included participants in the Gen 3, Omni2, and NOS cohorts who have undergone 3 research examinations every 6 to 8 years (exam 1: 2002-2005, exam 2: 2008-2011, exam 3: 2016-2019) [15]. At research exam 3 (2016 to 2019), English-speaking participants who owned a smartphone (including iPhone 4S [Apple Inc] or higher with iOS version 8.2 [Apple Inc] or higher or beginning October 30, 2017, an Android phone) were invited to enroll in the eFHS [16]. Most participants who agreed to participate in eFHS downloaded the eFHS smartphone app from the Apple App Store at the in-person research center visit with help from the FHS-trained staff. Some participants chose to register in eFHS after leaving the research center with written instructions provided by eFHS staff. The sample size of our main analysis included participants who attended FHS examinations 1-3 and had self-reported physical activity and valid Apple Watch step data (defined below) for the first month following their third examination (Figure 2).

Figure 2. Study sample selection and statistical analyses. CES-D: Center for Epidemiological Studies-Depression; eFHS: electronic Framingham Heart Study; PAI: physical activity index.



Ethics Approval

All study participants provided informed consent for the study. The protocols for eFHS and the FHS were approved by the Institutional Review Board (H-36586 and H-32132) at the Boston University Medical Center.

Center for Epidemiological Studies-Depression Score and Antidepressant Medications for Trajectory Building

We selected participants from Gen 3, OMNI2, and NOS cohorts who participated in all 3 FHS exams. Trained technicians administered the 20-item Center for Epidemiological Studies-Depression (CES-D) scale at each of the 3 research exams. The CES-D scale assesses how often individuals experienced depressive symptoms over the past week, with the total score ranging from 0 to 60. Each item has a score that ranges from 0 to 3 [17]. In 16 of the 20 CES-D questions, a higher score corresponds to a higher frequency of depressive symptoms. For the remaining 4 questions, a higher score corresponds to a lower frequency of depressive symptoms. We reversed the scores of these 4 questions and summed up the individual scores over the 20 questions to construct the composite CES-D score for each participant at every attended exam. A threshold of 16 or greater is commonly used to identify individuals who have depressive symptoms [17].

At each FHS examination, participants were asked to bring all medications to the research center in a medication bag. All antidepressant medications were recorded, and we identified antidepressant medications such as sertraline, fluoxetine, and citalopram. We considered that participants used antidepressant medication if any antidepressant medication was identified at a given FHS exam.

Assessment of Step Counts and Physical Activity Index

At exam 3, iPhone users were provided with the choice to participate in eFHS or not to participate. For users who chose to participate in the study, they had a choice to use the study smartwatch, an Apple Watch Series 0 (Apple Inc), or the first generation of Apple Watch. eFHS study staff assisted participants with pairing the watch to their smartphone in the Research Center. Participants were asked to wear the watch daily for at least 1 year [16].

We defined a person-day as a day for which a participant returned any amount of smartwatch data, and we defined "wear hours" as the number of hours during which a participant wore the smartwatch on a person-day. We excluded person-days for which participants had fewer than 5 wear hours (Figure 2) [18]. To minimize inaccurate data points, we further excluded person-days for which participants had less than 1000 step counts per day [19]. We then removed participants who failed to return watch data for at least 1 day per week in the first month after the participant started returning data in eFHS, because we planned to examine the association between antecedent depressive symptoms and average step counts during the first 30 days. In addition, we used data collected from the eFHS enrollment at the exam 3 to 1-year follow-up because one of the aims of this study was to investigate the association between antecedent depressive symptoms with habitual physical activity over an intermediate time span.

Daily step counts were collected by the smartwatch. Baseline average daily step counts were calculated as the mean of the daily step counts during the first 30 days after a participant's first return of the step data. Longitudinal step counts referred to the repeated daily step counts within the 1-year follow-up since the first return of the step data in eFHS.

In addition to the physical activity data that were objectively measured from the smartwatch, we also assessed the self-reported physical activity index (PAI) collected by the eFHS smartphone app for comparison. The eFHS smartphone app included a survey to assess physical activity using the Framingham PAI. The physical activity survey was sent through the app every 3 months for 1 year. The PAI was calculated as a composite score based on the number of hours spent sleeping, or in sedentary, slight, moderate, and heavy activities during a 24-hour period. We assigned weights of 1, 1.1, 1.5, 2.4, and 5 to sleep, sedentary, slight, moderate, and heavy activity, respectively [20-22]. Baseline PAI was defined as the first app-based PAI record collected within the first 3 months by the eFHS smartphone app. Longitudinal PAI referred to the repeated app-based PAI collected within 1-year follow-up since the baseline PAI.

Figure 1 displays the timeline of data collection for CES-D scores and antidepressant medication use at FHS research exams, as well as for app-based PAI and smartwatch-collected step counts.

Covariates

Demographic variables were measured during exam 3 at the research center. BMI was calculated by dividing body weight (kg) by height (meters) square. Obesity was defined as a BMI value of 30 kg/m^2 or higher. Using the highest education level the participant achieved at exam 3, we defined the education variable into 3 categories: "less than high school," "completed high school or some college," and "bachelor's degree or higher." Marital status was defined into 5 categories: "divorced," "married," "never married," "separated," and "widowed." Lipid-lowering treatment was defined as a self-report of receiving lipid treatment in the year before the exam. Hypertension was defined by systolic blood pressure of 130 mm Hg or higher, or diastolic blood pressure of 80 mm Hg or higher, or antihypertensive medication use. We also defined current smoking as participants who reported smoking at least 1 cigarette per day in the year before the exam. Several laboratory variables, including total cholesterol, high-density lipoprotein cholesterol, triglycerides, and fasting blood glucose, were also measured at exam 3. Diabetes was defined as the prior diagnosis, use of diabetes medications, hemoglobin A_{1c} (HbA_{1c}) ≥6.5%, fasting blood glucose ≥126 mg/dL, or random plasma glucose ≥200 mg/dL; prevalent cardiovascular disease (CVD) was defined as history of heart failure, myocardial infarction, angina, stroke, or intermittent claudication and coded.

Statistical Analyses

Overview

We first examined the main characteristics of the participants and compared the characteristics between those who remained in our study and those who were excluded. Next, we built the trajectory groups based on the longitudinal patterns of depressive symptoms (CES-D scores) and antidepressant medication use. Furthermore, we performed association analyses to examine the relationships between trajectory groups and physical activity variables during the first 30 days and during the 1-year follow-up.

Trajectory Modeling

We constructed the depressive symptom trajectories across exam 1 (2002-2005) to exam 3 (2016-2019) to capture different temporal patterns of CES-D scores and antidepressant medication use. We used group-based multi-trajectory [23] modeling to construct the trajectory patterns with both the antidepressants use and the continuous CES-D scores from research exams 1 to 3 (occurring over the 14 years before the eFHS). The trajectories of CES-D scores and antidepressant use were specified as a function of exam, adjusted for baseline age and sex. The group-based multi-trajectory modeling is an extension of univariate group-based trajectory modeling. The group-based multi-trajectory modeling identifies latent subgroups of individuals following similar longitudinal trajectories across multiple correlated outcome variables over time. This method is flexible to model the interrelationship of

multiple clinically relevant variables. Group-based multi-trajectory models with different trajectory shapes (ie, linear or quadratic) and with a different number of groups (ranging from 1 to 4) were tested. To determine the optimal number of trajectory groups, we started with a single group, and added 1 more group at a time. We used the Bayesian information criterion statistic to evaluate the model fit with different groups and shapes, with an additional requirement that each group consists of at least 5% of the sample [24].

Association Analysis of Depressive Symptom Trajectory Groups With the First App-Based PAI of eFHS or the Average Daily Step Counts During the First 30 Days of eFHS

We investigated both self-reported PAI and objectively measured step count in 3 trajectory groups. First, we performed a correlation analysis for the first app-based PAI and the average daily step count during the first 30 days. Next, we conducted analysis of covariance models to compare the first app-based PAI and the daily average step counts during the first 30 days (outcome) in the depressive symptom trajectory groups (predictor). Models were adjusted for age, sex (additionally adjusted for average daily wear hour for steps; model 1) and further adjusted for BMI, smoking status, prevalent CVD, current diabetes status, education, marital status, hypertension status, fasting blood glucose, high-density lipoprotein cholesterol, total cholesterol, triglycerides, and lipid treatment at exam 3 as additional variables to those in model 1 (model 2).

Association Analysis of the Depressive Symptom Trajectory Groups With the Longitudinal PAI or Daily Step Counts

We further examined the associations of the depressive symptom trajectory groups as the independent variable with the longitudinal PAI and daily step counts collected over a 1-year follow-up in the eFHS. We used a linear mixed model with longitudinal PAI values collected every 3 months as the outcome, adjusting for age, sex, and the number of days between baseline and a survey collection. We further adjusted for all the other covariates in model 2 in the previous section. We used linear mixed models with the repeated daily step counts as the outcome, adjusting for age, sex, the number of days between baseline and data collection, and the daily wear hour of the smartwatch use. Similarly, the model with daily step counts was further adjusted for other covariates.

All statistical analyses were conducted using SAS (version 9.3; SAS Institute) and R (version 4.0; R Foundation for Statistical Computing). Group-based multi-trajectory modeling was performed using the SAS Proc Traj [23]. We used 2-tailed $P<.05$ for significance.

Results

Participant Characteristics

A total of 3196 FHS participants (2873 from Gen 3, 49 from NOS, and 274 from Omni2) attended research exams 1 to 3. Of these 3196 participants, 2150 participants enrolled in the eFHS study, and 1186 participants returned app-based PAI and smartwatch-collected step counts data [25]. After applying 3 exclusion criteria to those who had both PAI and step counts records and removing those with missing values in any of the covariates, 722 individuals (median follow-up 14 years; mean age 53 years, SD 8.5 years; n=432, 60% women; n=656, 91% White) were included in the association analyses (Figure 2). Within a 1-year window, the participants had a median follow-up time of 52 weeks (Q1-Q3=50, 52). Compared to FHS participants who declined enrollment in eFHS, our study sample was significantly younger (mean 53, SD 8.5 years vs mean 57, SD 10 years; $P<.001$), had a higher proportion of women (432/722, 60% vs 632/1275, 50%; $P<.001$), lower CES-D scores (mean 5.79, SD 6.63 vs mean 7.27, SD 7.47; $P<.001$), lower BMI (mean 28.2, SD 5.6 vs mean 28.9, SD 6; $P=.04$), and a higher proportion of college graduates (495/722, 69% vs 605/1275, 47%; $P<.001$). Compared to other eFHS participants who accepted enrollment but did not meet the inclusion criteria, our study sample was significantly more likely to be women (432/722, 60% vs 670/1255, 53%; $P=.006$) and had a higher education level (bachelor's degree or higher 495/722, 69% vs 802/1255, 64%; $P=.04$; Multimedia Appendix 1).

Depressive Symptom Trajectory Groups

To identify the best trajectory model, we explored 2 polynomial orders (ie, linear and quadratic) with different numbers of trajectory groups in the group-based multi-trajectory modeling. The best model with 3 trajectory groups with a linear trend was chosen based on the model selection criteria (Figure 3). According to the trend across 3 exams, the patterns of the 3 trajectories were classified as "low symptom," "moderate symptom," and "high symptom" groups. The low symptom group (570/722, 79%; mean follow-up 287, SD 109 days) consisted of participants who had consistently low CES-D scores (mean CES-D 4.64 at exam 1, 3.72 at exam 2, and 4.06 at exam 3) and more than 95% of them did not report antidepressant medication use across exam 1 to exam 3. The moderate symptom group (71/722, 9.8%; mean follow-up 280, SD 118 days) comprised participants who exhibited intermediate CES-D scores (mean CES-D 6.56 at exam 1, 5.06 at exam 2, and 5.73 at exam 3). About 67.6% (48/71) of participants in the moderate symptom group took antidepressants at exam 1, 87.3% (62/71) at exam 2, and 91.5% (65/71) at exam 3. The high symptom group (81/722, 11.2%; mean follow-up 252, SD 116 days) consisted of participants who had the highest CES-D scores (mean CES-D 17.5 at exam 1, 17.2 at exam 2, and 18 at exam 3), surpassing the threshold CES-D score of 16 consistent with the presence of depressive symptoms. Among participants in the high symptom group, 21% (17/81) at exam 1, 33.3% (27/81) at exam 2, and 46.9% (38/81) at exam 3 reported the use of antidepressants. Both the moderate and high symptom groups increased the antidepressant medication use from exam 1 to exam 3. However, a larger proportion of participants reported antidepressants across exams 1 to 3 in the moderate symptom group than in the high symptom group.

We also observed substantial differences in the other study characteristics (Table 1). Compared to the low symptom group (mean age 52 years, mean BMI 27.9), the participants in the moderate (mean age 55 years, mean BMI 29.6) and high symptom groups (mean age 54 years, mean BMI 29.5) were

older and heavier. Among the 3 trajectory groups, the participants in the moderate symptom group were the most likely to be women (57/71, 80.3% compared to 317/579, 55.6% in the low symptom group and 58/81, 71.6% in the high symptom group). In addition, the participants in the high symptom group exhibited the lowest wear hours per day (mean wear hour per day was 12.1 compared to 13.1 hours per day in the low symptom group and 12.8 wear hours per day in the moderate symptom group). The high symptom group also had the shortest follow-up duration (mean follow-up days 252 compared to 287 in the low symptom group and 280 in the moderate symptom group).

Figure 3. The relationship of trajectory groups with CES-D scores and antidepressant medication use. The average CES-D scores and the proportions of participants with antidepressant use are shown in Figures A and B, respectively. CES-D: Center for Epidemiological Studies-Depression.

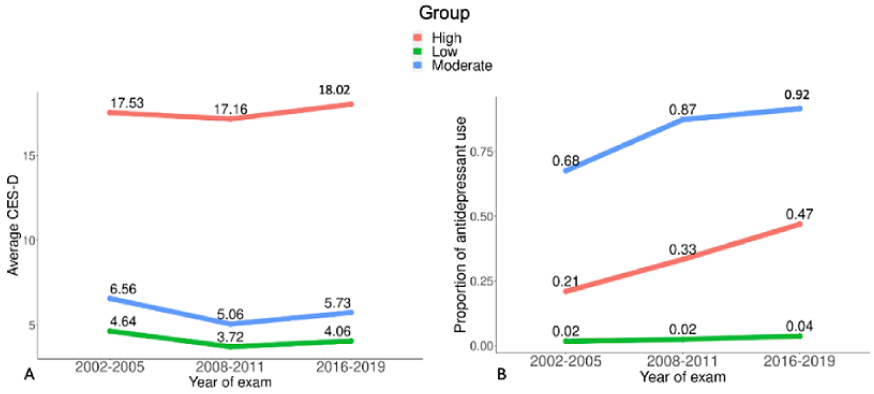


Table 1. Characteristics of 722 participants by trajectory group[a].

Characteristics	Low symptom	Moderate symptom	High symptom	*P* value
Total, n (%)	579 (79)	71 (9.8)	81 (11.2)	N/A[b]
Age (years), mean (SD)	52 (8.7)	55 (7.7)	54 (7.9)	.02
Sex (female), n (%)	317 (55.6)	57 (80.3)	58 (71.6)	<.001
Wear hours per day, mean (SD)	13.1 (2.12)	12.8 (2.07)	12.1 (2.31)	<.001
Follow-up days, mean (SD)	287 (109)	280 (118)	252 (116)	.03
BMI, mean (SD)	27.9 (5.28)	29.6 (6.82)	29.5 (6.29)	.004
Current smoking (yes), n (%)	28 (4.9)	1 (1.4)	7 (8.6)	.12
Marital status (married), n (%)	443 (77.7)	53 (74.6)	45 (55.6)	<.001
Hypertension (yes), n (%)	251 (44)	35 (49.3)	48 (59.3)	.03
Education, n (%)				.24
Less than high school	2 (0.4)	0 (0)	0 (0)	
Completed high school or some college	162 (28.4)	27 (38)	33 (40.7)	
Bachelor's degree or higher	403 (70.7)	44 (62)	48 (59.3)	
Fasting blood glucose, mean (SD)	97.7 (17.0)	98.1 (17.0)	95.5 (13.7)	.52
Diabetes status, n (%)	29 (5.1)	4 (5.6)	6 (7.4)	.69
Prevalent CVD[c], n (%)	24 (4.2)	3 (4.2)	2 (2.5)	.75
HDL[d], mean (SD)	61.2 (20.0)	62.7 (18.1)	60.8 (17.8)	.81
Total cholesterol, mean (SD)	188 (36.8)	199 (36.7)	194 (36.7)	.03
Triglycerides, mean (SD)	105 (76.6)	111 (46.3)	107 (66.6)	.85
Lipid treatment	111 (19.5)	22 (31.0)	12 (16.0)	.046

[a]Wear hour and follow-up days of step counts were measured during 1-year follow-up during eFHS, and all the other variables were measured at exam 3 (Year 2016-2019).

[b]N/A: not applicable.

[c]CVD: cardiovascular disease.

[d]HDL: high-density lipoprotein.

Association of Depressive Symptom Trajectory Groups With Physical Activity

We observed a weak positive correlation between the first app-based PAI and smartwatch-collected step counts during the first 30 days (Pearson correlation coefficient r=0.21; Figure 4). We further quantified the association between trajectory groups and the first PAI values collected from the eFHS smartphone app (Table 2). Adjusting for age and sex, participants in the high symptom group had, on average, a 0.83 smaller PAI compared to the low symptom group (95% CI –2.05 to 0.38), but the P value did not reach statistical significance (P=.18). After further adjusting for other covariates, a greater and significant decrease in the PAI for the participants in the high symptom group was observed. They had a 1.29 smaller PAI compared to the low symptom group (95% CI –2.52 to –0.06; P=.04). Furthermore, the PAI of participants in the moderate symptom group was 1.08 smaller than that of the low symptom group after we adjusted for all covariates. However, this difference was not statistically significant (P=.10).

We also observed a significant association between the trajectory groups and the average daily step counts during the first 30 days collected by smartwatch after adjusting for age, sex, and average daily wear hour (Table 2). On average, the participants in the moderate symptom group walked 945 fewer daily steps (95% CI –1573 to –318; P=.003) compared to the low symptom group. After we further adjusted for additional covariates, the association was attenuated: compared to the low symptom trajectory group, the participants in the moderate symptom group walked 733 fewer steps (95% CI –1344 to –121; P=.02), which was a 22% decrease in the magnitude of association (Table 2). The high symptom group, compared to the low symptom group, showed a decrease of 340 daily steps (95% CI –931 to 251; P=.26) in the base model that adjusted for age, sex, and average daily wear hour, and displayed a decrease of 208 daily steps (95% CI –790 to 373; P=.48) in the multivariable-adjusted model, but these differences were not statistically significant.

In the longitudinal study, we assessed the association between trajectory groups and the longitudinal PAI values collected by smartphone app multiple times within 1 year (Table 3). In the base model that adjusted for age, sex, and number of days, the difference in longitudinal PAI comparing the high symptom group to low symptom group was –0.7 and nonsignificant (95% CI –1.77 to 0.37; P=.20). After adjusting for relevant covariates, we observed a significant decrease of 1.09 in longitudinal PAI for the high symptom group compared to the low symptom group (95% CI –2.16 to –0.01; P=.048). In addition, the difference in the longitudinal PAI between moderate symptom group and low symptom group was not significant (mean difference –0.98, 95% CI –2.11 to 0.15; P=.09) after adjusting for all covariates.

Furthermore, we observed that the trajectory groups were significantly associated with the longitudinal daily step counts collected within 1-year of follow-up. The moderate symptom group, on average, had 1080 fewer daily steps than the low symptom group (95% CI –1695 to –465; P=.001) after adjusting for sex, age, average daily wear hour, and number of days. Adjusting for additional covariates attenuated the association: participants in the moderate symptom group walked 823 fewer daily steps on average compared to those in the low symptom group (95% CI –1421 to –226; P=.007). The high symptom group walked 586 steps fewer than the low symptom group in the longitudinal step count analysis in the base model that adjusted for sex, age, average daily wear hour, and number of days (95% CI –1161 to –10; P=.046). However, the difference was attenuated and no longer significant after adjusting for all covariates (mean difference –374, 95% CI –941 to 193; P=.20).

Figure 4. Correlation between the first app-based PAI and step counts during first 30 days in electronic Framingham Heart Study (eFHS). PAI: physical activity index.

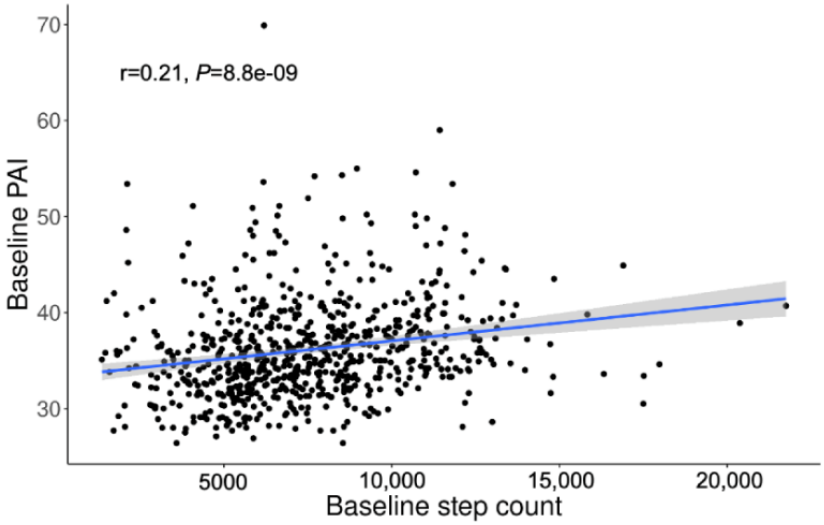


Table 2. Association between trajectory groups and baseline physical activity outcomes.

Outcome	Model 1[a]				Model 2[b]			
	High symptom vs low symptom, mean difference (95% CI)	*P* value	Moderate symptom vs low symptom, mean difference (95% CI)	*P* value	High symptom vs low symptom, mean difference (95% CI)	*P* value	Moderate symptom vs low symptom, mean difference (95% CI)	*P* value
Physical activity index	–0.83 (–2.05 to 0.38)	.18	–1.09 (–2.40 to 0.21)	.10	–1.29 (–2.52 to –0.06)	.04	–1.08 (–2.38 to 0.22)	.10
Daily step counts	–340 (–931 to 251)	.26	–945 (–1573 to –318)	.003	–208 (–790 to 373)	.48	–733 (–1344 to –121)	.02

[a]Model 1 covariates include sex, age, and daily wear hour (only for outcome daily step counts) at examination 3.

[b]Model 2 covariates include sex, age, daily wear hour (only for outcome daily step counts), BMI, smoking status, prevalent cardiovascular disease, current diabetes status, education, marital status, hypertension status, fasting blood glucose, high-density lipoprotein cholesterol, total cholesterol, triglycerides, and lipid treatment at examination 3.

Table 3. Association between trajectory groups and longitudinal physical activity outcomes during the 1-year follow-up.

Outcome	Model 1[a]				Model 2[b]			
	High symptom vs low symptom, mean difference (95% CI)	*P* value	Moderate symptom vs low symptom, mean difference (95% CI)	*P* value	High symptom vs low symptom, mean difference (95% CI)	*P* value	Moderate symptom vs low symptom, mean difference (95% CI)	*P* value
Physical activity index	–0.70 (–1.77 to 0.37)	.20	–1.03 (–2.17 to 0.12)	.08	–1.09 (–2.16 to –0.01)	.048	–0.98 (–2.11 to 0.15)	.09
Daily step counts	–586 (–1161 to –10)	.046	–1080 (–1695 to –465)	.001	–374 (–941 to 193)	.20	–823 (–1421 to –226)	.007

[a]Model 1 covariates at examination 3 include sex and age; covariates during 1-year follow-up include daily wear hour (for outcome daily step counts only) and number of days.

[b]Model 2 covariates at examination 3 include sex, age, daily wear hour (only for outcome daily step counts), BMI, smoking status, prevalent cardiovascular disease, current diabetes status, education, marital status, hypertension status, fasting blood glucose, high-density lipoprotein cholesterol, total cholesterol, triglycerides, lipid treatment, and number of days.

Discussion

Principal Findings

In this study, we constructed depressive symptom trajectory groups based on CES-D scores and antidepressant medication use collected from the FHS research exams. We identified 3 distinct trajectory groups using data from the 3 research exams conducted over a 14-year span prior to the initiation of eFHS. We observed that the low symptom group exhibited consistently low CES-D scores, with an extremely low proportion of participants reporting antidepressant medication use across all 3 exams. On the other hand, the moderate symptom group comprised participants with intermediate CES-D scores, who showed the highest and increasing likelihood of reporting antidepressant medications from exam 1 to exam 3. Last, the high symptom group consisted of participants with the highest CES-D scores that met the threshold for the presence of depressive symptoms. The proportion of participants reporting antidepressant medications in this group fell between the other 2 groups and was observed to be increasing across examinations. We further evaluated the associations of depressive symptom trajectory groups with self-reported PAI and objectively measured step counts collected from eFHS. In models adjusting for important confounders, compared to the low symptom group, the participants in the high symptom group had significantly lower self-reported PAI values collected from the eFHS smartphone app at first return and within the 1-year follow-up. In addition, the participants in the moderate symptom group walked significantly fewer average daily steps during the first 30 days and within the 1-year follow-up when compared to the low symptom group after accounting for important covariates. The moderate symptom group reported the highest antidepressant medication use with more than 90% of participants reporting use at the last timepoint. Our finding is consistent with another study reporting an association between antidepressant use and lower levels of physical activity among women with diabetes [26].

This study differs from most of the previous studies in the following 3 ways. First, the participants in eFHS have been followed for 14 years prior to their participation in eFHS. This enabled us to study the trajectory of depressive symptoms and antidepressant medication use over the course of 14 years and 3 examinations prior to the eFHS. In contrast, few previous studies considered longitudinal trends of depression and depressive symptoms including subclinical thresholds. Most previous studies performed an association of physical activity with depression at one time point [27,28]. Some other studies focused on the association between average levels of depression and physical activity in a longitudinal design with depression at a few time points as the predictor [11,29]. Furthermore, this

study used both self-reported PAI and objectively measured step counts within the 1-year follow-up in a community sample. These 2 physical activity measures captured different aspects of a participant's physical activity level, and they each have advantages and disadvantages. The smartwatch-collected step counts data provided a more objective measure of physical activity than self-reported questionnaires that were used in most of the previous studies, which can be affected by many factors, such as social desirability or recall bias. Compared to the questionnaires used in previous studies, our app-based PAI surveys allowed the participants to complete them in a home-based setting, and they were deployed multiple times within the 1-year period. Compared to the step count data, the PAI scores also quantified the amount and intensity of physical activity performed by an individual over a specific period, thus providing a more comprehensive evaluation of an individual's overall physical activity level [30]. Last, the integration of both CES-D scores and antidepressant medication across 3 research exams in trajectory modeling gave more complete information than using CES-D scores only. The classification of participants in trajectory groups did not depend solely on 1 exam or 1 variable, instead, the classification of participants revealed the long-term underlying patterns across exams and reflected both depressive symptoms and antidepressant medication use.

The main findings in this study were consistent with those from earlier studies, that is, depressive symptoms were associated with lower subsequent physical activity. The Whitehall study, the Coronary Artery Risk Development in Young Adults study, and other studies analyzed the association between depression at baseline and subsequent self-reported physical activity [11-13]. The 1958 British Birth Cohort used the depression phenotype at a few time points as an exposure of interest in a longitudinal setting to analyze the association between the average levels of depression and physical activity [11]. Previous findings also reported that both depressive symptoms and physical activity played an important role in promoting overall cardiovascular health and well-being. A previous study recruited 125 women from cardiac health screening events to investigate depressive symptoms and health-promoting lifestyle behaviors [31]. This study noted that higher levels of depressive symptoms were significantly associated with poorer health behaviors; early detection and treatment of depression were likely to improve quality of life [31]. Previous discoveries in affected science proposed a new theoretical framework called the upward spiral theory of lifestyle change, which explained how positive affect could contribute to long-term adherence to positive health behaviors [32]. A study that involved 204 patients during recovery from myocardial infarction reported that patients with depression after an acute myocardial infarction were more likely to fail to meet the recommended level of lifestyle changes; the failure to adopt a healthy lifestyle such as engaging in regular physical activity, in turn, contributed to a higher risk of cardiovascular events [33].

Our study represents an effective strategy to leverage health variables collected in traditional research exams with health outcomes collected by mHealth devices. First, the use of mHealth makes it possible for the participants to record their daily routines and lifestyles in a home-based setting, which is more representative of their true physiological behaviors. Therefore, mHealth provides the participants with more precise tracking of their health compared to traditional clinical health exams that record health at 1 point in time. Second, some studies reported that mHealth with an intervention on mood could improve physical activity, and subsequently, health outcomes. The BEHOLD-8 Controlled Clinical Trial examined the impact of a novel phone-delivered positive psychology-motivational interviewing intervention to promote physical activity in patients with type 2 diabetes. This study reported that the positive psychology-motivational interviewing intervention was feasible and well-accepted [34]. The Promoting Activity in Cardiac Patients via Text Messages pilot study identified the association between SMS text messaging focused on well-being and improvements in physical activity and mental health among 40 patients with prior acute coronary syndrome [35]. Therefore, the findings from our study and the previous studies suggest that mood interventions may be essential for mHealth studies that are designed to promote PA levels. Future mHealth studies to promote PA should intervene on mood through such strategies as improving sleep, counseling (well-based), supportive messaging, and activity promotion. Physical activity is a cornerstone of CVD prevention yet nearly half of Americans fail to meet physical activity guidelines. Addressing mood may hold promise to improve engagement in healthy behaviors such as increasing physical activity.

This study had several limitations. First, the sample size was substantively reduced from 1186 to 722 after applying the 3 exclusion criteria that were related to adherence. Low adherence has been a critical limitation in most mHealth studies [36,37]. Despite reduced adherence during the follow-up, the eFHS participants had higher adherence to smartwatch use compared to most mHealth studies [38]. In addition, this study was an observational study, and the causal relationship between depressive symptoms and physical activity cannot be determined although the depressive symptom trajectory groups were constructed prior to physical activity data. Furthermore, several previous studies have shown that physical activity and depressive symptoms have bidirectional associations. That is, people with depressive symptoms were more likely to have a lower level of physical activity at follow-up, and the lack of physical activity can predict an increased likelihood of depressive symptoms at follow-up. However, since we do not have CES-D scores collected in the health research center after mHealth physical activity, we are unable to investigate whether physical activity predicts CES-D scores at the next exam. Last, the eFHS participants were mostly White, middle-aged, with higher education, resided in New England, and were healthier than all participants attending examinations (Multimedia Appendix 1). Therefore, the generalization of the findings in this study to individuals in other age ranges, races or ethnicities, and different geographical areas may be limited.

Conclusions

We observed that antecedent depressive symptom trajectories were associated with subsequent physical activity collected in eFHS. Compared to the low symptom group with consistently low CES-D scores and proportions of antidepressant use, the participants in the high symptom group with high CES-D scores

and increasing likelihood of taking antidepressants had decreased levels of app-based PAI values. In addition, the participants in the moderate symptom group with intermediate CES-D scores and the highest reported antidepressant medication use walked fewer average daily step counts. Further exploration is needed to explore interventions aimed at enhancing mood, specifically through the use of mHealth technologies, with the goal of encouraging individuals to engage in regular physical activity.

Acknowledgments

This study was supported by an award from the Robert Wood Johnson Foundation (number 74624) and a grant from the National Heart Lung and Blood Institute (R01HL141434). The FHS is supported by a contract from the National Heart, Lung, and Blood Institute (principal investigator VR; 75N92019D00031), and investigator time from the following grants: R01HL126911 (EJB), 2R01 HL092577 (EJB), 1R01AG066010 (EJB), American Heart Association, 18SFRN34110082 (EJB), 2U54HL120163 (EJB), R01HL126911 (DDM), R01HL137734 (DDM), R01HL137794 (DDM), R01HL13660 (DDM), and U54-HL 143541 (DDM). The Apple Watches were provided to Boston University by Apple Inc at no cost to the study. Apple was not involved in the study design, analysis, interpretation, or reporting of study results.

Conflicts of Interest

DDM has received research support from Apple Inc, Bristol-Myers Squibb, Boehringer-Ingelheim, Pfizer, Flexcon, Samsung, Philips Health care, and Biotronik, and has received consultancy fees from Heart Rhythm Society, Bristol-Myers Squibb, Pfizer, Flexcon, Boston Biomedical Associates, and Rose Consulting. DDM also declares financial support for serving on the Steering Committee for the GUARD-AF study (NCT04126486) and Advisory Committee for the Fitbit Heart Study (NCT04176926). JMM was a guest lecturer or consultant for Merck Research Laboratories. VK is the principal, and CN is employees of CareEvolution, Inc, a health care technology company [39]. Starting in 2020, NLS received funding from Novo Nordisk for an investigator-initiated research grant unrelated to this study. Other authors have no relevant disclosures.

Multimedia Appendix 1
Comparison between included participants in our analysis and excluded participants.
[DOCX File , 22 KB - mental_v10i1e44529_app1.docx]

References

1. D'Agostino RB, Belanger AJ, Markson EW, Kelly-Hayes M, Wolf PA. Development of health risk appraisal functions in the presence of multiple indicators: the Framingham Study nursing home institutionalization model. Stat Med 1995 Aug 30;14(16):1757-1770. [doi: 10.1002/sim.4780141605] [Medline: 7481208]
2. Liu Q, He H, Yang J, Feng X, Zhao F, Lyu J. Changes in the global burden of depression from 1990 to 2017: findings from the Global Burden of Disease Study. J Psychiatr Res 2020;126:134-140 [FREE Full text] [doi: 10.1016/j.jpsychires.2019.08.002] [Medline: 31439359]
3. Murray CJL, Atkinson C, Bhalla K, Birbeck G, Burstein R, Chou D, Foreman, Lopez, Murray, Dahodwala, Jarlais, Famahi, U.S. Burden of Disease Collaborators. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. JAMA 2013;310(6):591-608 [FREE Full text] [doi: 10.1001/jama.2013.13805] [Medline: 23842577]
4. Ettman CK, Cohen GH, Abdalla SM, Sampson L, Trinquart L, Castrucci BC, et al. Persistent depressive symptoms during COVID-19: a national, population-representative, longitudinal study of U.S. adults. Lancet Reg Health Am 2022;5:100091 [FREE Full text] [doi: 10.1016/j.lana.2021.100091] [Medline: 34635882]
5. Teychenne M, Ball K, Salmon J. Physical activity and likelihood of depression in adults: a review. Prev Med 2008;46(5):397-411. [doi: 10.1016/j.ypmed.2008.01.009] [Medline: 18289655]
6. Lawlor DA, Hopker SW. The effectiveness of exercise as an intervention in the management of depression: systematic review and meta-regression analysis of randomised controlled trials. BMJ 2001;322(7289):763-767 [FREE Full text] [doi: 10.1136/bmj.322.7289.763] [Medline: 11282860]
7. Cooney GM, Dwan K, Greig CA, Lawlor DA, Rimer J, Waugh FR, et al. Exercise for depression. Cochrane Database Syst Rev 2013;2013(9):CD004366 [FREE Full text] [doi: 10.1002/14651858.CD004366.pub6] [Medline: 24026850]
8. Craft LL, Perna FM. The benefits of exercise for the clinically depressed. Prim Care Companion J Clin Psychiatry 2004;6(3):104-111 [FREE Full text] [doi: 10.4088/pcc.v06n0301] [Medline: 15361924]
9. Overdorf V, Kollia B, Makarec K, Alleva Szeles C. The relationship between physical activity and depressive symptoms in healthy older women. Gerontol Geriatr Med 2016;2:2333721415626859 [FREE Full text] [doi: 10.1177/2333721415626859] [Medline: 28138487]
10. Roshanaei-Moghaddam B, Katon WJ, Russo J. The longitudinal effects of depression on physical activity. Gen Hosp Psychiatry 2009;31(4):306-315. [doi: 10.1016/j.genhosppsych.2009.04.002] [Medline: 19555789]

11. Pinto Pereira SM, Geoffroy MC, Power C. Depressive symptoms and physical activity during 3 decades in adult life: bidirectional associations in a prospective cohort study. JAMA Psychiatry 2014;71(12):1373-1380 [FREE Full text] [doi: 10.1001/jamapsychiatry.2014.1240] [Medline: 25321867]
12. Azevedo Da Silva M, Singh-Manoux A, Brunner EJ, Kaffashian S, Shipley MJ, Kivimäki M, et al. Bidirectional association between physical activity and symptoms of anxiety and depression: the Whitehall II study. Eur J Epidemiol 2012;27(7):537-546 [FREE Full text] [doi: 10.1007/s10654-012-9692-8] [Medline: 22623145]
13. Zhang D, Pettee Gabriel K, Sidney S, Sternfeld B, Jacobs D, Whitaker KM. Longitudinal bidirectional associations of physical activity and depressive symptoms: the CARDIA study. Prev Med Rep 2021;23:101489 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101489] [Medline: 34336558]
14. Dias D, Paulo Silva Cunha J. Wearable health devices—vital sign monitoring, systems and technologies. Sensors (Basel) 2018;18(8):2414 [FREE Full text] [doi: 10.3390/s18082414] [Medline: 30044415]
15. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, et al. The third generation cohort of the national heart, lung, and blood institute's Framingham heart study: design, recruitment, and initial examination. Am J Epidemiol 2007;165(11):1328-1335 [FREE Full text] [doi: 10.1093/aje/kwm021] [Medline: 17372189]
16. McManus DD, Trinquart L, Benjamin EJ, Manders ES, Fusco K, Jung LS, et al. Design and preliminary findings from a new electronic cohort embedded in the Framingham heart study. J Med Internet Res 2019;21(3):e12143 [FREE Full text] [doi: 10.2196/12143] [Medline: 30821691]
17. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for epidemiologic studies depression scale (CES-D) as a screening instrument for depression among community-residing older adults. Psychol Aging 1997;12(2):277-287. [doi: 10.1037//0882-7974.12.2.277] [Medline: 9189988]
18. Sardana M, Lin H, Zhang Y, Liu C, Trinquart L, Benjamin EJ, et al. Association of habitual physical activity with home blood pressure in the electronic Framingham heart study (eFHS): cross-sectional study. J Med Internet Res 2021;23(6):e25591 [FREE Full text] [doi: 10.2196/25591] [Medline: 34185019]
19. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial. JAMA Intern Med 2017;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
20. Tan ZS, Spartano NL, Beiser AS, DeCarli C, Auerbach SH, Vasan RS, et al. Physical activity, brain volume, and dementia risk: the Framingham study. J Gerontol A Biol Sci Med Sci 2017;72(6):789-795 [FREE Full text] [doi: 10.1093/gerona/glw130] [Medline: 27422439]
21. Spartano NL, Davis-Plourde KL, Himali JJ, Murabito JM, Vasan RS, Beiser AS, et al. Self-reported physical activity and relations to growth and neurotrophic factors in diabetes mellitus: the Framingham offspring study. J Diabetes Res 2019;2019:2718465 [FREE Full text] [doi: 10.1155/2019/2718465] [Medline: 30729134]
22. Kannel WB, Belanger A, D'Agostino R, Israel I. Physical activity and physical demand on the job and risk of cardiovascular disease and death: the Framingham study. Am Heart J 1986;112(4):820-825. [doi: 10.1016/0002-8703(86)90480-1] [Medline: 3766383]
23. Nagin DS, Jones BL, Passos VL, Tremblay RE. Group-based multi-trajectory modeling. Stat Methods Med Res 2018;27(7):2015-2023. [doi: 10.1177/0962280216673085] [Medline: 29846144]
24. Andruff H, Carraro N, Thompson A, Gaudreau P, Louvet B. Latent class growth modelling: a tutorial. Tutor Quant Methods Psychol 2009;5(1):11-24 [FREE Full text] [doi: 10.20982/tqmp.05.1.p011]
25. Zhang Y, Pathiravasan CH, Hammond MM, Liu H, Lin H, Sardana M, et al. Comparison of daily routines between middle-aged and older participants with and those without diabetes in the electronic Framingham heart study: cohort study. JMIR Diabetes 2022;7(1):e29107 [FREE Full text] [doi: 10.2196/29107] [Medline: 34994694]
26. Danhauer SC, Brenes GA, Levine BJ, Young L, Tindle HA, Addington EL, et al. Variability in sleep disturbance, physical activity and quality of life by level of depressive symptoms in women with type 2 diabetes. Diabet Med 2019;36(9):1149-1157 [FREE Full text] [doi: 10.1111/dme.13878] [Medline: 30552780]
27. Juarbe TC, Gutiérrez Y, Gilliss C, Lee KA. Depressive symptoms, physical activity, and weight gain in premenopausal Latina and White women. Maturitas 2006;55(2):116-125. [doi: 10.1016/j.maturitas.2005.10.006] [Medline: 16846703]
28. Brummett BH, Babyak MA, Siegler IC, Mark DB, Williams RB, Barefoot JC. Effect of smoking and sedentary behavior on the association between depressive symptoms and mortality from coronary heart disease. Am J Cardiol 2003;92(5):529-532. [doi: 10.1016/s0002-9149(03)00719-7] [Medline: 12943871]
29. Panagiotakos DB, Pitsavos C, Lentzas Y, Skoumas Y, Papadimitriou L, Zeimbekis A, et al. Determinants of physical inactivity among men and women from Greece: a 5-year follow-up of the ATTICA study. Ann Epidemiol 2008;18(5):387-394. [doi: 10.1016/j.annepidem.2007.11.002] [Medline: 18083543]
30. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. Int J Behav Healthc Res 2011;2(4):320-332 [FREE Full text] [doi: 10.1504/IJBHR.2011.043414] [Medline: 25383095]
31. Savoy SM, Penckofer S. Depressive symptoms impact health-promoting lifestyle behaviors and quality of life in healthy women. J Cardiovasc Nurs 2015;30(4):360-372. [doi: 10.1097/JCN.0000000000000158] [Medline: 24850376]
32. Van Cappellen P, Rice EL, Catalino LI, Fredrickson BL. Positive affective processes underlie positive health behaviour change. Psychol Health 2018;33(1):77-97 [FREE Full text] [doi: 10.1080/08870446.2017.1320798] [Medline: 28498722]

33. Ziegelstein RC, Fauerbach JA, Stevens SS, Romanelli J, Richter DP, Bush DE. Patients with depression are less likely to follow recommendations to reduce cardiac risk during recovery from a myocardial infarction. Arch Intern Med 2000;160(12):1818-1823 [FREE Full text] [doi: 10.1001/archinte.160.12.1818] [Medline: 10871976]
34. Huffman JC, Golden J, Massey CN, Feig EH, Chung WJ, Millstein RA, et al. A positive psychology-motivational interviewing intervention to promote positive affect and physical activity in type 2 diabetes: the BEHOLD-8 controlled clinical trial. Psychosom Med 2020;82(7):641-649 [FREE Full text] [doi: 10.1097/PSY.0000000000000840] [Medline: 32665479]
35. Legler S, Celano CM, Beale EE, Hoeppner BB, Huffman JC. Use of text messages to increase positive affect and promote physical activity in patients with heart disease: the promoting activity in cardiac patients via text messages (PACT) pilot study. Curr Psychol 2020;39:648-655 [FREE Full text] [doi: 10.1007/s12144-018-9785-y] [Medline: 32982125]
36. Chan YFY, Wang P, Rogers L, Tignor N, Zweig M, Hershman SG, et al. The asthma mobile health study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017;35(4):354-362 [FREE Full text] [doi: 10.1038/nbt.3826] [Medline: 28288104]
37. McConnell MV, Shcherbina A, Pavlovic A, Homburger JR, Goldfeder RL, Waggot D, et al. Feasibility of obtaining measures of lifestyle from a smartphone app: the MyHeart counts cardiovascular health study. JAMA Cardiol 2017;2(1):67-76 [FREE Full text] [doi: 10.1001/jamacardio.2016.4395] [Medline: 27973671]
38. Pathiravasan CH, Zhang Y, Trinquart L, Benjamin EJ, Borrelli B, McManus DD, et al. Adherence of mobile app-based surveys and comparison with traditional surveys: eCohort study. J Med Internet Res 2021;23(1):e24773 [FREE Full text] [doi: 10.2196/24773] [Medline: 33470944]
39. Care Evolution. URL: https://careevolution.com [accessed 2023-06-24]

Abbreviations

CES-D: Center for Epidemiological Studies-Depression
CVD: cardiovascular disease
eFHS: electronic Framingham Heart Study
FHS: Framingham Heart Study
Gen 3: third generation
mHealth: mobile health
NOS: New Offspring Spouse
PAI: physical activity index
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by J Torous; submitted 22.11.22; peer-reviewed by D Brent, A AL-Asadi; comments to author 13.03.23; revised version received 02.06.23; accepted 05.06.23; published 14.07.23.

Please cite as:
Wang X, Pathiravasan CH, Zhang Y, Trinquart L, Borrelli B, Spartano NL, Lin H, Nowak C, Kheterpal V, Benjamin EJ, McManus DD, Murabito JM, Liu C
Association of Depressive Symptom Trajectory With Physical Activity Collected by mHealth Devices in the Electronic Framingham Heart Study: Cohort Study
JMIR Ment Health 2023;10:e44529
URL: https://mental.jmir.org/2023/1/e44529
doi:10.2196/44529
PMID:37450333

Original Paper

Examining Predictors of Depression and Anxiety Symptom Change in Cognitive Behavioral Immersion: Observational Study

Iony D Ezawa[1], PhD; Steven D Hollon[2], PhD; Noah Robinson[2], MSc

[1]Department of Psychology, University of Southern California, Los Angeles, CA, United States

[2]Department of Psychology, Vanderbilt University, Nashville, TN, United States

Corresponding Author:
Iony D Ezawa, PhD
Department of Psychology
University of Southern California
3620 McClintock Ave
Los Angeles, CA, 90089
United States
Phone: 1 213 740 2203
Email: ezawa@usc.edu

Abstract

Background: Depressive and anxiety disorders are the most common mental disorders, and there is a critical need for effective, affordable, and accessible interventions. Cognitive Behavioral Immersion (CBI) is a novel group-based cognitive behavioral skills training program delivered by lay coaches in the metaverse that can be accessed through various modalities including virtual reality (VR) head-mounted displays or flat-screen devices. Combining its ability to offer empirically supported therapy skills in a digital setting that can still facilitate interpersonal variables (eg, working alliance and sense of social support) with the aid of lay coaches, CBI has the potential to help fill this critical need.

Objective: This study had 2 primary aims. First, we aimed to examine changes in depression and anxiety symptoms in a sample of individuals who participated in CBI. Second, we aimed to examine 2 interpersonal process variables (working alliance and web-based social support) as predictors of symptom changes. We predicted CBI participants would experience depression and anxiety symptom improvements and that such improvements would be associated with an increase in both interpersonal process variables.

Methods: The study sample consists of 127 participants who endorsed clinical levels of depression or anxiety symptoms during their first CBI session and attended at least 2 sessions. Participants were asked to complete self-report measures of depression symptoms, anxiety symptoms, alliance, and web-based social support throughout their participation in CBI.

Results: Repeated measures ANOVAs determined that depression and anxiety symptom scores differed significantly across sessions (*P*s<.01). We also found participants' web-based social support predicted improvement in depression symptoms (*P*=.01), but neither the alliance nor web-based social support predicted change in anxiety symptoms (*P*s>.05). We also observed a significant difference in anxiety symptoms between participants who used a VR head-mounted display to access CBI and those who did not, such that participants who used VR head-mounted displays endorsed lower anxiety symptoms than those who did not at nearly every session (*P*=.04).

Conclusions: Participation in CBI is associated with both depression and anxiety symptom improvement. Web-based social support may play an important role in fostering changes in depression symptoms. Future studies are encouraged to continue examining the process of change in CBI with special attention paid to methods that can elucidate causal mechanisms of change.

(JMIR Ment Health 2023;10:e42377) doi:10.2196/42377

KEYWORDS

Cognitive Behavioral Immersion; virtual reality; metaverse; alliance; social support; cognitive behavioral; depression; anxiety; mood; mental health; mobile phone

Introduction

Depressive and anxiety disorders are the most common mental disorders, with an annual estimated 280 million people living with a depressive disorder and 301 million people living with an anxiety disorder [1]. Symptoms of depression and anxiety contribute to detriments in well-being, health, and functioning, and rank among the leading public health burdens worldwide [2]. While there are effective treatments for each, only a portion of those with these conditions receive treatment. There is a critical need for effective, affordable, and accessible interventions.

Cognitive behavioral therapy (CBT) is one of the most effective tools for the treatment of depression and anxiety. CBT is based on a model of how one's thoughts, emotions, behaviors, and physiology interact in one's environment [3]. According to the model, a person can indirectly change their emotions by changing maladaptive cognitions, behaviors, or physiological states. In this approach, therapists teach clients a host of cognitive and behavioral tools to identify, evaluate, and correct inaccurate negative thoughts, and increase behaviors that build efficacy and agency and boost positive affect [4]. In the treatment of depression, CBT has been rigorously tested in clinical trials and has been shown to be as efficacious as medications [5] and to have an enduring effect that lasts beyond the end of treatment [6]. CBT has also been shown to be efficacious and have enduring effects in the treatment of anxiety disorders [7]. Despite their efficacy, CBT-oriented interventions are often very expensive and difficult to access. Further, because of the stigma associated with psychological conditions, many individuals do not present for help [8].

To address problems of affordability and accessibility, the field is increasingly using lay therapists (also commonly referred to as coaches or peers) to provide cognitive behavioral coaching. Investigations of lay therapists provide empirical support that well-trained coaches can be trained to deliver effective cognitive behavioral interventions [9-11]. The use of lay therapists offers a way to reach millions of affected individuals who might otherwise not receive any care.

There also is increasing evidence suggesting that technology can be used to deliver affordable and accessible treatments with outcomes comparable to traditional face-to-face psychotherapy interventions [12]. In line with this, digital treatments are increasingly being developed and used as the "new normal" in a pandemic and postpandemic world. While some digital CBT programs have come in the form of self-guided computerized programs that waive the need for therapists altogether, a recent meta-analysis reported that the efficacy and engagement in digital CBT programs are enhanced when provided with coach- or peer-support via email or phone communication [12].

Opportunely, internet-connected virtual environments have now been developed that allow participants to interact with one another as avatars. These virtual environments are often referred to as the "metaverse" [13]. Most of these environments are videogames (eg, Fortnite, Minecraft, Roblox, SecondLife, and World of Warcraft) accessed through flat screens such as mobile phones or desktop computers. More recently, the proliferation of consumer virtual reality (VR) head-mounted displays extends experiences in the metaverse by further immersing individuals into virtual environments in which they can think, feel, and behave as if the virtual environment they are in is real [14]. Moreover, the use of VR to connect to the metaverse can simultaneously promote a sense of presence (ie, the subjective experience of being in a place or environment when one is physically situated in another) [15] and maintain a sense of anonymity via use of an avatar. In other words, participants can still reap the benefits of feeling as if they are in a shared social space with others while also benefitting from a veil of anonymity, which can keep them safe from the perceived judgment from others that may otherwise prevent them from trying out new skills, meeting new people, and taking therapeutic risks such as disclosing their own vulnerabilities.

Although there is a robust literature that demonstrates evidence for the use of VR without connecting to the metaverse in mental health treatment (specifically in treating posttraumatic stress disorder and specific phobias) [16,17], there remains a gap in the literature exploring the use of VR metaverse interventions for any mental health concern. Immersion into the metaverse may be an especially powerful modality to facilitate mental health programs. Besides acting as a digital medium for providing empirically supported cognitive behavioral skills training, the metaverse may help facilitate the nonspecific effects that help account for symptom change in traditional in-person psychological treatments [18]. Specifically, the interpersonal nature of the metaverse may help foster a working alliance between coaches and participants as well as nurture a broader sense of web-based social support. Each of these interpersonal elements has been shown to play a key role in promoting therapeutic outcomes in traditional in-person interventions [19,20], and the metaverse may be a uniquely positioned setting in which each of these components can be introduced digitally to help reduce depression and anxiety symptoms.

Cognitive behavioral immersion (CBI; accessed via the Innerworld app developed by Innerworld, Inc) is a novel synchronous cognitive behavioral skills training group delivered by coaches in the metaverse [13]. CBI was developed to combine the strengths of recent psychological and technological advances to potentially deliver efficacious, affordable, accessible, and scalable mental health support. Upon entering the Innerworld app, participants create an anonymous username and embody a custom, self-designed avatar to attend live events that are hosted by coaches trained in CBI. Participants communicate with one another by speaking out loud, and the audio is transmitted over the internet in real time. Besides voice communication, other social cues are communicated through head movement, hand movement, and emojis (used to communicate affect). CBI is accessible through various modalities, including VR technology (ie, Oculus Quest head-mounted displays) and non-VR modalities (ie, mobile phones and desktop computers).

In this study, we had 2 primary aims. First, we aimed to examine changes in depression and anxiety symptoms among a group of individuals who participated in CBI. Given the novelty of the interpersonal aspects of the metaverse, our second aim was focused on examining 2 interpersonal variables (working

alliance and web-based social support) as predictors of symptom change. In line with our first aim, we predicted participants would experience depression and anxiety symptom improvement throughout their participation in CBI. In line with our second aim, we predicted higher ratings of working alliance and perceived sense of web-based social support would be related to greater symptom improvement. In addition to these primary aims, we explored whether the magnitude of symptom change differed based on the type of device (VR vs non-VR) participants used.

Methods

Participants

Participants were selected based on the following criteria: (1) participated in at least 2 CBI sessions, (2) completed the symptom measures, and (3) reported a clinical level of depression or anxiety symptoms at their first CBI session (as evidenced by scoring a 10 or greater on the 9-item Patient Health Questionnaire (PHQ-9) [21] or 7-item Generalized Anxiety Disorder scale (GAD-7) [22]. A total of 127 participants met the study criteria out of the larger number of 468 who participated in at least 1 session of CBI. Specifically, 107 participants met the study criteria while reporting a clinical level of depressive symptoms, whereas 99 participants met study criteria while reporting a clinical level of anxiety symptoms (including 79 participants who reported clinical levels of both). See Figure 1 for a flowchart of participant recruitment.

Collectively, these participants were predominantly White, single, and employed males aged between 30 and 39 years. Participants attended an average of 13.80 (SD 26.61; median 6; mode 2) sessions, and approximately 79% (100/127) of the sample accessed CBI using a VR head-mounted display. See Table 1 for additional information on participant characteristics.

Figure 1. Flowchart of study participant recruitment. CBI: Cognitive Behavioral Immersion; GAD-7: 7-item Generalized Anxiety Disorder scale; PHQ-9: 9-item Patient Health Questionnaire.

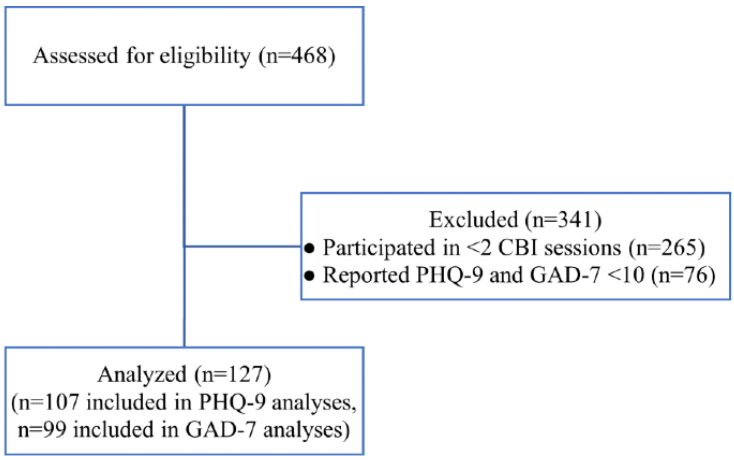


Table 1. Participant characteristics.

Variable	Participants, n (%)
Age (years)	
18-20	16 (12.6)
21-29	29 (22.8)
30-39	35 (27.6)
40-49	27 (21.3)
50-59	11 (8.7)
60-69	3 (2.4)
No response	6 (4.7)
Gender	
Female	43 (33.9)
Male	68 (53.5)
Other	9 (7.1)
No response	7 (5.5)
Race	
American Indian or Alaskan	2 (1.6)
Asian	13 (10.2)
Black or African American	5 (3.9)
Other	10 (7.9)
White	89 (70.1)
No response	8 (6.3)
Marital status	
Divorced	10 (7.9)
Married	31 (24.4)
Separated	3 (2.4)
Single	75 (59.1)
Widowed	1 (0.8)
No response	7 (5.5)
Employment status	
Disabled, not able to work	9 (7.1)
Full-time employed	53 (41.7)
Part-time employed	21 (16.5)
Not employed, searching for work	11 (8.7)
Not employed, not searching for work	23 (18.1)
Retired	2 (1.6)
No response	8 (6.3)
Income (US $)	
20,000	38 (29.9)
20,000-49,999	39 (30.7)
50,000-74,999	20 (15.8)
≥75,000	17 (13.4)
No response	13 (10.2)
Main device used to access CBI[a]	

Variable	Participants, n (%)
VR head-mounted display	100 (78.7)
Non-VR device (phone or computer)	27 (21.3)
Sessions attended	
2	23 (18.1)
3	15 (11.8)
4	14 (11.0)
5	11 (8.7)
6	4 (3.2)
7	6 (4.7)
8	6 (4.7)
9	6 (4.7)
10	5 (3.9)
≥11	37 (29.1)

[a]CBI: Cognitive Behavioral Immersion.

Measures

Demographics Survey

Participants were asked to complete a self-report survey assessing a variety of demographic variables including age, gender, race, marital status, employment status, and annual household income. This survey was administered upon entering the Innerworld app for the first time.

PHQ-9

The PHQ-9 [21] is a 9-item self-report measure based on the diagnostic criteria for major depressive disorder from the Diagnostic and Statistical Manual Fourth Edition. Each item is scored on a 0- to 3-point basis, and total scores can range from 0 to 27 with greater scores indicating greater severity of depression symptoms. A cutoff score of at least 10 has been recommended to detect clinical levels of depression symptom severity [23]. Participants were asked to complete this measure after each CBI session. Internal consistency for PHQ-9 scores in this study was acceptable (α=.72 at the first session).

GAD-7

The GAD-7 [22] is a 7-item self-report measure based on the diagnostic criteria for generalized anxiety disorder from the Diagnostic and Statistical Manual Fourth Edition. It has been used in primary care and research settings to assess anxiety severity, symptomatology, and functional impairment. Each item is again scored on a 0- to 3-point basis, and total scores can range from 0 to 21 with greater scores indicating greater severity of anxiety symptoms. As on for the PHQ-9, a total cutoff score of at least 10 has been recommended to detect clinical levels of anxiety symptom severity [22]. Participants again were asked to complete this measure after each CBI session. Internal consistency for GAD-7 scores in this study was acceptable (α=.71 at the first session).

Group Session Ratings Scale

The Group Session Ratings Scale (GSRS) is a 4-item self-report measure of the working alliance measured on 10-point sliding scale [24]. Total scores can range from 4 to 40 with higher scores indicating greater levels of the group alliance. It has demonstrated acceptable validity and reliability [24,25]. Participants again were asked to complete this measure after each CBI session. Internal consistency for GSRS scores in this study was good (α=.85 at the first session).

Online Social Support Scale

The Online Social Support Scale (OSSS) is a 40-item self-report measure of web-based social support one has received from sessions [26]. We removed 11 items that did not apply to the current intervention, including items that covered instrumental support (eg, the provision of financial aid, material resources, and needed services). Therefore, participants completed a 29-item version of the OSSS, measuring constructs of esteem or emotional support, social companionship, and informational support. Each item is scored from 1 to 5, and total scores on our 29-item version could range from 29 to 145 with higher scores reflecting greater web-based social support. The OSSS has demonstrated evidence of reliability and validity [26]. Participants were asked to complete this measure only once per month. Internal consistency for OSSS scores in this study was excellent (α=.98 at the first session).

CBI Intervention and Coach Training

CBI was conducted in a metaverse app called Innerworld, which was developed by Innerworld, Inc. [13]. Participants could connect to Innerworld with an Oculus Quest VR head-mounted display, a desktop computer, or mobile phone. Participants are depicted through customizable avatars with anonymous usernames. Participants can view each other's avatars and body movements while communicating with their actual voices.

The CBI program was created based on the principles of the cognitive behavioral model [4]. Innerworld, Inc handled all

recruitment and training of coaches providing CBI. All coaches had to be at least 18 years of age, but no prior professional training in psychology or therapy was required for coaches prior to undergoing CBI training. Coaches received a minimum of 18 hours of intensive training in cognitive behavioral principles, cognitive behavioral techniques, basic psychological skills (eg, using open-ended questions and developing an empathetic tone), and ethics training. Coaches also participated in ongoing weekly supervision. Each CBI session began with a "check-in" during which each participant had the opportunity to share a recent challenge and success followed by a discussion on a cognitive behavioral tool. CBI sessions were offered daily, and participants were encouraged to attend as many sessions as they liked.

Procedures

No specific recruitment efforts took place to recruit participants for this study. Upon joining Innerworld, participants were provided a tutorial that explained the app's routine data collection and monitoring for the purposes of quality improvement and research. The tutorial also noted that the app had funding from the National Institutes of Health. Following the tutorial and before being able to participate in CBI, participants were required to provide consent to these data collection efforts and to agree to Innerworld, Inc's privacy policy and terms of service. In addition to following the tutorial, applicants were asked to complete an optional demographics survey that appeared as a pop-up window within the Innerworld app.

CBI sessions were available daily, and participants were encouraged to attend as many CBI sessions as they liked. Symptom and interpersonal measures were distributed to participants through pop-up windows that appeared in the app interface. See Table 2 detailing the timing of measure distribution. No incentive was provided for the completion of measures. In addition to these self-report measures, the Innerworld app also automatically tracked the number of sessions attended and the type of device participants used to enter the app.

Table 2. Timing of measure distribution.

Measure	Time of distribution
Demographics survey	Upon entering Innerworld app
Symptom measures	
PHQ-9[a]	After each CBI[b] session
GAD-7[c]	After each CBI session
Interpersonal measures	
GSRS[d]	After each CBI session
OSSS[e]	Once per month

[a]PHQ-9: 9-item Patient Health Questionnaire.

[b]CBI: Cognitive Behavioral Immersion.

[c]GAD-7: 7-item Generalized Anxiety Disorder scale.

[d]GSRS: Group Session Ratings Scale.

[e]OSSS: Online Social Support Scale.

Ethical Considerations

Innerworld, Inc passed only deidentified data to this study's researchers for the purpose of quality improvement as determined by Vanderbilt University's institutional review board (#200327).

Analytic Strategy

Changes in Symptoms Across CBI

A repeated measures ANOVA was conducted to examine differences in symptom scores between each session on the intent-to-treat sample. We conducted a model for each of the 2 symptom measures (ie, PHQ-9 and GAD-7). Given that there was a considerable amount of variability in the number of sessions each participant completed and subsequently the number of measures completed, the session comparisons were limited in the repeated measures ANOVA to the first 10 sessions each participant may have completed. We chose this cutoff because most participants (n=90, 70.1%) completed 10 sessions or less. Based on the recommendation of statisticians analyzing the best ways to handle missing values when conducting an intent-to-treat analysis, it was decided to take a mixed model approach without any ad hoc imputation to run this analysis [27]. We report η_p^2 as a measure of effect size for these analyses. These models were conducted using PROC MIXED in Statistical Analysis System (SAS; version 9.4; SAS Institute).

Relation of Interpersonal Variables With Symptom Change

For our analyses concerning whether the interpersonal variables (ie, GSRS and OSSS) predicted change in symptoms, we first calculated residual change scores for each symptom measure. Residual change scores are calculated by taking the residual from a regression model in which first session symptom scores are entered as predictors of the most recent symptom scores. Given the difference in the number of times participants were asked to complete the GSRS and OSSS (ie, once per session for the GSRS and once per month for the OSSS), average GSRS

and OSSS scores were calculated for each participant. We then used general linear models to examine the interpersonal variables as predictors of symptom scores (measured as residual change scores) while controlling for the total number of sessions each participant completed. One model was run for each symptom measure. Models were run using PROC GLM in SAS 9.4.

Magnitude of Symptom Change Between VR and Non-VR Users

We conducted 2-group repeated measures ANOVAs to examine the effects of device type used to access CBI (VR vs non-VR device) and time (ie, session) on symptom scores. We ran 1 model for each symptom measure. Similar to our previous ANOVA models, the session comparisons were limited to the first 10 sessions. These models were conducted using PROC MIXED in SAS 9.4.

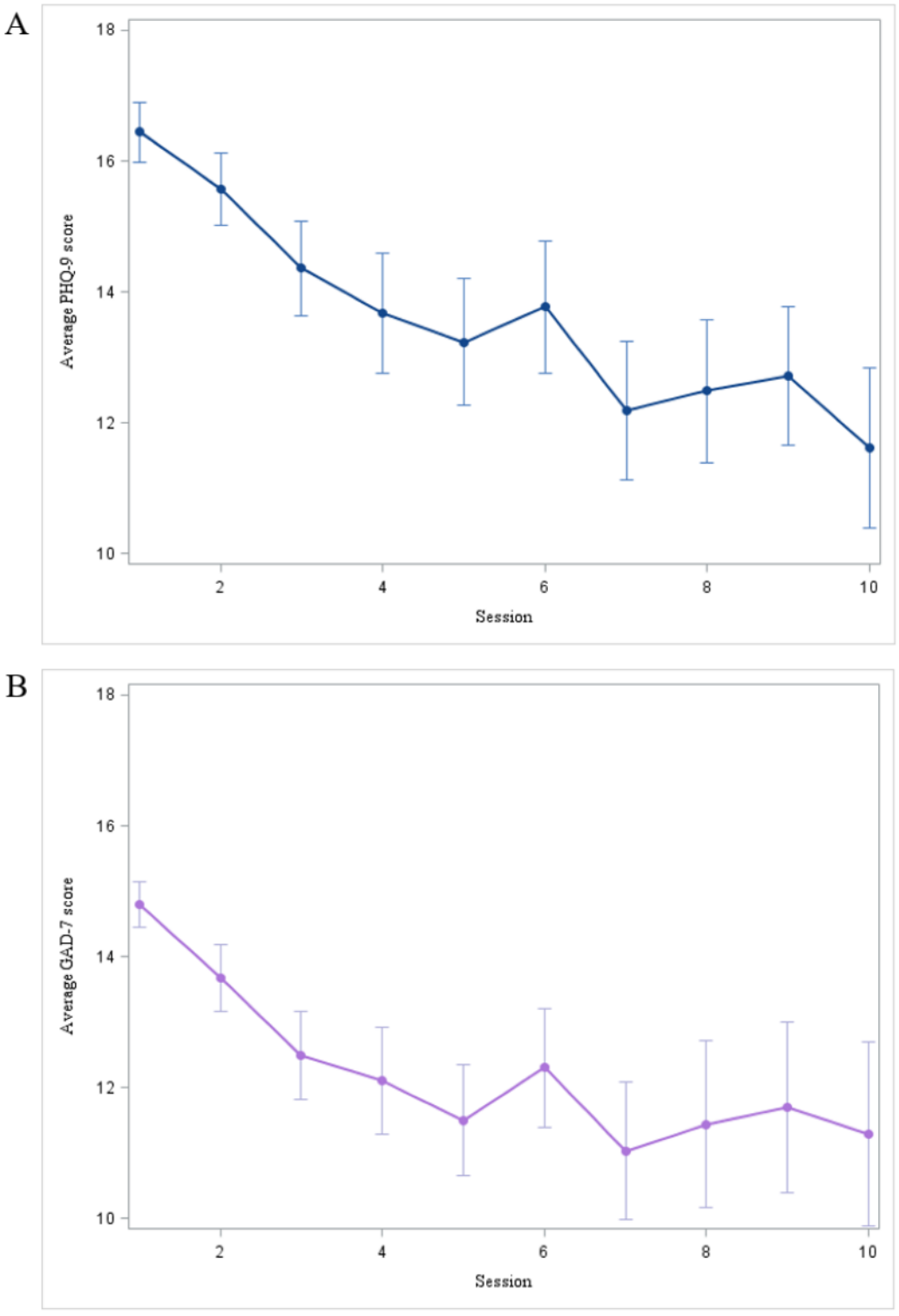
Results

Changes in Symptoms Across CBI

A repeated measures ANOVA determined that depression symptom scores differed significantly across sessions ($F_{9,422}$=5.19; P<.01; η_p^2=0.16). A post hoc pairwise comparison using the Tukey correction showed a significant decrease in PHQ-9 scores between a number of sessions, including a significant difference between participants' first and tenth session (P=.002). Notably, participants endorsed moderate to severe depression symptom severity at their first session (mean 16.44, SD 4.77; range 10-27) [21]. Participants' reports of depression symptoms were significantly lower at the tenth session (mean 11.62, SD 5.59; range 1-24).

A repeated measures ANOVA also showed that anxiety symptom scores differed significantly across sessions ($F_{9,416}$=4.45; P<.01; η_p^2=0.26). After applying the Tukey correction, we also found a significant decrease in GAD-7 scores between a number of sessions, including a significant difference between participants' first and tenth session (P<.01). Similar to the depression scores, participants also tended to endorse moderate to severe anxiety symptom severity at their first session (mean 14.79, SD 3.45; range 10-21) [22]. Participants' reports of anxiety symptoms were substantially lower at their 10th session (mean 11.29, SD 6.42; range 0-21). A table of the significant session comparisons from Tukey post hoc comparisons can be found in Multimedia Appendix 1. See Figure 2 for a graph charting average symptom scores across the first 10 sessions.

Figure 2. (A) Depression and (B) anxiety symptom change across cognitive behavioral immersion. GAD-7: 7-item Generalized Anxiety Disorder scale; PHQ-9: 9-item Patient Health Questionnaire.



Relation of Interpersonal Variables With Symptom Change

We next examined whether alliance or web-based social support (represented by average GSRS and OSSS scores across the course of treatment) predicted change in depression symptoms (represented by a residual change in PHQ-9 scores from participants' first to most recent session) while controlling for participants' total number of sessions attended. We found a significant effect for OSSS scores predicting change in PHQ-9 scores, such that greater OSSS scores predicted greater symptom improvement ($\beta_{standardized}$=–1.50; β=–.07; SE=0.03; t_{85}=–2.06; *P*=.04). However, GSRS scores did not significantly predict change in PHQ-9 scores ($\beta_{standardized}$=–.69; β=–.08; SE=0.08; t_{85}=–0.94; *P*=.35).

We next examined whether the alliance or web-based social support (averaged GSRS and OSSS scores across the course of treatment) predicted change in anxiety symptoms (represented by a residual change in GAD-7 scores from participants' first to most recent session). Neither OSSS ($\beta_{standardized}$=–.14; β=–.04; SE=0.03; t_{79}=–1.11; *P*=.27) nor GSRS scores predicted change in GAD-7 scores ($\beta_{standardized}$=.04; β=.02; SE=0.09; t_{79}=0.28; *P*=.78).

Magnitude of Symptom Change Between VR and Non-VR Users

We first conducted a 2-group repeated measures ANOVA to examine the effects of VR versus non-VR usage and session on PHQ-9 scores. We did not find a significant main effect of device type used on PHQ-9 scores ($F_{1,105}$=2.24; *P*=.14; η_p^2=0.002) or a significant interaction between device type used and session ($F_{9,413}$=1.79; *P*=.07; η_p^2=0.005). We then conducted a 2-group repeated measures ANOVA to examine the effects of VR versus non-VR usage and session on GAD-7 scores. The results of this analysis revealed a significant main effect of device type used on GAD-7 scores ($F_{1,97}$=4.14; *P*=.04; η_p^2=0.002), indicating that there was a significant difference in anxiety symptoms between participants who used a VR head-mounted display to access CBI versus to those who did not, regardless of time. Further examination of the least squares means revealed that, on average, participants who used VR head-mounted displays had lower anxiety symptoms than those who did not at nearly every session (the exception being session 7). Of note, we did not find a significant interaction between device type used and session ($F_{9,407}$=1.54; *P*=.13; η_p^2=0.04).

Discussion

Principal Findings

Given recent advances in technology, the metaverse has been transformed from a science fiction idea into a reality [28]. Because of its seemingly limitless possibilities, the metaverse is gaining attention across many fields including the mental health field [29]. Using the metaverse has the potential to improve the accessibility and affordability of mental health interventions while maintaining their fidelity and (most importantly) their efficacy. Several metaverse apps are being built to seize the opportunities of this technology, but there remains little regulation as to what they do or knowledge regarding their effectiveness. To our knowledge, this study is the first to evaluate symptom outcomes in a mental health intervention conducted in the metaverse. We found evidence of significant reduction in depressive (η_p^2=0.16) and anxiety symptoms (η_p^2=0.26) across CBI—a novel coach-led cognitive behavioral skills training group conducted in the metaverse. Given the absence of a control group, we cannot attribute these reductions to CBI, but their magnitude was impressive and, in the range, reported in the literature for bona fide treatments [30,31]. This is especially exciting given that the coaches who were trained to lead CBI sessions were not therapists and required no prior professional training in providing mental health services. We are preparing to conduct a randomized controlled trial to evaluate the efficacy of CBI, but these initial pilot data suggest that such a trial is worthy of pursuit.

The primary aim of this study was to examine symptom change across the course of CBI participation, but we were also interested in the process through which these effects might be achieved. One unique contribution of the metaverse as compared to many previous digital modalities is the ability to interact with others while engaging in the electronic app. In doing so, interpersonal elements that so often play an important role in psychological interventions (ie, nonspecific effects) can be maintained [18]. Therefore, we assessed participant's perception of the working alliance and web-based social support. We found that only participants' perceived sense of web-based social support was significantly related to depressive symptom improvement, while neither interpersonal variable predicted anxiety symptom change. Both the alliance and social support are similar in that they are interpersonal variables, but it appears that the relationship that is fostered with peers in the web-based community may play a more important role in symptom improvement as compared to the group alliance, particularly in the context of alleviating depression symptoms.

Another strength of CBI is its compatibility to be accessed with different devices, including VR head-mounted displays that can create a truly immersive experience for participants who enter the metaverse. The vast majority (n=100, 78.7%) of participants in the sample for this study used a VR head-mounted display to participate in CBI, while the remainder accessed CBI through the use of their smartphone or desktop computer. Although we did not randomly assign the device participants used, we found that participants who tended to use the VR head-mounted displays also reported a greater magnitude of anxiety symptom change on average than those who used non-VR devices at nearly every session. It may be possible that the level of immersion or extent to which a participant interacts with their virtual environment as if it was real [14] may be an important element of digital interventions for these symptoms. Future research is needed to further explore the effect that the extent of immersion may have on CBI outcomes and other metaverse mental health interventions. Outcomes from such research could be used to assess whether the increased cost of a head-mounted display may be warranted to lead to better outcomes.

Notably, our sample was predominantly male (n=68, 53.5%). This contrasts with the predominantly female composition of typical treatment-seeking samples [32,33]. Men are typically less likely to seek treatment than women, thus their mental health is often overlooked and undertreated. This is a critical problem, especially due to the troubling fact that depression and suicide are among the leading causes of death among men [34]. Thus, our sample composition is especially exciting because it suggests that the metaverse may be a particularly accessible and appealing setting for men to access mental health services, potentially increasing the reach of mental health intervention for this underserved population.

Limitations

We note some key limitations of this study. First, we used self-report measures of the alliance, web-based social support,

and symptoms. While each of the self-report measures used in this study has been validated, there are alternative measurement approaches that can provide more nuanced assessments of these variables. For example, researchers suggest that the congruence between multiple perspectives in rating the alliance may provide more useful information than a single perspective alone [35,36] and that observer-rated alliance may provide an even less biased perspective that is more related to outcomes [37]. Relatedly, we only assessed 2 interpersonal variables and 2 symptom measures. Given the transdiagnostic nature of the CBI program, additional process and outcome variables will be important to assess. We also believe it will be important to assess attitudes related to the provision of mental health services in the metaverse, as attitudes can affect interest in, engagement with, and expectations of the intervention. Second, we cannot establish causality of relations between variables in this study. Without a control group, we are unable to say with certainty that participation in CBI causes symptom change. Given the timing of assessments, we are unable to say whether the alliance or web-based social support occurred prior to the symptom change with which it was associated. It will be important for future studies to consider the timing of assessments. It will also be useful to consider confounding variables that might influence these process-outcome relations (such as educational levels of coaches and participant demographic variables). Third, our sample was comprised of predominantly White adults 50 years or younger of age. Therefore, there should be caution in generalizing these findings to other samples with respect to ethnicity or race and age (17 years or younger and 50 years or older).

Future Directions

Given the significant depression and anxiety symptom improvements our sample experienced, it will be important to continue examining the process of change so we can continue to promote and improve these outcomes. Nonspecific effects are one avenue through which symptom outcomes can be achieved [18]. However, specific factors such as cognitive or behavioral techniques are also frequently examined process variables that can elevate symptom change [38]. Although our research group is aware of the methodological challenges of rating process variables through rigorous traditional observer-coding projects [39], the digital nature of this intervention may open another future opportunity to automate and enhance the coding procedures of process variables. Future studies are encouraged to not only measure additional process variables but also to pay special attention to the timing of variables to ensure temporal precedence of variables in line with causal relations. This will allow us to better understand the process of change as well as elucidate mediators and moderators of process-outcome relations [40].

Moreover, it will be important for future researchers to investigate metaverse mental health interventions in more diverse samples. The CBI intervention can be accessed virtually anywhere at any time, so it has the potential to reach various populations who may otherwise not have access to empirically supported mental health services. More targeted recruitment of underserved populations is needed.

One exciting feature about the Innerworld app used in this study is that it was designed to be a clinical research platform that welcomes research collaborations to continue investigating these relations with sophisticated methods and continuous data collection. We encourage collaborative efforts with team expertise in both technology and therapeutic processes to continue the investigation into and enhancement of metaverse mental health interventions.

Conclusions

This study evaluated the symptom outcomes in a sample of individuals who participated in a coach-led cognitive behavioral skills training group conducted in the metaverse. These participants experienced significant depression and anxiety symptom improvements, and these improvements were related to higher levels of web-based social support. With the increasing popularity of the metaverse and VR, we strongly encourage researchers to continue investigating its potential for further accessibility, affordability, and efficacy of mental health interventions.

Acknowledgments

This work was supported, in part, by a National Institute of Mental Health training grant (T32 MH018921).

Conflicts of Interest

NR is the CEO of Innerworld, Inc, the company that created the platform that is the subject of this study and is the sole author with a conflict of interest. NR played no role in the data collection and analysis. The design and data management were conducted under the supervision of IDE and SDH, and data analysis was conducted by IDE.

Multimedia Appendix 1
Results from Tukey post hoc comparisons.
[DOCX File , 15 KB - mental_v10i1e42377_app1.docx]

References

1. World Health Organization. Mental disorders. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/mental-disorders [accessed 2023-06-15]

2. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. Lancet 2013;382(9904):1575-1586 [FREE Full text] [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
3. Beck A, Haigh EAP. Advances in cognitive theory and therapy: the generic cognitive model. Annu Rev Clin Psychol 2014;10:1-24 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032813-153734] [Medline: 24387236]
4. Beck A, Rush AJ, Shaw BF, Emery G. Cognitive Therapy of Depression. New York: Guilford Press; 1979.
5. DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, et al. Cognitive therapy vs medications in the treatment of moderate to severe depression. Arch Gen Psychiatry 2005;62(4):409-416 [FREE Full text] [doi: 10.1001/archpsyc.62.4.409] [Medline: 15809408]
6. Hollon SD, DeRubeis RJ, Shelton RC, Amsterdam JD, Salomon RM, O'Reardon JP, et al. Prevention of relapse following cognitive therapy vs medications in moderate to severe depression. Arch Gen Psychiatry 2005;62(4):417-422 [FREE Full text] [doi: 10.1001/archpsyc.62.4.417] [Medline: 15809409]
7. Cuijpers P, Sijbrandij M, Koole S, Huibers M, Berking M, Andersson G. Psychological treatment of generalized anxiety disorder: a meta-analysis. Clin Psychol Rev 2014;34(2):130-140 [FREE Full text] [doi: 10.1016/j.cpr.2014.01.002] [Medline: 24487344]
8. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015;45(1):11-27 [FREE Full text] [doi: 10.1017/S0033291714000129] [Medline: 24569086]
9. Singla DR, Weobong B, Nadkarni A, Chowdhary N, Shinde S, Anand A, et al. Improving the scalability of psychological treatments in developing countries: an evaluation of peer-led therapy quality assessment in Goa, India. Behav Res Ther 2014;60:53-59 [FREE Full text] [doi: 10.1016/j.brat.2014.06.006] [Medline: 25064211]
10. Patel V, Weobong B, Weiss HA, Anand A, Bhat B, Katti B, et al. The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. Lancet 2017;389(10065):176-185 [FREE Full text] [doi: 10.1016/S0140-6736(16)31589-6] [Medline: 27988143]
11. van Ginneken N, Tharyan P, Lewin S, Rao GN, Meera SM, Pian J, et al. Non-specialist health worker interventions for the care of mental, neurological and substance-abuse disorders in low- and middle-income countries. Cochrane Database Syst Rev 2013(11):CD009149 [FREE Full text] [doi: 10.1002/14651858.CD009149.pub2] [Medline: 24249541]
12. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
13. Robinson N, Mahapatra A, Jean-Baptiste B, Mallard A, Yang A, Hollon SD, et al. Cognitive Behavioral Immersion for Substance Use Disorders: A Feasibility and Pilot Study of a Peer-Based Coaching Program in the Metaverse. Games Health J 2023 Jun 09. [doi: 10.1089/g4h.2022.0214] [Medline: 37294540]
14. Slater M, Wilbur S. A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. Presence: Teleoperators Virtual Environ 1997;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
15. Witmer BG, Singer MJ. Measuring presence in virtual environments: a presence questionnaire. Presence: Teleoperators Virtual Environ 1998;7(3):225-240 [FREE Full text] [doi: 10.1162/105474698565686]
16. Botella C, Fernández-Álvarez J, Guillén V, García-Palacios A, Baños R. Recent progress in virtual reality exposure therapy for phobias: a systematic review. Curr Psychiatry Rep 2017;19(7):42 [FREE Full text] [doi: 10.1007/s11920-017-0788-4] [Medline: 28540594]
17. Kothgassner OO, Goreis A, Kafka JX, Van Eickels RL, Plener PL, Felnhofer A. Virtual reality exposure therapy for posttraumatic stress disorder (PTSD): a meta-analysis. Eur J Psychotraumatol 2019;10(1):1654782 [FREE Full text] [doi: 10.1080/20008198.2019.1654782] [Medline: 31489138]
18. Cuijpers P, Driessen E, Hollon SD, van Oppen P, Barth J, Andersson G. The efficacy of non-directive supportive therapy for adult depression: a meta-analysis. Clin Psychol Rev 2012;32(4):280-291 [FREE Full text] [doi: 10.1016/j.cpr.2012.01.003] [Medline: 22466509]
19. Leibert TW, Smith JB, Agaskar VR. Relationship between the working alliance and social support on counseling outcome. J Clin Psychol 2011;67(7):709-719 [FREE Full text] [doi: 10.1002/jclp.20800] [Medline: 21480227]
20. Mallinckrodt B. Change in working alliance, social support, and psychological symptoms in brief therapy. J Couns Psychol 1996;43(4):448-455 [FREE Full text] [doi: 10.1037/0022-0167.43.4.448]
21. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
22. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097 [FREE Full text] [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
23. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002;32(9):509-515 [FREE Full text] [doi: 10.3928/0048-5713-20020901-06]
24. Duncan BL, Miller SD, Sparks JA, Claud DA, Reynolds LR, Brown J, et al. The Session Rating Scale: preliminary psychometric properties of a “working” alliance measure. J Brief Ther 2003;3(1):3-12 [FREE Full text]

25. Quirk K, Miller S, Duncan B, Owen J. Group Session Rating Scale: preliminary psychometrics in substance abuse group interventions. Couns Psychother Res 2013;13(3):194-200 [FREE Full text] [doi: 10.1080/14733145.2012.744425]
26. Nick EA, Cole DA, Cho SJ, Smith DK, Carter TG, Zelkowitz RL. The Online Social Support Scale: measure development and validation. Psychol Assess 2018;30(9):1127-1143 [FREE Full text] [doi: 10.1037/pas0000558] [Medline: 29781664]
27. Chakraborty H, Gu H. A Mixed Model Approach for Intent-to-Treat Analysis in Longitudinal Clinical Trials With Missing Values. Research Triangle Park, NC: RTI Press; 2019.
28. Sparkes M. What is a metaverse. New Scientist 2021;251(3348):18 [FREE Full text] [doi: 10.1016/s0262-4079(21)01450-0]
29. Ifdil I, Situmorang DDB, Firman F, Zola N, Rangka IB, Fadli RP. Virtual reality in metaverse for future mental health-helping profession: an alternative solution to the mental health challenges of the COVID-19 pandemic. J Public Health (Oxf) 2023;45(1):e142-e143 [FREE Full text] [doi: 10.1093/pubmed/fdac049] [Medline: 35467746]
30. Cuijpers P, Smit F, Bohlmeijer E, Hollon SD, Andersson G. Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. Br J Psychiatry 2010;196(3):173-178 [FREE Full text] [doi: 10.1192/bjp.bp.109.066001] [Medline: 20194536]
31. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007;37(3):319-328 [FREE Full text] [doi: 10.1017/S0033291706008944] [Medline: 17112400]
32. Gonzalez JM, Alegría M, Prihoda TJ, Copeland LA, Zeber JE. How the relationship of attitudes toward mental health treatment and service use differs by age, gender, ethnicity/race and education. Soc Psychiatry Psychiatr Epidemiol 2011;46(1):45-57 [FREE Full text] [doi: 10.1007/s00127-009-0168-4] [Medline: 19921079]
33. Psychological therapies: annual report on the use of IAPT services—England, 2015-16. NHS Digital. 2015. URL: http://webarchive.nationalarchives.gov.uk/20180328133700/http://digital.nhs.uk/catalogue/PUB22110 [accessed 2023-06-15]
34. Leading causes of death—all races and origins—males—United states, 2016. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/healthequity/lcod/men/2016/all-races-origins/index.htm [accessed 2023-06-15]
35. Atzil-Slonim D, Bar-Kalifa E, Rafaeli W, Lutz W, Rubel J, Schiefele AK, et al. Therapeutic bond judgments: congruence and incongruence. J Consult Clin Psychol 2015;83(4):773-784. [doi: 10.1037/ccp0000015] [Medline: 25664641]
36. Whelen ML, Ezawa ID, Strunk DR. Congruence in patient and therapist-rated alliance predicts subsequent depressive symptoms in cognitive behavioral therapy for depression. Psychother Res 2022:1-10 [FREE Full text] [doi: 10.1080/10503307.2022.2156304] [Medline: 36525598]
37. Fenton LR, Cecero JJ, Nich C, Frankforter TL, Carroll KM. Perspective is everything: the predictive validity of six working alliance instruments. J Psychother Pract Res 2001;10(4):262-268 [FREE Full text] [Medline: 11696653]
38. Pfeifer B, Strunk D. A primer on psychotherapy process research: a review of cognitive therapy of depression. Aust Psychol 2020;50(6):411-415 [FREE Full text] [doi: 10.1111/ap.12147]
39. Ezawa ID, Bartels GC, Pfeifer BJ, Strunk DR. Pay no attention to the man behind the curtain: the overlooked issue of rater selection in psychotherapy observer ratings. Psychother Res 2021;31(4):548-556 [FREE Full text] [doi: 10.1080/10503307.2020.1780510] [Medline: 32552478]
40. Lorenzo-Luaces L, German RE, DeRubeis RJ. It's complicated: the relation between cognitive change procedures, cognitive change, and symptom change in cognitive therapy for depression. Clin Psychol Rev 2015;41:3-15 [FREE Full text] [doi: 10.1016/j.cpr.2014.12.003] [Medline: 25595660]

Abbreviations

CBI: Cognitive behavioral immersion
CBT: cognitive behavioral therapy
GAD-7: 7-item Generalized Anxiety Disorder scale
GSRS: Group Session Ratings Scale
OSSS: Online Social Support Scale
PHQ-9: 9-item Patient Health Questionnaire
SAS: Statistical Analysis System
VR: virtual reality

Edited by J Torous; submitted 01.09.22; peer-reviewed by L Trimpop; comments to author 24.01.23; revised version received 20.03.23; accepted 26.04.23; published 14.07.23.

Please cite as:
Ezawa ID, Hollon SD, Robinson N
Examining Predictors of Depression and Anxiety Symptom Change in Cognitive Behavioral Immersion: Observational Study
JMIR Ment Health 2023;10:e42377
URL: https://mental.jmir.org/2023/1/e42377
doi:10.2196/42377
PMID:37450322

Original Paper

A Novel and Integrated Digitally Supported System of Care for Depression and Anxiety: Findings From an Open Trial

Kate Wolitzky-Taylor[1], PhD; Richard LeBeau[2], PhD; Inna Arnaudova[1], PhD; Nora Barnes-Horowitz[2], MA; Elizabeth Gong-Guy[1], PhD; Scott Fears[1,3], MD; Eliza Congdon[1], PhD; Nelson Freimer[1], MD; Michelle Craske[1,2], PhD

[1]Department of Psychiatry and Biobehavioral Sciences, University of California - Los Angeles, Los Angeles, CA, United States

[2]Department of Psychology, University of California - Los Angeles, Los Angeles, CA, United States

[3]Greater Los Angeles Veterans Administration, Los Angeles, CA, United States

Corresponding Author:
Kate Wolitzky-Taylor, PhD
Department of Psychiatry and Biobehavioral Sciences
University of California - Los Angeles
760 Westwood Blvd.
Los Angeles, CA, 90025
United States
Phone: 1 3102675339
Email: kbtaylor@mednet.ucla.edu

Abstract

Background: The global burden of anxiety and depression has created an urgent need for scalable approaches to increase access to evidence-based mental health care. The Screening and Treatment for Anxiety and Depression (STAND) system of care was developed to meet this need through the use of internet-connected devices for assessment and provision of treatment. STAND triages to level of care (monitoring only, digital therapy with coaches, digital therapy assisted by clinicians in training, and clinical care) and then continuously monitors symptoms to adapt level of care. Triaging and adaptation are based on symptom severity and suicide risk scores obtained from computerized adaptive testing administered remotely.

Objective: This article discusses how the STAND system of care improves upon current clinical paradigms, and presents preliminary data on feasibility, acceptability, and effectiveness of STAND in a sample of US-based university students.

Methods: US-based university students were recruited and enrolled in an open trial of the STAND system of care. Participants were triaged based on initial symptom severity derived from a computerized adaptive test and monitored over 40 weeks on anxiety, depression, and suicide risk to inform treatment adaptation and evaluate preliminary effectiveness.

Results: Nearly 5000 students were screened and 516 received care. Depression and anxiety severity scores improved across all tiers (P<.001 in all cases). Suicide risk severity improved in the highest tier (ie, clinical care; P<.001). Acceptability and feasibility were demonstrated.

Conclusions: STAND is a feasible and acceptable model of care that can reach large numbers of individuals. STAND showed preliminary effectiveness on all primary outcome measures. Current directions to improve STAND are described.

(JMIR Ment Health 2023;10:e46200) doi:10.2196/46200

KEYWORDS

depression; anxiety; cognitive behavioral therapy; digital mental health; stratified models of care; model of care; care model; depressive; mental health; CBT; psychotherapy; university; postsecondary; student; college; service delivery; care system; system of care; mHealth; eHealth; online support; student mental health

Introduction

Background

The public health burden of depression and anxiety is enormous and rising, with effects extending to unemployment, risky substance use, and of greatest concern, suicide. In particular, young adults have seen a substantial increase in anxiety, depression, and suicidal thoughts or behaviors over the past decade [1,2]. Mental health care systems, globally, face significant challenges in meeting the staggering need for care created by this rising public health burden. In particular, access

to traditional forms of mental health care is limited by cost, time, effort, and stigma [3], and the number of clinicians trained in evidence-based treatments is grossly inadequate to meet current needs. The urgency of these needs has stimulated the development of 2 broad strategies for increasing access to care: stepped care approaches and digital provision of therapies.

Increasing Access to Care Using Stepped Care Approaches and Digital Provision of Therapies

In stepped care models, typically, all patients begin with low-intensity treatments and only nonresponders are moved to the next (more intensive) step of care [4]. Low-intensity treatments often include digital or online therapies, with or without support, as multiple trials targeting anxiety and depressive disorders have shown that these treatments can be as effective as face-to-face therapy [5-8]. While this "fail-first" approach more efficiently allocates resources than traditional approaches that strive to provide the highest level of clinician-delivered care to all patients, it results in delays to more intensive treatment for those who need it [9], and such delays may lead to symptom worsening and increased disability [10], discouragement for new treatments after initial options fail [11], or dropout before a second course of treatment [12]. An emerging body of literature indicates that *stratified* models, in which individuals are matched to level of care based on predetermined criteria (typically current symptom severity), produce superior outcomes and are more cost-effective in treating depression and anxiety compared with fail-first stepped care models [13-16].

Digital Technologies for Delivering Measurement-Based Care and Treatment Adaptation

The use of data collected from patients during the course of treatment for clinical decision-making (measurement-based care) has the potential to greatly improve outcomes but remains the exception rather than the rule [17]. Specifically, measurement-based care typically outperforms usual care on both symptom reduction and reduced dropout rate [18] and is associated with decreased costs and lower odds of patient deterioration during treatment [19]. Without measurement-based care, clinicians have difficulty identifying patients who are at risk for nonresponse or deterioration [20,21]. Consequently, interventions become reactive to "crisis" needs rather than being proactive. Further, decisions about when treatment has been successfully completed tend to be determined by when the patient or their provider "feels ready" to terminate therapy, which may lead to extending treatment far beyond the point when symptom gains have been attained and maintained.

Even when measurement-based care is utilized, guidelines on how to adjust treatment over time are lacking. As discussed, stepped care approaches typically personalize level of care *after* a patient has shown nonresponse. Waiting several months to make adaptation decisions could have dire consequences for individuals with depression and anxiety [22]. Moreover, typical stepped care approaches step up only for patients who remained actively engaged in treatment, thereby not addressing the critical issue of how to manage and engage individuals who prematurely discontinue treatment [23].

Furthermore, typical care models fail to fully consider changes in clinical status after the end of acute treatment; relapse is not uncommon (especially for depression), yet rarely are patients monitored in order to identify for whom reinitiation of care is warranted. By identifying those at risk for nonresponse or worsening of symptoms during or after treatment, adaptations can be made to increase treatment intensity or reenter treatment and prevent further worsening of symptoms and their potential consequences. Routine monitoring of clinical status enables dynamic adaptation of treatment as needed, which has the potential to improve effectiveness and reduce attrition, as patients may be more engaged in treatment when they are receiving what they need most at the time they most need it [24-26]. Adaptive interventions can also enhance cost-effectiveness by increasing the efficiency of service delivery and reducing downstream service costs [27].

Digital Technologies for Identification and Management of Suicide Risk

Suicidal thoughts and behaviors are a major public health problem, particularly among young adults [28]. Rates of suicide attempts, nonsuicidal self-injury, and suicidal ideation increase markedly during adolescence [29-31], with suicide death rates increasing as youth move into adulthood. Across all adult age groups, the prevalence of serious suicidal thoughts is highest among young adults aged 18-25 (11%) [32]. There is a critical need to embed methods for detection of risk and strategies for suicide/self-harm prevention within systems of care, which go beyond relying on patients to reach out to providers or to call crisis lines, and beyond monitoring at frequencies that fail to capture tipping points toward high risk. Digital tools may be particularly useful in assessing and monitoring such risk, given that they can promote increased self-disclosure of sensitive topics compared with face-to-face assessment and allow for more rapid, in-the-moment identification and response [33].

STAND: A Novel, Scalable, Dynamic, Digitally Assisted Evidence-Based Solution to Mental Health Care Delivery

Screening and Treatment for Anxiety and Depression (STAND) is a stratified stepped care model that incorporates online screening, continuous symptom monitoring over ~10 months, and tiered treatment for anxiety and depressive symptoms, with ongoing suicide risk detection and management (Figure 1). After individuals complete a brief online adaptive assessment of symptom severity and suicide risk (selected modules from the Computer Adaptive Test for Mental Health [CAT-MH]; [34]), they are provided feedback, routed to an appropriate level of care, and scheduled for an orientation or intake session, all within minutes of completion of initial screening. The CAT-MH assessment continuously tracks symptoms (both during and after active treatment) to adapt treatment (ie, move to a higher level of care when symptoms worsened or to a lower level of care for maintenance and relapse prevention) and to rapidly detect and respond to signs of elevated suicide risk. For the open trial described herein, there were 4 levels of care, or tiers.

Tier 0 was for those with no or minimal symptoms of depression or anxiety on the initial CAT-MH assessment. These participants were offered the option to continue completing the CAT-MH

biweekly [34]. Tier 1 was offered to those with mild depression or mild anxiety on the CAT-MH, and included digital therapy. Coaching was offered to tier 1 participants given evidence for its positive effects on retention and clinical outcomes from digital therapy [8]. Tier 2 involved digital cognitive behavioral therapy (CBT) with advanced (doctoral student) coaches, and was offered to individuals with moderate depressive symptoms or moderate to severe anxiety symptoms. Evidence-based psychological treatment with option for medication management was offered to those with severe depressive (or manic) symptoms or suicidality (tier 3). Core active ingredients of CBT (see the "Methods" section) were selected to match each participant's problems areas and were delivered by clinical psychology graduate students. Therapy was complemented by protocolized medication management as needed (modified from the STAR*D trial; [35]), delivered by psychiatry residents.

Across all tiers, treatment was adapted according to weekly or biweekly CAT-MH scores (Figure 2). If participants' scores indicated a need for a higher tier of care at any time, they were contacted to initiate switching to the appropriate level of care. Similarly, participants in tier 3 could be switched to a lower tier and were thus offered access to digital therapy materials when they completed a course of tier 3 treatment and symptoms had shown improvement. Moreover, within tier 3, clinicians used weekly CAT-MH scores to guide decisions regarding treatment strategy (such as switching from behavioral activation to cognitive restructuring).

Frequent monitoring of suicide risk was conducted using a standardized protocol for responding to risk in real time (see the "Measures" section for details). This approach to risk management was implemented for all levels of care within STAND. In addition to outreach, a positive suicidal triggering alert indicated adaptation to tier 3 if the participant was in a lower tier.

Figure 1. The STAND system of care. STAND: Screening and Treatment for Anxiety and Depression.

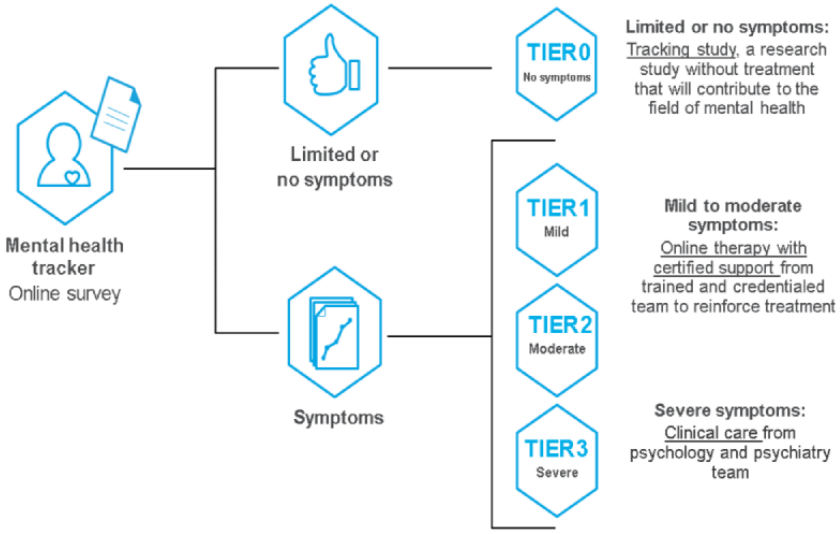
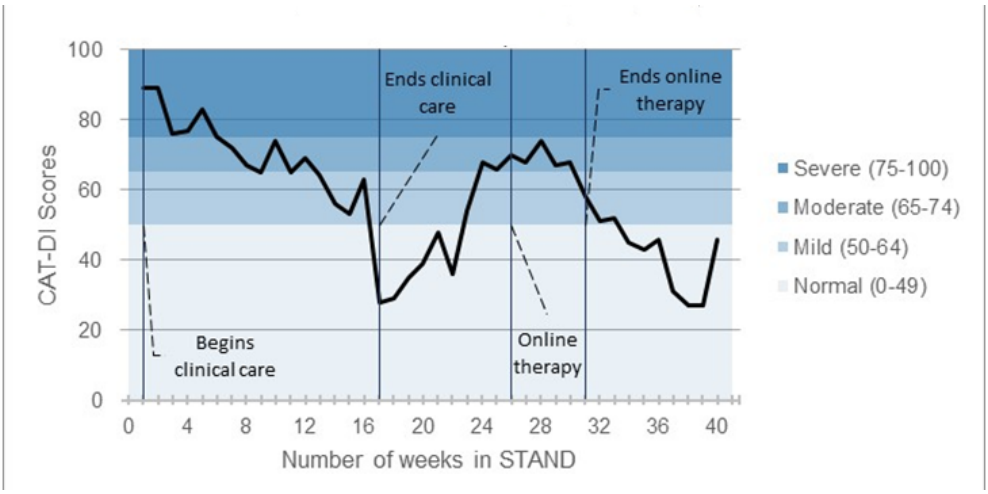


Figure 2. Treatment adjustment of a simulated patient in clinical care. CAT-DI: Computerized Adaptive Test-Depression Inventory; STAND: Screening and Treatment for Anxiety and Depression.



Summary of STAND Improvements to Clinical Practice Paradigms

STAND improves standard clinical practice paradigms in several ways. First, STAND provides increased access to evidence-based care by utilizing digital tools for a large proportion of the patient experience from screening through service enrollment to service delivery and risk management. Second, the emphasis on adapting both the level of care and intervention type as needs change is a major improvement over existing approaches (eg, stepped care), which typically wait until nonresponse is established at predefined points that are months apart before offering an alternative, if at all [16]. Third, STAND focuses on maintaining gains and minimizing relapse, which are rare in systems of care, despite the many benefits to individuals and a health system [36]. Finally, the STAND program focuses on personalization of services based on continuous symptom monitoring and clinical decision-making (both with regard to tier triaging and adaptation as well as choice of treatment modality within tier 3, the highest level of care) that is expected to reduce attrition and improve symptomatic and functional outcomes [23].

This study aimed to develop and evaluate the preliminary acceptability, feasibility, and effectiveness of the STAND program in a large sample of college students in an open trial. It was hypothesized that participants would report high satisfaction with the program, and would show significant improvement in anxiety, depression, and suicide risk severity scores over the course of participation in STAND.

Methods

Participant Recruitment and Initial Triaging

Recruitment and enrollment into the open trial occurred between Fall 2017 and Spring 2020. All University of California, Los Angeles (UCLA) students were made aware of the screener through announcements from UCLA's Chancellor, tabling at campus events, campus residence hall flyers, classroom announcements, web-based advertisements, banners, and email blasts. Students completed the CAT-MH online and received immediate feedback regarding their scores and recommended tier of care. Those who were interested in and eligible for the study and who provided informed consent were then scheduled for a baseline visit. At the baseline visit, research staff gathered pertinent information about treatment history and other demographics, after which participants were notified about how to start care in their assigned tier, with ongoing completion of the CAT-MH and other assessments over the course of 40 weeks.

Inclusion/Exclusion Criteria

All matriculating UCLA students who were at least 18 years of age and had access to the internet were eligible to complete the initial screener. Additional inclusion criteria for screening were English fluency, willingness to install an app on their phone, and no plans (eg, upcoming long vacations or studying abroad) that would interfere with participation. Additional exclusion criteria following screening for those entering clinical care (ie, tier 3) were severe psychopathology requiring intensive day-treatment or inpatient care (eg, severe eating disorder or substance use disorder, multiple [2+] suicide attempts leading to hospitalization within the past 6 months, significant psychotic symptoms not part of a severe major depressive episode, or bipolar manic episode) and were determined at the initial assessment through further screening. Participants in tier 3 were also excluded if they were unwilling to provide a blood sample (for secondary analyses outside of the scope of this manuscript) or had a current psychiatrist or provider and were unwilling to transfer their care to the study team for the period that they will receive treatment in STAND.

Interventions

Tier 0

Participants whose Computerized Adaptive Test-Depression Inventory (CAT-DI) and Computerized Adaptive Test-Anxiety Inventory (CAT-ANX) scores were in the normal range were offered tier 0 or monitoring of their status using biweekly CAT-MH.

Tier 1

Those whose depression or anxiety scores were in the mild range were offered tier 1 that included an evidence-based digital CBT called *ThisWayUp* [37]. *ThisWayUp* provides 6 lessons on CBT strategies for anxiety and depression including behavioral activation, cognitive restructuring, and exposure. This digital therapy program was selected, as randomized controlled trials have demonstrated evidence to support its efficacy [37,38]. Digital CBT was accompanied by the offer of up to 8 phone or video coaching sessions by extensively trained undergraduate student coaches (see the "Methods" section for details) to support the digital therapy material and to assist with troubleshooting (rather than to teach clinical content). As human support has been shown to increase efficacy and retention in digital therapies [8], coaching sessions were offered each week after participants independently completed a *ThisWayUp* lesson. Coach training had 4 sequential components (see [39] for details). Didactic, interactive classroom meetings and small-group coaching support meetings focused on content and general coaching skills. Discussions were held weekly with opportunities to practice coaching skills under close supervision and with small-group feedback, and to achieve core competencies in professionalism, diversity awareness, ethical standards, and reflective practice and self-care. Advancement to the Certified Coach status was required before coaches could provide video-chat coaching to study participants, under supervision from program faculty. Coaching was initially provided to participants in a group format, and later was provided in an individual 1-on-1 format.

Tier 2

Tier 2 was offered to those who scored in the moderate range for depression or moderate to severe range for anxiety. Participants were assigned to complete *ThisWayUp* and were strongly encouraged to attend weekly face-to-face, video-chat, or phone coaching sessions following their independent completion of each *ThisWayUp* lesson. In addition to being trained in *ThisWayUp*, advanced (doctoral students) coaches participated in the tier 3 clinical training (described below) to

gain a deeper knowledge of cognitive and behavioral principles. Given their more advanced training, tier 2 advanced coaches had more flexibility to support the CBT material in the coaching sessions. They received weekly supervision by licensed clinical psychologists.

Tier 3

Evidence-based psychological treatment with option for medication management was offered to those with severe depressive (or manic) symptoms or suicidality. Evidence-based psychological treatments often consist of multicomponent packages of different cognitive and behavioral strategies, such as cognitive restructuring, self-monitoring, problem solving, relaxation training, assertiveness training, or exposure therapy. Such combination packages prohibit understanding of which ingredient was most accountable for therapeutic change for each individual and the processes responsible for therapeutic change [40,41]. Approaches that rely upon specific active ingredients may enable better treatment personalization and greater treatment efficacy (eg, [42]). Thus, core cognitive behavioral ingredients [43] (Table 1) were delivered to patients according to their individualized needs, as was ascertained during a functional assessment. The goal of this clinician-administered, semistructured interview assessment was to identify the primary processes (eg, fear/avoidance and deficits in extinction; inactivity/sadness and deficits in response-contingent positive reinforcement) that were driving or maintaining symptoms, and then to select the therapeutic strategy that most directly targeted those processes (eg, exposure therapy and behavioral activation). Therapy was complemented by protocolized medication management as needed (modified from the STAR*D trial; [35]), delivered by psychiatry residents. Weekly CAT-MH scores made available to clinicians on a dashboard enabled continuous monitoring of treatment response and need for treatment adaptation. For example, if CAT-MH depression scores did not show improvement within 6 weeks of behavioral activation, the functional assessment was repeated, followed by possible switching to either cognitive restructuring or mindfulness-based approaches. Similarly, medication management changes were considered in light of poor response on the CAT-MH scores.

Weekly individual therapy sessions were delivered by clinical psychology doctoral students and psychiatry residents, as part of their in-house training programs. They were trained through a series of workshops and didactics on each of the evidence-based active ingredients (Table 1), consisting of didactic material, demonstrations of skills by the experts, and role-play practices of skills. They were taught to use the functional assessment and treatment decision-making algorithm to identify the principal process to target and the first-line treatment ingredient that mapped onto that target (eg, worry—cognitive restructuring; low activity/sadness—behavioral activation); and the process for monitoring symptoms to identify when a functional assessment and treatment plan should be updated. After initial training, psychology trainees participated in weekly supervision and an interdisciplinary case conference.

Psychiatry trainees were trained in the STAND medication management algorithm and received regular supervision from attending psychiatrists. Psychiatrists also participated in the same weekly interdisciplinary case conference as psychology trainees, which also included members of the campus crisis response team who could assist with coordination of outreach for high-risk participants.

Table 1. Matching of clinical care to patient problem area.

Problem area	Process targeted	First-line therapy module + medication as appropriate	Process targeted	Second-line therapy module + medication as appropriate
Low activity/sadness	Low response-contingent positive reinforcement	Behavioral activation (mood monitor, activity schedule, problem solve barriers, sleep schedule for barriers)	Cognitive distortions and rumination	Cognitive restructuring or mindfulness and value-driven action and problem solving
Anhedonia	Reward hyposensitivity	Pleasant event scheduling (hedonic and eudaimonic rewards), memory specificity recounting	Reward hyposensitivity	Cognitive restructuring with positive focus and cultivating positivity
Fear/phobia	Deficits in extinction, safety learning, avoidance	Exposure therapy	Negative cognitive bias and poor social skills	Cognitive restructuring, mindfulness, value-driven action or social skills training
Worry	Negative cognitive bias	Cognitive restructuring or mindfulness, value-driven action	Avoidance (experiential and in vivo)	Exposure therapy or mindfulness and value-driven action or social skills training
Sleep dysregulation	Sleep dysregulation	Brief behavioral therapy for insomnia	Negative cognitive bias	Cognitive restructuring
Trauma—fear	Deficits in extinction, safety learning, avoidance	Imaginal and in vivo exposure	Negative cognitive bias	Cognitive restructuring and impact statement
Trauma—guilt, shame, cognitive distortions	Negative cognitive bias	Trauma narrative with cognitive restructuring and impact statement	N/A[a]	N/A
Chronic suicidality, self-harm, affective instability	Low tolerance of distress	Distress tolerance skills in DBT[b]	Poor emotion regulation and interpersonal difficulties	Emotion regulation skills and interpersonal effectiveness in DBT
Mania	Circadian dysregulation	Brief behavioral therapy for insomnia	N/A	N/A
Major life stressors (any symptom profile)	Poor coping	Problem solving for controllable stressor, mindfulness, value-driven action for uncontrollable stressor	N/A	N/A
Interpersonal relations (any symptom profile)	Social skills deficits	Interpersonal effectiveness training in DBT	N/A	N/A

[a]N/A: not applicable (ie, no second-line therapy module for trauma, mania).

[b]DBT: dialectical behavioral therapy.

Tier Adaptation

Across all tiers, treatment was adapted according to weekly or biweekly CAT-MH scores. If participants' scores indicated a need for a higher tier of care at any time (ie, showed a 30% worsening of symptom severity from baseline over 2 consecutive assessments), they were contacted to initiate switching to the appropriate level of care. Similarly, participants in tier 3 were switched to a lower tier when their symptoms remained consistently low (ie, in the moderate to normal range of severity following ~12-16 weeks of treatment). Thus, when participants completed treatment in tier 3 they were offered access to ThisWayUp materials from tiers 1 and 2. Moreover, within tier 3, clinicians used weekly CAT-MH scores to switch treatment strategies (such as switching from behavioral activation to cognitive restructuring).

Suicide Risk Monitoring

The CAT-MH [34] digitally identified participants at risk of suicide; scores above a preset threshold triggered an alert to initiate a risk management protocol. The CAT-MH includes (but is not limited to) the screener module from the Columbia Suicide Severity Rating Scale (CSSRS; [44]). The CSSRS yields a positive case as indicated by the presence of current suicide ideation *plus* current intent *or* current method *or* recent suicidal behavior (eg, suicide attempt, steps taken toward an attempt). A positive case automatically triggered a 24/7 crisis service that made up to 3 attempts to contact the participant to conduct a risk assessment and appropriate care management. All participants in the STAND program consented to be contacted by this crisis service at study entrance. Relevant STAND staff and clinicians were also informed, and could make additional outreach efforts if the crisis service outreach attempts were unsuccessful. In the event that the participant was unable to be reached by the crisis service or STAND staff and clinicians, the deployment of other community-based crisis units was considered on a case-by-case basis.

Measures

Clinical Outcomes

The primary outcome measure was the CAT-MH [34]. The CAT-MH is a commercially available (Adaptive Technologies) computerized adaptive test for assessing a variety of symptoms including those of anxiety (CAT-ANX; [45]), depression (CAT-DI; [46]), and suicide risk severity (Computerized Adaptive Test-Suicide Scale [CAT-SS]; [47]). The CAT-MH scores included symptom severity as a cumulative score of endorsed symptoms (0-100) or severity categories per symptom cluster for the CAT-DI and CAT-ANX (normal, mild, moderate, and severe) based on empirically derived cut points.

Depression scores (CAT-DI severity) of <50.0 were considered normal, 50.0-65.0 mild, >65.0-75.0 moderate, and ≥75.0 severe for the majority of the recruitment period (through June 2019) [46,48], and anxiety scores (CAT-ANX severity) of <35.0 were considered normal, 35.0-50.0 mild, >50.0-65.0 moderate, and >65.0 severe [45,48]. From June 2019 to June 2020 (ie, end of recruitment), CAT-DI categories were slightly changed based on findings from [48] that led to recommendations from the CAT-MH developer team (R. D. Gibbons, PhD, personal communication, June 9, 2019). During this latter period, the threshold for mild depressive symptoms was lowered such that scores <35.0 were considered normal, and the mild range was 35.0-65.0. The CAT-MH was administered weekly (tier 3) or biweekly (tiers 0-2) across the entire course of participation in STAND. The CAT-MH categories were used to triage participants into initial tier, and were used to inform adaptation to tiers (ie, tier switching) during the next 40 weeks. Therefore, the primary outcomes gathered from the CAT-MH (ie, CAT-ANX severity score, CAT-DI severity score, and CAT-SS severity score) were the dependent variables. See [34,45-47] for details regarding the excellent psychometric properties of the CAT-ANX, CAT-DI, and CAT-SS.

Participant Acceptability

Participants were asked to provide feedback on their experience in the STAND program when they reached the final 40-week assessment (this assessment was initiated halfway through the study). Participants were asked to rate how logical the program seemed (on a scale from 1 to 9, where 1=not at all logical, 5=somewhat logical, and 9=very logical), to what extent the program met their expectations (on a scale from 1 to 5, where 1=completely failed to meet expectations, 3=met expectations, and 5=greatly exceeded expectations), how satisfied they were with the program (on a scale from 1 to 7, where 1=extremely dissatisfied, 4=neither satisfied nor dissatisfied, and 7=extremely satisfied), and how likely they would be to recommend the program to a friend (on a scale from 1 to 7, where 1=extremely unlikely, 4=neither likely nor unlikely, and 7=extremely likely).

Statistical Analysis

Descriptive statistics were used to characterize the data and to identify indices of feasibility and acceptability. With regard to the primary outcome data, a series of hierarchical linear models (1 per tier per outcome measure) were conducted with time as the level 1 predictor and the outcome of interest (CAT-ANX, CAT-DI, and CAT-SS scores for tier 3, and CAT-ANX and CAT-DI scores for tiers 1 and 2). In line with an effectiveness paradigm, we included all participants who were assigned to a tier and had at least one CAT-MH datapoint, regardless of how many sessions/lessons were completed or how much missing data on the CAT-MH was observed. Parallel sets of analyses were run with and without a covariate (number of sessions/lessons completed on level 2). The survey data were not consistently collected for participants in tiers 1 and 2 after completion of the *ThisWayUp* program. Therefore, results for tiers 1 and 2 reflect CAT-MH scores over the course of *ThisWayUp* program completion only (ie, 8 weeks). Further, of the 11 participants who completed a baseline assessment for tier 0, only 5 participants completed more than 3 additional assessments. For this reason, CAT-MH scores are not reported for tier 0.

As a secondary analysis, the percentages of those within each tier who achieved at least 30% improvement on each CAT-MH index (ie, CAT-DI, CAT-ANX, and CAT-SS) were calculated among participants considered at least moderately engaged (ie, completed at least four sessions/lessons).

Ethics Approval

The project was approved by the UCLA Institutional Review Board (approval numbers 17-001938, 16-001395, and 17-001365).

Results

Indices of Feasibility

Coach and Clinician Training

A total of 530 student coaches (including undergraduates in the resilience peer network for tier 1 and doctoral students for tier 2) were trained (to an adequate or better level, thus allowing them to deliver coaching; see [39]), and 47 psychology and psychiatry clinicians were trained to deliver psychological and psychiatric care in tier 3. For further details on student coach training and competency assessment, see [39].

STAND Participant Uptake, Engagement, and Adaptations

A total of 4845 unique students completed screening using the CAT-MH; 3580 out of the 4845 students who completed screening (73.89%) were eligible and offered care in tiers 1, 2, or 3; 516 out of the 3580 eligible students who were offered care (14.41%) initiated therapy in one of the tiers. As many as 327 individual suicide risk alerts were detected at screening, and additional risk assessments (after screening) were conducted for 1054 alerts indicating a risk of suicide or severely worsening depression over the course of the study.

A total of 180 students received care (ie, completed at least one lesson) in tier 1 and 197 in tier 2. Those in tier 1 completed on average 4.22 (SD 1.91) out of 6 digital therapy lessons, which was similar or better than adherence reported in previous studies [37,38]. Tier 1 participants completed on average 0.88 (SD 2.07) coaching sessions. The correlation between number of coaching sessions and digital therapy lessons completed in tier 1 was small but significant and positive (r=0.25, P<.001).

Students in tier 2 completed on average 4.09 (SD 2.06) out of 6 digital therapy lessons and 2.09 (SD 2.56) group or 1-on-1 coaching sessions. The correlation between number of coaching sessions and digital therapy lessons completed in tier 2 was moderately large and significant (r=.54, P<.001). With regard to "tier switching"/adaptation, only 5 participants who were in tiers 1 or 2 were moved up to tier 3 due to a symptom worsening or suicide risk alert. A total of 139 students received clinical care in tier 3 (ie, attended at least one therapy session), and attended on average 13.86 (SD 7.94) therapy sessions. Following the acute course of tier 3 treatment, all tier 3 participants were offered online CBT via *ThisWayUp* (tier 1 without coaching).

Demographic, diagnostic, and baseline clinical data for tiers 0-3 are presented in Table 2.

Table 2. Demographic characteristics.

Characteristics[a]	Tier 0	Tier 1	Tier 2	Tier 3
Completed baseline assessment, n	11	225	239	144
Age (years), mean (SD)	22.91 (6.56)	23.01 (5.81)	21.93 (4.47)	21.14 (3.36)
Assigned sex at birth, n				
Female	5	160	184	107
Male	6	63	53	36
Current gender identity, n				
Female	5	159	181	105
Male	6	61	51	35
Transgender	0	1	2	0
Does not identify as either	0	2	3	3
Sexual orientation[b], n				
Asexual	0	1	0	1
Bisexual	0	12	25	27
Heterosexual/straight	8	49	82	81
Homosexual, gay or lesbian	1	2	5	7
Queer	0	3	6	6
Questioning or unsure	1	8	12	9
Not listed	1	0	3	0
Prefers not to answer	0	1	2	5
Marital status, n				
Single, never married	10	195	211	131
Living with partner	1	10	9	4
Domestic partnership	0	0	2	1
Married	0	14	11	4
Separated	0	0	0	0
Divorced	0	2	0	1
Do not know	0	0	1	1
Prefers not to answer	0	2	2	1
First language, n				
English	11	159	161	104
Spanish	0	22	23	15
Other	0	42	53	24
Country of birth, n				
United States	10	171	174	111
Other	1	52	63	32
Immigration status, n				
Domestic	11	194	201	126
International	0	25	31	13
Undocumented	0	4	4	2
Other	0	0	0	0
Prefers not to answer	0	0	1	2

Characteristics[a]	Tier 0	Tier 1	Tier 2	Tier 3
Racioethnic group, n				
Hispanic	1	40	59	36
Non-Hispanic White	3	85	71	35
Non-Hispanic Black	2	4	10	4
Non-Hispanic Asian	4	72	71	52
Non-Hispanic multiple	1	17	17	12
Non-Hispanic Native American/Pacific Islander	0	0	0	0
Unknown	0	5	9	4
Highest level of education, n				
High school graduate	2	34	44	43
General educational development or equivalent	0	1	0	0
Some college	4	85	96	59
Associate's degree (occupational, technical, or vocational)	0	1	3	0
Associate's degree (academic program)	0	12	27	14
Bachelor's degree	5	66	46	18
Master's degree	0	23	18	8
Professional degree	0	1	2	0
PhD	0	0	1	0
Prefers not to answer	0	0	0	1
Employment, n				
Working	3	20	26	16
Only temporarily laid off, sick leave, or maternity leave	0	0	0	1
Looking for work, unemployed	0	6	6	2
Disabled, permanently or temporarily	0	0	0	1
Student	8	194	201	121
Other	0	3	4	2
Student enrollment status, n				
Full time	10	215	229	138
Part-time	0	4	7	5
College level, n				
Freshman	3	30	35	34
Sophomore	0	29	41	25
Junior	1	42	61	32
Senior	2	37	36	26
Graduate student	4	70	55	22
Professional student	0	9	5	4
Other	0	2	3	0
Baseline CAT-MH[c] Depression				
Depression severity, mean (SD)	33.63 (9.04)	49.90 (11.44)	59.96 (12.25)	73.21 (11.32)
Minimal to none depression, n	10	96	36	5
Mild depression, n	1	109	116	21

Characteristics[a]	Tier 0	Tier 1	Tier 2	Tier 3
Moderate depression, n	N/A[d]	19	65	61
Severe depression, n	N/A	N/A	21	57
Baseline CAT-MH Anxiety				
Anxiety severity, mean (SD)	18.42 (12.74)	39.65 (12.51)	55.60 (14.37)	60.14 (16.10)
Minimal to none anxiety, n	10	82	19	8
Mild anxiety, n	1	104	74	32
Moderate anxiety, n	N/A	33	79	45
Severe anxiety, n	N/A	6	66	59
Baseline CAT-MH Suicidality				
Suicidal ideation diagnosis, n	N/A	2	6	24
Suicidality severity, mean (SD)	18.49 (12.95)	39.62 (12.67)	50.29 (11.99)	62.19 (11.76)

[a]Participants were able to enroll in the study at both wave 1 and wave 2 of recruitment. Therefore, some participants may be represented twice if they chose to reenroll in the study.

[b]This question was introduced partway through the study, therefore not all participants had the opportunity to respond.

[c]CAT-MH: Computer Adaptive Test for Mental Health.

[d]N/A: not applicable.

STAND Clinical Outcomes (Preliminary Effectiveness)

Overview

There was no difference in the pattern or significance of any findings when the covariate (ie, number of sessions/lessons completed) was included. Therefore, analyses without statistically adjusting for the number of sessions/lessons completed are reported below as a conservative approach. Specific trajectories for each tier/outcome combination are described below in more detail. Figures 3 and 4 show the decline slopes for the CAT-ANX and CAT-DI within each tier, and Figure 5 shows the CAT-SS decline slope in tier 3.

Figure 3. Tier 1 changes in CAT-DI and CAT-ANX scores over time. CAT-ANX: Computerized Adaptive Test-Anxiety Inventory; CAT-DI: Computerized Adaptive Test-Depression Inventory; SI: suicidal ideation.

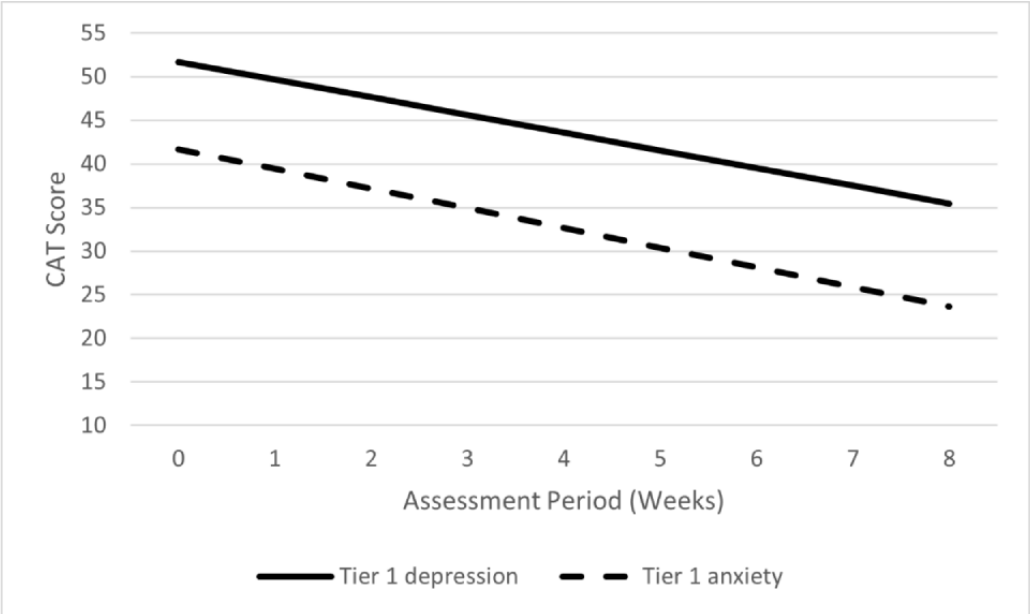


Figure 4. Tier 2 changes in CAT-DI and CAT-ANX scores over time. CAT-ANX: Computerized Adaptive Test-Anxiety Inventory; CAT-DI: Computerized Adaptive Test-Depression Inventory.

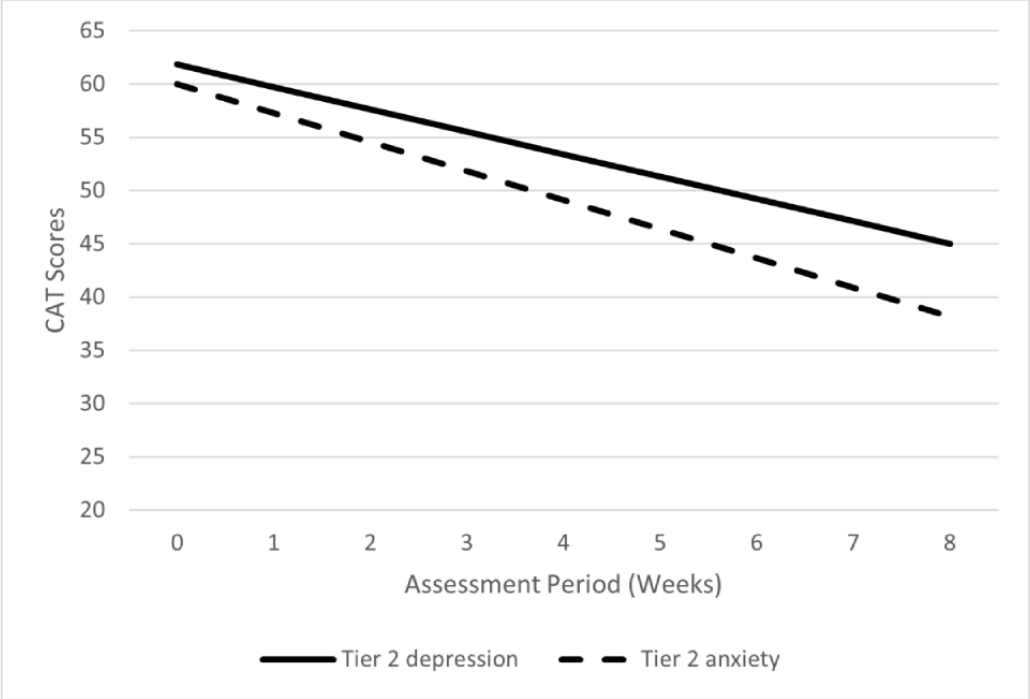
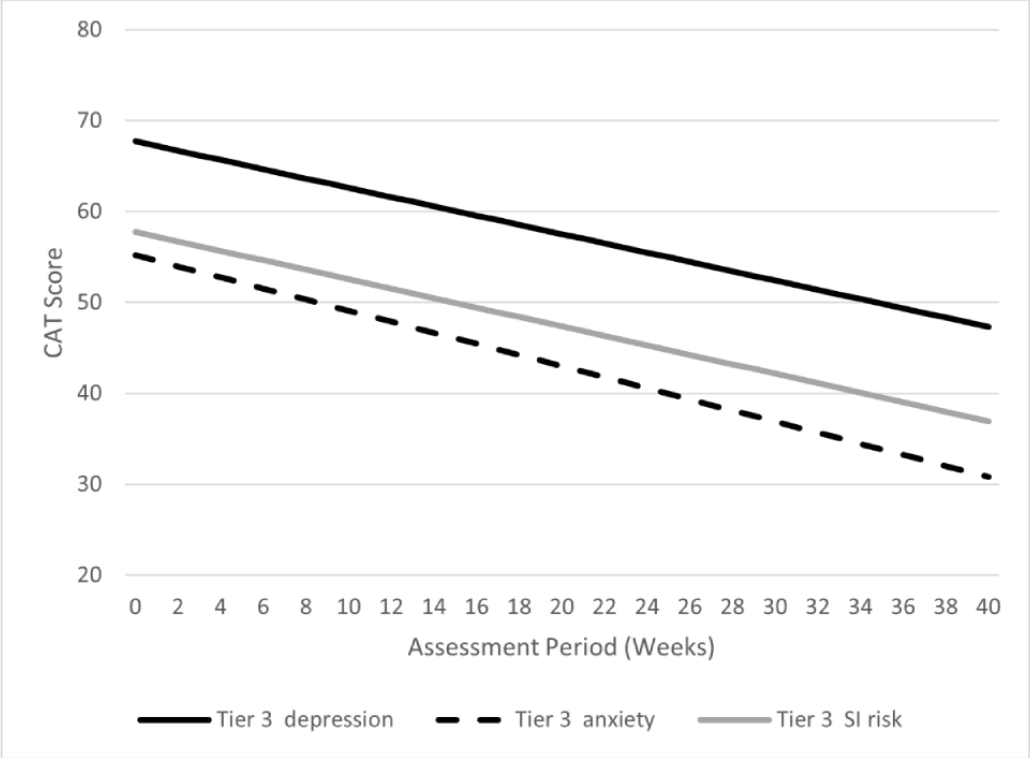


Figure 5. Tier 3 changes in CAT-DI, CAT-ANX, and CAT-SS scores over time. CAT-ANX: Computerized Adaptive Test-Anxiety Inventory; CAT-DI: Computerized Adaptive Test-Depression Inventory; CAT-SS: Computerized Adaptive Test-Suicide Scale.



Tier 1

The CAT-DI and CAT-ANX scores significantly declined over the course of STAND participation (b=–2.03, t_{732}=–10.71, P<.001 and b=–2.26, t_{732}=–12.47, P<.001, respectively). Based on the regression lines, scores moved from the mild to normal range for both depression and anxiety severity (Figure 3). Among participants who completed at least four digital therapy lessons (n=110), 51 participants (46.4%) showed ≥30% improvement in CAT-DI scores and 62 participants (56.4%) showed ≥30% improvement in CAT-ANX scores, notable for a sample likely to have a floor effect, with minimal room for significant symptom improvement.

Tier 2

The CAT-DI and CAT-ANX scores significantly declined over the course of STAND participation (b=–2.10, t_{696}=–9.64, P<.001 and b=–2.73, t_{696}=–12.42, P<.001, respectively). As shown in Figure 4, although scores for depression began and ended in the mild range (following the adjustment of categories in June 2019), they moved from the top of the mild range into the lower end of the mild range (and into what was considered the normal range prior to June 2019). Anxiety severity scores moved from the moderate range to the mild range. Among participants who completed at least four digital therapy lessons (n=106), 33 participants (31.1%) showed ≥30% improvement in CAT-DI scores and 47 participants (44.3%) showed ≥30% improvement in CAT-ANX scores.

Tier 3

The CAT-DI and CAT-ANX scores significantly declined over the course of STAND participation (b=–0.61, t_{2805}=–11.91, P<.001 and b=–0.51, t_{2805}=–9.64, P<.001, respectively). The CAT-SS severity also declined over time (b=–0.52, t_{2805}=–11.05, P<.001). As depicted in Figure 5, the regression line shows depression scores moved from the moderate range to the mild range (and what was considered the normal range prior to June 2019) and anxiety scores moved from the moderate range to the normal range. Among participants who completed at least four therapy sessions (n=131), 74 participants (56.5%) showed ≥30% improvement in CAT-DI scores, 81 participants (61.8%) showed ≥30% improvement in CAT-ANX scores, and 78 participants (59.5%) showed ≥30% improvement in CAT-SS scores.

Participant Acceptability

Results on the subsample (n=79) who completed the satisfaction survey are reported in Table 3. In sum, participants across tiers 1-3 rated their care as highly logical (scores ranging from 6.65 to 7.60 out of 9), that the program succeeded in meeting their expectations (scores ranging from 3.47 to 3.95 out of 5), that they were satisfied with the program (scores ranging from 5.24 to 5.90 out of 7), and that they would recommend it to a friend (scores ranging from 5.47 to 6.14 out of 7). Tier 0 participants also rated these categories highly but scores should be interpreted with caution, given the small sample size in tier 0 (Table 3).

Table 3. Program feedback.

Variables	Tier 0	Tier 1	Tier 2	Tier 3
Participants, n	3	17	17	42
Logical, mean (SD)	7.33 (0.58)	6.65 (2.23)	6.59 (1.28)	7.60 (1.62)
Expectation, mean (SD)	3.00 (0.00)	3.47 (1.01)	3.76 (0.83)	3.95 (1.04)
Satisfaction, mean (SD)	4.67 (0.58)	5.24 (1.56)	5.47 (1.01)	5.90 (1.25)
Recommendation, mean (SD)	4.67 (0.58)	5.65 (1.37)	5.47 (0.94)	6.14 (1.18)

Discussion

Summary of Outcomes

This preliminary study demonstrated that STAND, a technology-assisted and scalable system of care, is feasible and effective on a college campus. STAND facilitated the deployment of a significant number of suicide risk assessments as needed in real-time, a critical index of risk prevention. As part of the STAND model, a large number of coaches (n=530) and clinicians (n=47; in psychiatry and psychology) were trained to provide support for or deliver evidence-based treatment for anxiety and depression, respectively. Care was delivered to 516 UCLA students (ie, those who completed at least one session of therapy/digital therapy lesson), who, on average, experienced significant reductions in their anxiety and depression symptoms across all tiers. Participant acceptability on the satisfaction survey revealed moderately high ratings of satisfaction with participants' assigned tier. Importantly, the stratified model of initial triaging to the appropriate level of care was successful in that the vast majority of participants showed significant improvement in their originally assigned tier. Indeed, only 5 participants required a higher level of care during their initial participation in a lower tier.

Although only a small percentage of participants who completed the initial CAT-MH screening assessment and were offered care actually initiated care in one of the tiers (516/3580, 14.41%), there is a wide range of uptake reported in similar digital interventions, with the rate in this study falling within that range [5,49,50], and promising outcomes emerged among those who did participate. Additionally, a prior analysis of racial/ethnic differences in STAND treatment uptake midway through the recruitment period found that students from ethnic/racial minority groups were not less likely to enroll or engage in treatment compared with non-Hispanic White students (see [51]), indicating that STAND has the potential to reach and be acceptable to a diverse student population, which is important for our current work that aims to implement STAND in underserved communities. Of note, although significant improvement over time was observed across all outcome measures within each tier, a relatively lower percentage of participants in tier 2 showed clinically meaningful change across outcomes than tiers 1 and 3, suggesting that there may be some participants who were assigned to tier 2 who may have benefited more from a higher level of care. Our current work is examining whether including other variables (eg, life stress, trauma, demographics) in multivariate predictive models alongside symptom severity can improve upon and personalize tier triaging to improve clinical outcomes.

Taken together, these findings support the feasibility, acceptability, and preliminary effectiveness of STAND; and this initial demonstration of the STAND model of care provided

an opportunity to identify areas for improvement and refinement of the model, including increasing engagement and retention in care.

Limitations

The data presented represent the initial implementation of this system in an open trial. Therefore, conclusions about the effectiveness of STAND compared with another treatment cannot be drawn. Further, as is common practice with initial implementations that aim to iteratively refine and improve upon methodology, some minor data-driven changes were made throughout the course of the study (eg, changing CAT-DI categories based on incoming psychometric information, shifting from group to individual coaching sessions). Another limitation was the low response rate, particularly in tier 0 and following the *ThisWayUp* intervention in tiers 1 and 2. This low response rate precludes our ability to evaluate the longer-term outcome in the lower tiers. Finally, with regard to booster treatment during the monitoring period following the initial acute phase of care, in the present demonstration of STAND, we opted to provide the digital therapy to all participants who had completed tier 3 treatment. However, we did not track usage during this preventative/maintenance phase. Future work we have planned will more precisely identify those who may need this additional booster and at what point in their trajectory. Finally, we did not systematically collect demographic data (eg, gender, race/ethnicity) on coaches and clinicians, which would have been interesting to examine both descriptively and correlations between participant/clinician match on demographic factors and clinical outcomes/engagement. Our current work is now collecting these data on coaches in tier 2 as well as conducting a randomized trial to examine the potential impact of racial/ethnic matching between coaches and participants on engagement and clinical outcomes.

Lessons Learned and Current Directions in STAND

Two notable observations in this initial implementation of STAND were the low initial uptake of STAND services, and the relatively modest ongoing engagement in the digital therapy tiers. With regard to the former, the vast minority of students who were screened and eligible to receive services in STAND actually engaged in care. Despite our multipronged approach to recruitment, and the possibility that a lack of financial incentive to participants in the study may have contributed to some extent in low uptake, improvements to our recruitment strategies are clearly needed. Specifically, additional research is needed to identify recruitment strategies (possibly tailored to each new population that adopts STAND) that provide messaging that resonates with the student population. We are currently partnering with social marketing experts to use inclusive, nonstigmatizing messaging and improve upon our outreach efforts; and current research projects underway are seeking feedback from students who did not engage in care to identify barriers and develop solutions.

With regard to the latter, consistent with prior evidence [49,52], retention in digital therapy was somewhat problematic, and utilization of coaching support was particularly low. This is a ubiquitous challenge in digital therapy research for which creative solutions are being developed and evaluated (eg, the use of gamification, avatars, personalization, feedback, and individual support; see [53] for a review). Although our digital therapy approach already incorporates several of these features, we have made several changes to improve retention in digital therapy and coaching, which are currently being implemented and evaluated in ongoing trials. These include automatic scheduling of all digital therapy participants for coaching sessions, rather than making it an add-on option, and personalizing digital therapy by switching from a unified, one-size-fits-all approach to a modularized suite of digital components that are selected using embedded measurement systems to match an individual's presenting concerns. The change to a personalized approach to module selection is expected to increase retention given evidence for students to prefer tailored online therapy [54]. Current studies are also underway to develop and evaluate engagement strategies (with an emphasis on culturally responsive messaging) and to implement text messaging as a method to increase engagement.

Other changes to our current iteration of STAND include collapsing tiers 1 and 2 into 1 tier (tier 2), with streamlined coach training and supervision, as well as provision of a self-guided wellness digital program in the original tier 0 (now called tier 1). Given that we found such low engagement in tier 0 (the monitoring-only program), we were unable to analyze data collected from tier 0 participants. Possibly, students did not see the benefit of participating in a monitoring-only program when they had minimal to no symptoms (and were not being financially compensated). Our expectation is that by providing digital prevention tools framed as stress management and wellness, students with no symptoms will be more interested in participating in STAND.

Conclusions

The need for effective and scalable treatments for depression and anxiety is enormous given the high (and rising) rates of these disorders and pervasive lack of availability of accessible, evidence-based care. The Screening and Treatment of Anxiety and Depression (STAND) system of care was developed to address this need. STAND is based on the principles of stratified stepped care, adaptation of care in response to continuous symptom monitoring, and routine detection and prevention of suicidality. The STAND system of care was tested in a large sample of UCLA students, where it led to significant reductions in depression and anxiety symptoms and suicide risk. Efforts are underway to continue to refine and enhance the STAND system of care, while increasing its reach, with an emphasis on reaching underserved, diverse populations, conducting randomized clinical trials to identify the optimal ways to deliver STAND, and developing a pathway for sustainability of STAND as it is rolled out on a larger scale. These sustainability efforts will include stakeholder input and economic analyses to identify the specific resources needed to carry out each component of STAND, including ensuring that a sufficient clinical and coaching workforce is available and trained to fidelity in partnership with a college campus that wishes to implement STAND.

Acknowledgments

We thank Ariel Seroussi, MD, Jonathan Heldt, MD, Amanda Gorlick, PhD, and Raphael Rose, PhD, for their clinical supervisory roles in the project, as well as the numerous clinicians, coaches, and research staff who supported this project.

Conflicts of Interest

None declared.

References

1. Fang Y, Ji B, Liu Y, Zhang J, Liu Q, Ge Y, et al. The prevalence of psychological stress in student populations during the COVID-19 epidemic: a systematic review and meta-analysis. Sci Rep 2022 Jul 15;12(1):12118 [FREE Full text] [doi: 10.1038/s41598-022-16328-7] [Medline: 35840641]
2. Lipson S, Lattie EG, Eisenberg D. Increased Rates of Mental Health Service Utilization by U.S. College Students: 10-Year Population-Level Trends (2007-2017). Psychiatr Serv 2019 Jan 01;70(1):60-63 [FREE Full text] [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]
3. Andrade L, Alonso J, Mneimneh Z, Wells J, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol. Med 2013 Aug 09;44(6):1303-1317 [FREE Full text] [doi: 10.1017/s0033291713001943]
4. Cross S, Hickie I. Transdiagnostic stepped care in mental health. Public Health Res Pract 2017 Apr 27;27(2):1 [FREE Full text] [doi: 10.17061/phrp2721712] [Medline: 28474049]
5. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
6. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: A meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
7. Pauley D, Cuijpers P, Papola D, Miguel C, Karyotaki E. Two decades of digital interventions for anxiety disorders: a systematic review and meta-analysis of treatment effectiveness. Psychol. Med 2023;53(2):267-279 [FREE Full text] [doi: 10.1017/s0033291721001999]
8. Schueller S, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: The efficiency model of support. Clinical Psychology: Science and Practice 2017 Mar;24(1):27-45 [FREE Full text] [doi: 10.1037/h0101740]
9. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 2006 Jan;163(1):28-40. [doi: 10.1176/appi.ajp.163.1.28] [Medline: 16390886]
10. Gunn J, Wachtler C, Fletcher S, Davidson S, Mihalopoulos C, Palmer V, et al. Target-D: a stratified individually randomized controlled trial of the diamond clinical prediction tool to triage and target treatment for depressive symptoms in general practice: study protocol for a randomized controlled trial. Trials 2017 Jul 20;18(1):342 [FREE Full text] [doi: 10.1186/s13063-017-2089-y] [Medline: 28728604]
11. LaLonde L, Good J, Orkopoulou E, Vriesman M, Maragakis A. Tracing the missteps of stepped care: Improving the implementation of stepped care through contextual behavioral science. Journal of Contextual Behavioral Science 2022 Jan;23:109-116 [FREE Full text] [doi: 10.1016/j.jcbs.2022.01.001]
12. Wells J, Browne MO, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, Angermeyer MC, et al. Drop out from out-patient mental healthcare in the World Health Organization's World Mental Health Survey initiative. Br J Psychiatry 2013 Jan;202(1):42-49. [doi: 10.1192/bjp.bp.112.113134] [Medline: 23174514]
13. Delgadillo J, Ali S, Fleck K, Agnew C, Southgate A, Parkhouse L, et al. Stratified Care vs Stepped Care for Depression: A Cluster Randomized Clinical Trial. JAMA Psychiatry 2022 Feb 01;79(2):101-108 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.3539] [Medline: 34878526]
14. Firth N, Barkham M, Kellett S. The clinical effectiveness of stepped care systems for depression in working age adults: a systematic review. J Affect Disord 2015 Jan 01;170:119-130 [FREE Full text] [doi: 10.1016/j.jad.2014.08.030] [Medline: 25240141]
15. Ho FY, Yeung W, Ng TH, Chan CS. The Efficacy and Cost-Effectiveness of Stepped Care Prevention and Treatment for Depressive and/or Anxiety Disorders: A Systematic Review and Meta-Analysis. Sci Rep 2016 Jul 05;6:29281 [FREE Full text] [doi: 10.1038/srep29281] [Medline: 27377429]
16. van Straten A, Hill J, Richards D, Cuijpers P. Stepped care treatment delivery for depression: a systematic review and meta-analysis. Psychol. Med 2014 Mar 26;45(2):231-246 [FREE Full text] [doi: 10.1017/s0033291714000701]
17. Zimmerman M, McGlinchey JB. Why don't psychiatrists use scales to measure outcome when treating depressed patients? J Clin Psychiatry 2008 Dec;69(12):1916-1919. [doi: 10.4088/jcp.v69n1209] [Medline: 19192467]

18. de Jong K, Conijn JM, Gallagher RA, Reshetnikova AS, Heij M, Lutz MC. Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. Clin Psychol Rev 2021 Apr;85:102002 [FREE Full text] [doi: 10.1016/j.cpr.2021.102002] [Medline: 33721605]
19. Lewis C, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing Measurement-Based Care in Behavioral Health: A Review. JAMA Psychiatry 2019 Mar 01;76(3):324-335 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
20. Constantino MJ, Boswell JF, Bernecker SL, Louis Castonguay MG. Context-Responsive Psychotherapy Integration as a Framework for a Unified Clinical Science: Conceptual and Empirical Considerations. Journal of Unified Psychotherapy and Clinical Science 2013;2(1):1-20 [FREE Full text]
21. Hickie I, Scott EM, Cross SP, Iorfino F, Davenport TA, Guastella AJ, et al. Right care, first time: a highly personalised and measurement-based care model to manage youth mental health. Med J Aust 2019 Nov;211 Suppl 9:S3-S46. [doi: 10.5694/mja2.50383] [Medline: 31679171]
22. Carpenter JS, Iorfino F, Cross SP, Davenport TA, Hermens DF, Rohleder C, et al. Combining clinical stage and pathophysiological mechanisms to understand illness trajectories in young people with emerging mood and psychotic syndromes. Medical Journal of Australia 2019;211(supplement 9):S12-S22. [doi: 10.5694/mja2.50383]
23. Cross SP, Davenport TA, Scott EM, Iorfino F, Sawrikar V, Hickie IB. A service delivery model to support highly personalised and measurement-based care in youth mental health. Med J Aust 2019;211(Supplement 9):S42-S46 [FREE Full text]
24. Hamilton M, Hetrick SE, Mihalopoulos C, Baker D, Browne V, Chanen AM, et al. Identifying attributes of care that may improve cost-effectiveness in the youth mental health service system. Med J Aust 2017 Nov 20;207(10):S27-S37. [doi: 10.5694/mja17.00972] [Medline: 29129184]
25. Hickie I, Carpenter JS, Iorfino F, Scott E, Cross S, Hermens DF. The Utility of Clinical Staging in Youth Mental Health Settings: Neurobiological and Longitudinal Data from Sydney-Based Studies of Transdiagnostic Cohorts. In: McGorry P, Hickie I, editors. Clinical Staging in Psychiatry: Making Diagnosis Work for Research and Treatment. Cambridge, MA: Cambridge University Press; 2019:81-102.
26. Iorfino F, Cross SP, Davenport T, Carpenter JS, Scott E, Shiran S, et al. A Digital Platform Designed for Youth Mental Health Services to Deliver Personalized and Measurement-Based Care. Front Psychiatry 2019;10:595 [FREE Full text] [doi: 10.3389/fpsyt.2019.00595] [Medline: 31507465]
27. Adrian M. Adaptive Interventions Are Required for Efficient and Personalized Adolescent Depression Treatment. J Am Acad Child Adolesc Psychiatry 2019 Jan;58(1):16-17. [doi: 10.1016/j.jaac.2018.07.892] [Medline: 30577932]
28. Centers for Disease Control and Prevention (CDC). Web-based Injury Statistics Query and Reporting Syste4m (WISQARS). CDC. 2020. URL: https://www.cdc.gov/injury/wisqars/index.html [accessed 2022-12-01]
29. Jacobson CM, Gould M. The epidemiology and phenomenology of non-suicidal self-injurious behavior among adolescents: a critical review of the literature. Arch Suicide Res 2007;11(2):129-147. [doi: 10.1080/13811110701247602] [Medline: 17453692]
30. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
31. Klonsky ED. Non-suicidal self-injury in United States adults: prevalence, sociodemographics, topography and functions. Psychol. Med 2011 Jan 05;41(9):1981-1986. [doi: 10.1017/s0033291710002497]
32. Substance Abuse and Mental Health Service Administration. Key substance use and mental health indicators in the United States: Results from the 2018 National Survey on Drug Use and Health (HHS Publication No. PEP19-5068, NSUDH Series H-54). Substance Abuse and Mental Health Service Administration. Rockvile, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2019. URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf [accessed 2023-07-19]
33. Braciszewski J. Digital Technology for Suicide Prevention. Advances in Psychiatry and Behavioral Health 2021 Sep;1(1):53-65 [FREE Full text] [doi: 10.1016/j.ypsc.2021.05.008]
34. Achtyes ED, Halstead S, Smart L, Moore T, Frank E, Kupfer DJ, et al. Validation of Computerized Adaptive Testing in an Outpatient Nonacademic Setting: The VOCATIONS Trial. Psychiatr Serv 2015 Oct;66(10):1091-1096 [FREE Full text] [doi: 10.1176/appi.ps.201400390] [Medline: 26030317]
35. Rush AJ, Trivedi MH, Wisniewski SR, Stewart JW, Nierenberg AA, Thase ME, et al. Bupropion-SR, Sertraline, or Venlafaxine-XR after Failure of SSRIs for Depression. N Engl J Med 2006 Mar 23;354(12):1231-1242. [doi: 10.1056/nejmoa052963]
36. World Health Organization (WHO). WHO menu of cost-effective interventions for mental health. WHO. Geneva, Switzerland: WHO; 2021. URL: https://apps.who.int/iris/handle/10665/343074 [accessed 2023-07-11]
37. Newby J, Mewton L, Williams AD, Andrews G. Effectiveness of transdiagnostic Internet cognitive behavioural treatment for mixed anxiety and depression in primary care. J Affect Disord 2014 Aug;165:45-52. [doi: 10.1016/j.jad.2014.04.037] [Medline: 24882176]

38. Mahoney A, Li I, Haskelberg H, Millard M, Newby JM. The uptake and effectiveness of online cognitive behaviour therapy for symptoms of anxiety and depression during COVID-19. J Affect Disord 2021 Sep 01;292:197-203 [FREE Full text] [doi: 10.1016/j.jad.2021.05.116] [Medline: 34126311]
39. Rosenberg B, Kodish T, Cohen ZD, Gong-Guy E, Craske MG. A Novel Peer-to-Peer Coaching Program to Support Digital Mental Health: Design and Implementation. JMIR Ment Health 2022 Jan 26;9(1):e32430 [FREE Full text] [doi: 10.2196/32430] [Medline: 35080504]
40. Craske M. Honoring the Past, Envisioning the Future: ABCT's 50th Anniversary Presidential Address. Behav Ther 2018 Mar;49(2):151-164. [doi: 10.1016/j.beth.2017.05.003] [Medline: 29530256]
41. Holmes EA, Ghaderi A, Harmer CJ, Ramchandani PG, Cuijpers P, Morrison AP, et al. The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. The Lancet Psychiatry 2018 Mar;5(3):237-286. [doi: 10.1016/s2215-0366(17)30513-8]
42. Craske MG, Herzallah MM, Nusslock R, Patel V. From neural circuits to communities: an integrative multidisciplinary roadmap for global mental health. Nat. Mental Health 2023 Jan 17;1(1):12-24 [FREE Full text] [doi: 10.1038/s44220-022-00012-w]
43. Craske MG. Cognitive Behavior Therapy, 2nd Edition. New York, NY: American Psychological Association; 2017.
44. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
45. Gibbons R, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, et al. Development of the CAT-ANX: a computerized adaptive test for anxiety. Am J Psychiatry 2014 Feb;171(2):187-194 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13020178] [Medline: 23929270]
46. Gibbons R, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, et al. Development of a computerized adaptive test for depression. Arch Gen Psychiatry 2012 Nov;69(11):1104-1112 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2012.14] [Medline: 23117634]
47. Gibbons R, Kupfer D, Frank E, Moore T, Beiser DG, Boudreaux ED. Development of a Computerized Adaptive Test Suicide Scale—The CAT-SS. J. Clin. Psychiatry 2017 Dec 27;78(9):1376-1382 [FREE Full text] [doi: 10.4088/jcp.16m10922]
48. Graham A, Minc A, Staab E, Beiser DG, Gibbons RD, Laiteerapong N. Validation of the Computerized Adaptive Test for Mental Health in Primary Care. Ann Fam Med 2019 Jan;17(1):23-30 [FREE Full text] [doi: 10.1370/afm.2316] [Medline: 30670391]
49. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
50. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol. Med 2008 Sep 24;39(5):705-712 [FREE Full text] [doi: 10.1017/s0033291708004224]
51. Kodish T, Lau AS, Gong-Guy E, Congdon E, Arnaudova I, Schmidt M, et al. Enhancing Racial/Ethnic Equity in College Student Mental Health Through Innovative Screening and Treatment. Adm Policy Ment Health 2022 Mar;49(2):267-282 [FREE Full text] [doi: 10.1007/s10488-021-01163-1] [Medline: 34505211]
52. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
53. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding Engagement Strategies in Digital Interventions for Mental Health Promotion: Scoping Review. JMIR Ment Health 2021 Dec 20;8(12):e30000 [FREE Full text] [doi: 10.2196/30000] [Medline: 34931995]
54. Batterham PJ, Calear AL. Preferences for Internet-Based Mental Health Interventions in an Adult Online Sample: Findings From an Online Community Survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]

Abbreviations

CAT-ANX: Computerized Adaptive Test-Anxiety Inventory
CAT-DI: Computerized Adaptive Test-Depression Inventory
CAT-MH: Computer Adaptive Test for Mental Health
CAT-SS: Computerized Adaptive Test-Suicide Scale
CSSRS: Columbia Suicide Severity Rating Scale
STAND: Screening and Treatment for Anxiety and Depression
UCLA: University of California, Los Angeles

Edited by J Torous; submitted 01.03.23; peer-reviewed by A Ghosh, S Locke; comments to author 16.05.23; revised version received 25.05.23; accepted 26.05.23; published 24.07.23.

Please cite as:
Wolitzky-Taylor K, LeBeau R, Arnaudova I, Barnes-Horowitz N, Gong-Guy E, Fears S, Congdon E, Freimer N, Craske M
A Novel and Integrated Digitally Supported System of Care for Depression and Anxiety: Findings From an Open Trial
JMIR Ment Health 2023;10:e46200
URL: https://mental.jmir.org/2023/1/e46200
doi:10.2196/46200
PMID:37486735

Original Paper

Relationship Factors in Internet-Delivered Psychological Interventions for Veterans Experiencing Postpartum Depression: Qualitative Analysis

Cara L Solness[1], PhD; Paul J Holdefer[2], BA; Ti Hsu[2], MS; Emily B K Thomas[2], PhD; Michael W O'Hara[2], PhD

[1]Department of Psychiatry, School of Medicine, University of Colorado, Aurora, CO, United States

[2]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, United States

Corresponding Author:
Cara L Solness, PhD
Department of Psychiatry, School of Medicine
University of Colorado
1890 N Revere Court
Anschutz Medical Campus
Aurora, CO, 80045
United States
Phone: 1 319 331 3566
Email: cara.solness@cuanschutz.edu

Abstract

Background: Internet-delivered psychological interventions (IPIs) have been shown to be effective for a variety of psychological concerns, including postpartum depression. Human-supported programs produce better adherence and larger effect sizes than unsupported programs; however, what it is about support that affects outcomes is not well understood. Therapeutic alliance is one possibility that has been found to contribute to outcomes; however, the specific mechanism is not well understood. Participant perspectives and qualitative methodology are nearly absent from the IPI alliance research and may help provide new directions.

Objective: In this study, we aimed to provide participant perspectives on engagement with an IPI for postpartum depression to help inform alliance research, development of new IPIs, and inform resource allocation.

Methods: A qualitative methodology was used to explore participant perspectives of veteran women's engagement with the MomMoodBooster program, a human-supported internet-delivered intervention for postpartum depression. Participants were asked 4 open-ended questions with the 3-month postintervention survey, "In what ways did you find the MomMoodBooster most helpful?" "How do you think the MomMoodBooster could have been improved?" "In what ways did you find the personal coach calls to be helpful?" and "How do you think the personal coach calls could have been improved?"

Results: Data were collected from 184 participants who responded to at least 1 of the open-ended questions. These were analyzed using thematic analysis and a process of reaching a consensus among coders. The results suggest that not only the engagement with the support person is perceived as a significant contributor to participant experiences while using the MomMoodBooster content but also the relationship factors are particularly meaningful. The results provide insights into the specific qualities of the support person that were perceived as most impactful, such as warmth, empathy and genuineness, and feeling normalized and supported. In addition, the results provide insight into the specific change processes that can be targeted through support interactions, such as encouraging self-reflection and self-care and challenging negative thinking.

Conclusions: These data emphasize the importance of relationship factors between support persons and an IPI program for postpartum depression. The findings suggest that focusing on specific aspects of the alliance and the therapeutic relationship could yield fruitful directions for the training of support personnel and for future alliance-based research of internet-delivered treatments.

(JMIR Ment Health 2023;10:e46061) doi:10.2196/46061

KEYWORDS

internet-delivered treatment; postpartum depression; therapeutic alliance; qualitative methodology; veteran

Introduction

Background

Internet-delivered psychological interventions (IPIs; see the *Terminology* section) for postpartum depression are an important part of expanding the reach of specialized mental health services and thus reducing barriers to treatment. Less than half of those affected by mental health concerns seek or receive evidence-based treatments owing to costs, limited access, stigma, and preference to self-manage [1]. IPIs address these issues and offer anonymity, flexibility, and cost-effectiveness [2]. Several meta-analyses have found IPIs to be efficacious for treating depression [3-6] and specifically for postpartum depression [7].

Previous research has demonstrated that human-supported IPIs are more effective than unsupported ones [7-9]. However, what it is about support that makes treatment more effective is not well understood [10]. Research incorporating the perspectives of those who use IPIs may provide insight into these mechanisms. Given the importance of resource allocation, more work is needed to understand the specific mechanisms that account for the effect of support, and qualitative methods are central to this endeavor [10,11].

Therapeutic alliance is one possible explanation for the effects of support. Therapeutic alliance, as most commonly defined in the IPI literature, is based on the definition by Bordin [12,13] who suggests the conceptualization of a positive emotional bond between a therapist and a client, including mutual agreement on the goals and tasks of the treatment. In face-to-face (FTF) therapy, the therapeutic alliance is widely regarded as an important pantheoretical ingredient and accounts for approximately 8% of the variability in outcomes in FTF therapy [14,15]. Meta-analyses examining the association between therapeutic alliance and outcomes for IPIs indicate that there is a small and significant correlation between therapeutic alliance and various treatment outcomes (eg, anxiety disorders, anxiety related to preterm labor, obsessive compulsive disorder, depression, posttraumatic stress disorder, psychosis, schizophrenia, and tinnitus) [16,17]. Furthermore, this association does not appear to depend on the frequency of contact or on the mode of contact with the therapist or on the availability of self-help content [16]. The association between therapeutic alliance and treatment outcomes also does not depend on the amount of contact, the occurrence of FTF contact, or the timing of the alliance assessment [17].

Within the growing body of literature examining the therapeutic alliance in IPIs in general [16,18] and those examining depression outcomes specifically [10], some studies explore the alliance between participants and the computerized content [19], whereas others examine the alliance between participants and the support personnel [20,21]. In general, ratings of alliance in IPIs have been found to be similar to those in FTF therapies, and these ratings are positively associated with outcomes, although they do not reach statistical significance in some studies [22,23]. The systematic review of the alliance literature for guided IPIs by Pihlaja et al [23] examined only studies for which support was delivered via email and similarly found that the strength of the alliance was predictive of outcomes in some studies, whereas in others, the direction was positive but did not reach statistical significance. Other authors have concluded that the IPI alliance literature has provided mixed results [18]. However, more recent work on the alliance in IPIs such as the meta-analysis by Probst et al [17] found high homogeneity of effect sizes among their included studies. To summarize, although previous research on the role of the therapeutic alliance in IPIs provided some support but with mixed results, newer meta-analyses are less ambiguous about its contribution to various mental health outcomes [16,17].

Out of various IPIs, internet-delivered cognitive behavioral therapy (iCBT) has emerged as the most common and most studied intervention [18,24]. Previous research on iCBT has also examined the contribution of therapeutic alliance in this specific approach. For instance, a recent meta-analysis of phone-supported iCBT found a large effect size between alliance and outcomes [16], with the strength of the association only slightly smaller than that of FTF cognitive behavioral therapy (CBT; r=0.20 and r=0.29, respectively). The narrative review of the IPI alliance literature by Berger [18] concludes that guidance in iCBT plays a crucial role in outcomes and that the specific aspects of the alliance and therapist factors that facilitate positive outcomes are not yet well understood in the IPI literature.

Previous studies have examined therapist factors that contribute to the alliance. For example, Hilsenroth et al [25] found positive associations between the alliance and therapist characteristics such as openness, warmth, confidence, flexibility, honesty, tension, use of self-disclosure and negative associations with the alliance for characteristics such as rigidity, defensiveness, self-focus, and others in FTF therapy. The meta-analysis by Nienhuis et al [14] examined the associations between therapist empathy and genuineness and therapeutic alliance in individual FTF therapy and found a moderate association. When examining email-supported iCBT for generalized anxiety disorder, Paxling et al [26] identified the most common therapist behaviors in written correspondence as deadline flexibility, task reinforcement, alliance bolstering, task prompting, psychoeducation, self-disclosure, self-efficacy shaping, and empathetic utterances. Similarly, Holländare et al [27] examined common behaviors within email content between therapists and patients and found positive associations between therapists' behaviors, such as encouraging, guiding, and urging the patients and affirming their thoughts, and outcomes in iCBT for depression. Similar work on factors affecting alliance has not yet been completed specifically for telephone-supported IPIs. Furthermore, it is not possible to establish best practices for the provision of differently delivered support without first understanding the specific mechanisms identified by those using these interventions. As such, there is much to be gained from participant perspectives of their experiences with a telephone-supported IPI.

Questions related to the broad conceptualization of the alliance for IPIs and the extent to which traditional measures accurately represent what is occurring in IPIs are just beginning to be asked. For example, Askjer and Mathiasen [28] raised the question as to how to conceptualize the therapeutic alliance in

blended treatments and suggested that an aggregated alliance (ie, to the program and to the support person or therapist) may be a more accurate representation of alliance in the blended format. Blended formats consist of self-directed content interspersed with FTF sessions. It may be that IPIs with regular, synchronous, phone-based support are more similar to blended treatments than asynchronous, text-supported IPIs are, which raises the question of how alliance is being measured for each IPI format. To address this question, Linder et al [29] compared telephone- and email-supported interventions for depression and found no difference, but their study was limited by a small sample size. The meta-analysis by Kaiser et al [16] provides additional information but leaves some questions that may be addressed via participant perspectives. For example, their meta-analysis found no impact on the association between alliance and outcomes for the frequency of contact with a therapist, mode of contact, or availability of self-help content. However, no information was provided about approach to the contact (ie, accountability based, support based, review based, education based, etc), which leaves open the possibility that categorical coding of support type may not be sufficient for detecting differences. More work is needed in this area.

Current literature examining the alliance commonly uses the Working Alliance Inventory (WAI) [30], which was developed for FTF therapy. Recent literature suggests that the task and goal subscales of the therapeutic alliance are more important than the therapeutic bond subscale in supported IPIs [10,17,18,31]. However, the literature has yet to examine the psychometric properties of the WAI, specifically with IPI users. Askjer and Mathiasen [28] found that clients and therapists rated the goal dimension higher than task and bond and that therapists rated bond higher than clients did; however, the literature does not address whether there is a difference in these subscales depending on the format of support. For example, it may be that IPI users perceive the therapeutic bond differently between these 2 types of support models. The question becomes whether the WAI is sensitive enough to be able to make this kind of distinction and whether it is in fact measuring the same construct in each case. In order to begin to address questions such as these, participant perceptions and experiences of IPIs, the support relationship, and the relative importance of each will be helpful toward the endeavor of conceptualizing alliance within synchronous, telephone-supported IPIs.

Objectives of This Study

To address the gap in the literature of participant experiences with IPIs and the conceptualization of alliance in IPIs, this study examined participant-identified program and coach contributions to their experiences of using a human-supported iCBT delivered to veteran women with postpartum depression. The MomMoodBooster (MMB) program is an empirically supported iCBT intervention [32-34] administered to veteran women in the postpartum period [35] for which support was provided synchronously by telephone. The program from which these data were collected consisted of 6 self-directed, postpartum-specific, web-based CBT–based modules and 6 weekly concurrent phone coaching calls. A previous study described quantitative metrics of engagement and outcomes and provided further information regarding the methods of the intervention study [35]. On average, veterans engaged in 3.35 coaching calls and spent an average of 49.19 minutes on the phone with their coach over the course of their engagement with the program. Effect sizes for depressive symptoms, behavioral activation, and automatic thoughts were large [35]. Because alliance effects for IPIs have been attributed to participant relationships with the IPI programs in some studies [19,28], this analysis quantifies the frequency of responses related to each for comparison.

The findings bear on IPI treatment design, future directions for IPI alliance research, and the training of support personnel. As qualitative analysis is an inductive process, a priori hypotheses were not made.

Terminology

IPIs are used in this manuscript as an umbrella term for mental health interventions that are wholly or partly administered through technology such as the internet or other electronic means. The use of IPIs throughout this paper is intended to refer to the broader literature of electronic-delivered interventions and to acknowledge that alliance is a pantheoretical construct that is present to greater or lesser degrees between users of IPIs and the program itself and the support person. We will use more specific terminology such as iCBT when referring to the data from this study and when relevant in reference to specific cited studies.

Methods

Participants

Participants in this study were a subset of those included in the program evaluation for MMB for veteran mothers [35]. Of the final sample (N=200) from the program evaluation, 184 (92%) veterans provided at least 1 response to the 4 open-ended questions and were therefore included in this analysis. A total of 758 individual responses were provided.

Inclusion and Exclusion Criteria

The study sample consisted of veteran women enrolled in the MMB program who had completed a follow-up assessment at the 3-month postenrollment time point and had provided responses to at least 1 of the qualitative questions. Participants were excluded from the analysis if they did not provide a response to any of the 4 qualitative questions (16/200, 8%). Individual responses (10/758, 1.3%) were excluded if they were uninterpretable or ambiguous due to indecipherable words or uninterpretable brevity.

Procedure

Participants for the larger MMB study were identified nationwide through the Veterans Affairs database of veterans whose private obstetrical care was paid for by the Veterans Affairs. Full recruitment and enrollment descriptions can be found in the initial program evaluation [35]. Participants were sent a follow-up questionnaire 3 months after the initial screening and enrollment in the MMB program, and the questionnaire included self-rated outcomes measures and 4 open-ended questions about the program, for which participants had unlimited characters to comment. Participants were free to

skip questions that they did not wish to answer. Several participants opted not to answer any of the qualitative questions, whereas others answered only select questions. All the responses for the cohort included in the program analysis paper were included in this analysis [35]. The questions are referenced throughout the manuscript using the headings *program helpful*, *program improve*, *coach helpful*, and *coach improve* and are as follows:

1. Program helpful: "In what ways did you find the MomMoodBooster most helpful?"
2. Program improve: "How do you think the MomMoodBooster could have been improved?"
3. Coach helpful: "In what ways did you find the personal coach calls to be helpful?"
4. Coach improve: "How do you think the personal coach calls could have been improved?"

Analysis

A modified iterative thematic analysis approach was used for data analysis. Coders used a consensual qualitative approach to the coding process and tallied themes to indicate the relative representation within the thematic structure. The six stages of thematic analysis delineated by Rennie [36] are as follows: (1) familiarization with the data via deep emersion, (2) systematic coding with collation of similar codes, (3) deriving themes and gathering relevant codes, (4) reviewing themes and checking for fit across individual codes and entire data set, (5) defining themes and ongoing refinement, and (6) producing the report. The first 2 authors independently assigned codes to each datum and then met regularly for an iterative process of reaching a consensus. Once a consensus was reached, an internal auditor reviewed the codes and provided feedback. The coders then reviewed the auditor comments until consensus was reached again, wherein the 2 coders resumed independent coding of domains and themes before meeting again to build consensus. The processes of internal audit and subsequent review and revision were undertaken at this stage as well.

Positionality

The first author identifies as a mother who experienced postpartum depression and is a coach for the MMB program with 6 years of experience, including working with 105 participants in this sample. In addition, the first author has significant clinical experience working with women in the postpartum period. The second author identifies as a preparenthood man who functioned as the primary recruiter and contact person for the participants and was responsible for data management. The first and second authors acknowledged their experiences and power differences and worked to mitigate the effect of these factors during the coding process. The third author, who functioned as the auditor for the coding process, identifies as a woman without direct experience of motherhood or postpartum depression. This author, however, is familiar with the structure and content of the program in her role as a phone coach, commencing after this sample was collected. The fourth author identifies as a woman and is the principal investigator of this project. The fifth author identifies as a man and was the creator of the larger MMB for veteran women research project. He has worked in the area of perinatal depression research for 42 years. The fourth and fifth authors were involved in manuscript review and did not participate in data analysis. As a team, all authors are invested in women's and veteran's mental health and as such are most interested in how to improve the services provided through the MMB program.

Given the coders' close ties to the program and direct experience with participants, the research team anticipated that relational aspects of the coaching interactions would emerge as important in the data. In order to check their biases related to the importance of the relationship, ambiguous codes referencing the coach were not coded as interpersonal unless the participant directly named their specific coach; however, they may have eventually been subsumed under interpersonal during the process of distilling down to themes. Coders worked hard to challenge each other with respect to the interpretation of interpersonal themes.

Trustworthiness

The research team took several steps to attend to trustworthiness. The first was identifying positionality [37] to be transparent about their connection to the study participants and potential biases. Second, the team used an internal auditor and the process of consensus [38] throughout all the stages of coding and assigning to themes and domains. Third, the team journaled about their process during coding to document thought processes, challenges, and decision points. These have been included in the general audit trial [37]. Finally, an external auditor, a qualitative method expert who has no affiliation with the MMB program performed an audit of the coding process and randomly selected codes, the final thematic structure, and overall adherence to trustworthiness. No concerns were noted or changes were suggested.

Ethics Approval

This study was approved by the University of Iowa Institutional Review Board (201310766 [IRB-03] and 201603779 [IRB-01]).

Results

Overview

The final sample consisted of 184 veterans who provided at least 1 qualitative response. Responses to the 4 questions from these participants were coded into a total of 748 codes due to multiple codes being assigned to some responses. Of the 748 total codes, 390 (52.1%) were program related and 358 (47.9%) were coach related. The 184 participants answered the 4 questions that are referred by the following headings: program helpful (n=176, 95.7%); program improve (n=142, 77.2%); coach helpful (n=162, 88%); and coach improve (n=118, 64.1%).

The thematic analysis resulted in a final structure consisting of five domains: (1) program accessibility and functionality, (2) content, (3) coaching, (4) change processes, and (5) barriers.

Table 1 delineates the thematic structure by including frequencies for each domain, and Multimedia Appendix 1 displays the thematic structure with definitions and example quotes. Quotes from participants have been numbered in the order they appear in this manuscript to further protect

anonymity. Each participant number represents a unique individual, with relevant quotes provided by a wide range of participants.

Table 1. Relative representation of codes.

Domains, themes, and subthemes	Relative frequency, n (%)[a]
Program accessibility and functionality	
All program-related codes (n=390)	
Positive	17 (4.4)
Improve (n=55)	55 (14.1)
Access	13 (23.6)
Delivery mode	16 (29.1)
Functionality	11 (20)
Technical	15 (27.3)
Content (n=390)	
Positive	31 (7.9)
Improve	61 (15.6)
Specific helpful content	45 (11.5)
Coaching	
All coach-related codes (n=358)	
General positives	36 (10.1)
Amount (n=34)	34 (9.5)
Increase	26 (76.5)
Did not want, did not receive, or got no benefit	6 (17.6)
Decrease	2 (5.9)
Change processes	
Program (n=390)	
Intrapersonal	108 (27.7)
Coach	34 (8.7)
Accountability	3 (0.8)
Coach (n=358)	
Intrapersonal	41 (11.5)
Relationship or interpersonal	98 (27.4)
Qualities	69 (19.3)
Practical assistance	32 (8.9)
Accountability	33 (9.2)
Barriers (n=748)	
Time	12 (1.6)
Motivation and mood	3 (0.4)
Contextual factors	7 (0.9)
Scheduling or rescheduling problems	16 (2.1)
Coaching delivery method	6 (0.8)

[a]Percentages represent the relative number of responses for each theme or subtheme with the respective sample.

Domain 1: Program Accessibility or Functionality

The *program accessibility/functionality* domain refers to comments addressing how participants accessed or would like to have accessed the content and how it functioned from a technical standpoint. This domain includes two themes: (1) *positive* (17/390, 4.4% of the program-related codes), which refers to nonspecific positive comments such as the one shared by participant 1: "Convenient, self-paced"; and (2) *improve* (55/390, 14.4% of the program-related codes). The *improve* theme captures four subthemes: (1) *access*, (2) *delivery mode*, (3) *functionality*, and (4) *technical*. The subtheme *access*, which represented 24% (13/55) of the *improve* codes, refers to recommendations related to the timing or duration of access to the program. For example, participant 2 suggested, "given at discharge from the hospital." The subtheme *delivery mode*, which represented 29% (16/55) of the *improve* codes, refers to alternative modes of delivery that participants felt would have been beneficial such as those noted by participant 3: "Offered in a classroom setting." The *functionality* subtheme, which represented 20% (11/55) of the *improve* codes, identified electronic formatting that would have facilitated greater engagement; for example, participant 4 suggested the program be, "More game like, not literature." Finally, the *technical* subtheme, which represented 27% (15/55) of the *improve* codes, identified technological issues that affected participant experience such as the one commented by participant 5: "I had issues with inputting my daily mood."

Domain 2: Content

The *content* domain refers to comments identifying specific content from the program or more general comments made under the *program helpful and program improve* questions. This domain includes three themes: (1) *positive*, (2) *improve*, and (3) *specific helpful content*. *Positive*, referring to positive comments that were specific to content, accounted for 7.9% (31/390) of the program-related codes; for example, participant 6 noted, "I enjoyed the interactive [aspects]." *Improve*, accounting for 15.6% (61/390) of the program-related codes, refers to specific suggestions for ways the content could have better met participant needs such as the one requested by participant 7: "Make the program more diverse." The *specific helpful content* subtheme, representing 11.5% (45/390) of the program-specific codes, refers to codes in which participants named specific pieces of content that they appreciated; for example, participant 8 noted, "Identifying negative thoughts and downward mood spiral."

Domain 3: Coaching

The *coaching* domain refers to general comments relating to the coach or the amount of coaching from any of the 4 questions. This domain consists of two themes: (1) *general positives*, accounting for 10.1% (36/358) of the coach-related codes, and (2) *amount*, accounting for 9.5% (34/358) of the coach-related codes. The codes under the theme *general positives* were nonspecific satisfaction comments such as the one commented by participant 9: "No improvements needed." Three subthemes emerged under the *amount* theme: (1) the *increase* subtheme, which accounted for 77% (26/34) of the codes, was exemplified by participant 10, who expressed, "Continuation a week or two after the program ended," and participant 11, who shared, "More phone calls"; (2) the *didn't want, didn't receive, or got no benefit* subtheme was less common, representing only 18% (6/34) of the *amount* codes; for example, participant 12 expressed, "What calls?? Received none," and participant 13 stated, "I never spoke with anyone, that is not what I wanted from the program"; and (3) the *decrease* subtheme was the least common, representing only 6% (2/34) of the *amount* codes; for example, participant 14 stated, "More online activities, less calls." (see Figure S1 in Multimedia Appendix 2 for visual depiction).

Domain 4: Change Processes

The fourth domain, *change processes*, refers to specific intrapersonal, relational, and practical influences that participants identified. These are captured by two themes: (1) *program* and (2) *coach*, with 3 and 5 subthemes, respectively. The *program* theme refers to codes that identified change processes either specifically related to program content or nonspecifically under the *program helpful* question. The *coach* theme refers to codes that specifically indicated a link to coaching or were responses under the *coach helpful* or *coach improve* questions. Replicated subthemes (interpersonal and accountability) across *program* and *coach* themes emerged through the inductive coding process. Coding these subthemes separately, rather than collapsing, was prioritized to compare the relative contributions across themes (see Figure S2 in Multimedia Appendix 2 for visual depiction).

Under the *program* theme, the first subtheme, *intrapersonal*, refers to the internal process of personal growth and change identified by the participants. Of the 390 program-related codes, 108 (27.7%) codes were included under this subtheme. These included codes were managing negative thoughts, increased self-awareness, learning coping strategies, increased agency, and others. Participant 15 commented, "I learned how to be calm and take care of my baby." Some participants indicated that the coach was the most helpful aspect of the program; for example, in response to the *program helpful* question, participant 16 commented, "Helpful to talk to someone who is supporting." Of the 390 program-related codes, 34 (8.7%) codes referenced the coach and were therefore included as a subtheme of the program-related change processes. These were coded separately from the *coach* theme for 2 reasons. First, these comments were more general than those captured under the *coach* theme and second, because it felt important to differentiate and identify how often the coach was named as the salient helpful aspect of the program as a whole. The third subtheme under *program* theme was *accountability*. This subtheme refers to comments indicating that being held accountable was the most helpful aspect of the program. Only 3 respondents identified *accountability* in response to the *program helpful* question, and it represented 0.8% (3/390) of the program-related codes. Interestingly, 2 of these participants also named relationship factors in other responses; for example, participant 17 who commented under *program helpful*, "Having a system to be accountable to," which was coded as *accountability*, also commented under *coach helpful* "having a live person to check in with is very comforting," which was coded as *relationship/interpersonal* (which are subthemes under *coach* theme).

The second theme under the *change processes* domain is *coach*, which included 5 subthemes. *Intrapersonal* (41/358, 11.5% of the coach-related codes) refers to growth and change identified by the participant specifically credited to the coaching interaction. Some of the intrapersonal processes include self-reflection, challenging negative thinking, perspective-taking, increased self-compassion, normalizing, and processing problems. For example, participant 18 commented, "Helped me reflect on myself and identify negative talk to myself." (see Figure S3 in Multimedia Appendix 2 for visual depiction).

The second *coach* subtheme, *relationship/interpersonal* (98/358, 27.4%), refers to codes that referenced the importance of the relationship between the participant and the coach. Some of the codes in this domain were requests from participants for what they would like to be different in the coaching interactions such as the one shared by participant 19: "more input, deeper conversation," which was coded as *increase depth*. Others were experienced by the participant; for example, participant 20 shared "[specific coach name] listens even if I haven't completed the week and gives praise and good thoughts," which was coded as *relationship* because of the use of the coach's name and *coach qualities, active listener and supportive*. Participant 21 shared, "I really liked my coach" which was coded as *interpersonal relationship*.

The third *coach* subtheme, *qualities*, represented 19.3% (69/358) of the coach-related codes and refers to specific attributes named by participants. For example, the comment by participant 22, "Felt very genuine," was coded as *genuine*, and the comment by participant 23, "No matter what was going on in my life she always empathized and worked with me," was coded as *empathetic*. Note that this response, which is similar to the one shared by participant 20 above, was not coded as *relationship*, as the participant did not name her specific coach (see Figure S4 in Multimedia Appendix 2 for visual depiction).

The fourth *coach* subtheme, *practical assistance*, refers to codes that identified specific practical ways that coaches helped participants such as assistance with technical difficulties and clarification of content. This subtheme accounted for 8.9% (32/358) of the coach-related codes.

The fifth *coach* subtheme, *accountability*, refers to codes that specifically identified the coach as the agent of holding participants accountable. These codes represented 9.2% (33/358) of the coach-related codes. Interestingly, 24% (8/33) of the participants who provided responses that were coded under *coachaccountability* subtheme also provided responses that were coded under *relationship/interpersonal* subtheme.

Domain 5: Barriers

The final domain refers to the challenges that respondents named that prevented them from engaging with the content and coaching. Five themes were captured under *barriers* including (1) time, (2) motivation and mood, (3) coaching delivery method, (4) contextual factors, and (5) scheduling and rescheduling problems (representing n=12, 1.6%; n=3, 0.4%; n=7, 0.9%; n=16, 2.1%; and n=6, 0.8%; respectively, of all 748 codes). Table 1 provides the relative frequency of the themes under *barrier* domain, and Multimedia Appendix 1 provides descriptions and selected quotes for each of these themes.

Discussion

This study aimed to illuminate participant perspectives regarding engagement with a telephone-supported IPI for postpartum depression, the iCBT MMB program. Results of this study will inform future IPI alliance research, training of support personnel, program development, and allocation of resources for supported IPIs.

Principal Findings

Our thematic analysis of 184 veterans' responses to 4 open-ended questions regarding MMB program content and impressions of the support person provided a unique look at participant perspectives of this supported iCBT. Furthermore, we were able to compare participant perspectives across program content and the support person specifically. Overall, 3 of the 5 identified domains, *program content*, *accessibility/functionality*, and *barriers*, provide ideas for program developers regarding particularly helpful aspects of program content and design. In addition, some of the information found in the *intrapersonal* theme under *change processes* domain may identify specific processes that can be targeted via content in future human-supported programs.

More significantly, these data illuminate some of the important change processes facilitated by the support relationship and specific qualities of the support person that elicit participant responses. For example, the intrapersonal subthemes and relationship or interpersonal subthemes represent the largest proportion of codes. This indicates that interpersonal processes and relationship factors may warrant further investigation in the context of alliance-based research for iCBT specifically and IPIs more generally. Although some program-related intrapersonal processes are clearly attributable to the program content (eg, video content that served a normalizing purpose), others may be attributed to either the content or the coach. Without a clear reference to one or the other, such processes were assigned to the content. However, given the number of participants who referred to the coach under the *program helpful* question, it is clear that some were considering the entirety of the program, rather than strictly the content, when referencing intrapersonal processes. Future IPI alliance research may want to be mindful of construct validity issues and whether questionnaires probe intrapersonal or interpersonal processes. Similarly, qualitative researchers may want to be mindful of how they query these processes.

Our findings also suggest that the coach is perceived as one of the most important aspects of the program, as approximately 20% of the responses under *program helpful* referenced the coach. Furthermore, of the participants who identified the amount of time with the coach under *coach improve*, approximately 3 times as many indicated wanting more contacts compared with the combined total across *no change*, *didn't need*, or *decrease* codes, suggesting that participants valued the support time and, in many cases, wished for more. Significantly, approximately 60% of the coach codes (from the *coach helpful*

and *coach improve* questions) identified relationship or interpersonal factors, and >40% identified coach qualities that cannot exist outside of an interpersonal relationship. In summary, the participants acknowledged the importance of the coaching relationship. This is in line with previous research emphasizing the importance of the therapeutic alliance in supported IPIs [10,16,17].

In contrast, fewer participants identified accountability as an important aspect of the program or coaching compared with other themes under *change process* domain. *Accountability* was identified in <1% of the *program helpful* codes, and in response to the *coach helpful* question, *accountability* was identified in <20% of the codes. However, because *accountability* has been found to be related to better adherence to IPIs [39], it is important to consider this finding more closely. Previous findings have shown that clinician contact during iCBT has been positively associated with adherence [40] and that adherence is positively associated with outcomes [41], suggesting that greater doses of the intervention improve outcomes.

Limitations

This study has several limitations that should be acknowledged. First, the 4 questions that were used to collect these data were not crafted with the intention of completing a qualitative analysis and were not designed for exploring the importance of the coach relationship in comparison with the program content. As a result, there is some ambiguity in the responses provided under the *program helpful* question. In addition, some of the emphases in the results on coaching and the relationship could be reflecting the specificity of one set of questions compared with the other. Second, participant responses were very brief in many cases, meaning that coders had to interpret meaning using familiarity with this sample and the program overall. Despite constantly challenging assumptions and biases, elements of bias may have certainly influenced the final coding and thematic structure. Finally, the final thematic structure does not represent the women who were noncompleters in the study or who did not provide follow-up data. It is possible that the nonresponders had very different experiences with the program and coaches that were not captured in this study.

Comparison With Prior Work

One model of support, Supportive Accountability [42], a manualized model developed to address the low adherence rates found for unsupported iCBT [41], posits that reciprocity exists in the participant-coach relationship, and a strong therapeutic bond should enhance the effects of accountability [39]; however, the mechanism of the association between clinician contact and adherence is unclear. Therapeutic alliance is one possibility; however, previous findings suggest that the therapeutic alliance and adherence may contribute separately to outcomes. For instance, Bur et al [10] found that alliance and adherence mediated the effect of guidance; however, alliance contributed to better outcomes separately from adherence. This is consistent with the results of this analysis, which suggests that the focus of the support person is more importantly placed on alliance-building factors than on accountability and adherence by most participants.

It should be noted that accountability and alliance are not mutually exclusive, and one should not be used at the expense of the other. Our findings suggest that being responsive to client needs in IPIs is likely as important as it is in FTF therapy [43], and support should be approached from a client-centered perspective. Although some researchers have found that standardized feedback was as effective as individualized guidance [10], participants in this study, by contrast, identified the importance of experiencing the program as responsive and named ways by which a supported IPI could meet individual needs. One of these is to understand the individual needs for accountability, as seemingly few participants are motivated by accountability alone. Other program-based responsiveness strategies identified by these results includes maximizing access options, representing diverse identities, including social components, and providing interactive components. Other support-based responsiveness strategies included translating tools into real life, deepening the conversations, and focusing on the unique dyadic relationship and unique needs of program users.

Inherent in the adherence literature is the idea that specific ingredients of the IPI or iCBT content are responsible for outcomes, and higher “doses” of these ingredients are associated with more positive outcomes. In contrast to this medical model of therapeutic change is the contextual model [15], which emphasizes therapeutic alliance. Within the IPI literature, the alliance has thus far been defined in terms of task, goal, and bond subscales given by Bordin [12,13]. The present results suggest that more recent work on therapeutic alliance may be a better fit in the case of IPIs. For example, in the tripartite model of the therapeutic relationship (ie, the real relationship, the working alliance, and the transference configuration) given by Gelso [44], the real relationship is defined as “the personal relationship between therapist and patient marked by the extent to which each is genuine with the other” [44] and perceives and experiences the other in (realistic) ways. The results of this study appear to reflect this definition, as participants identified coach genuineness as an important aspect of their experience and expressed the desire for the coaching interactions to be personal and deeper than is possible with standardized support or an emphasis on accountability to completing content.

Further refinement of the model given by Gelso [44] has resulted in the tripartite model of the real relationship composed of belongingness, empathy, and expectations [45]. Belongingness, in this framework, references “Attachment Theory” by Bowlby and Base [46] and as such is proposed to influence positive outcomes through the effects of a participant feeling connected to another person who cares about their well-being. This is consistent with the suggestion by Bur et al [10] that it is not the actual contact itself that increases alliance but that the participant knows that a real person will support them during treatment. Our data appear to align with this idea as codes emerged such as feeling connected, human connection, and feeling supported, even when total time in contact with the coach was relatively low. However, the preponderance of codes indicating relationship factors suggests that additional work will be needed to further define the role of attachment and relationship factors in supported IPIs. The results of this study, therefore, provide

future research directions, including more nuanced definitions of the alliance and the real relationship, that align more closely with the current literature.

Further refinement may also be necessary in terms of differentiating alliances based on the mode of contact. The meta-analysis by Kaiser et al [16] determined that there was no difference in outcomes based on the frequency of contact with the support person, mode of contact, or availability of self-help content. However, most of the studies included in the meta-analysis by Kaiser et al [16] used email as the mode of support. It may be that different forms of support elicit different reactions to the task, goal, and bond subscales of the WAI or that different alliance measures are needed for differently supported IPIs. The results of this study suggest that the bond subscale could emerge as more important for support provided synchronously by phone. In addition, other researchers have suggested that the relevance of the alliance for therapy outcomes could differ among different kinds of disorders and different client groups [17]. The emphasis on the relationship with the coach in this study could be a reflection of the specific population (ie, veteran women in the postpartum period) and the specific concern (ie, postpartum depression).

Conclusions

These findings provide a unique perspective regarding participant experiences in a supported iCBT program for postpartum depression. These data suggest that the most helpful coaches are those who are empathetic and validating, supportive and nonjudgmental, flexible, patient, and genuine. These qualities are highly relevant for trainees in health service delivery and research. Similarly, coaches might avoid emphasizing accountability alone unless the participant specifically identifies this as their primary need from support. Several participants commented that they wanted their coaching sessions to feel personal, to include conversation beyond simply what was happening with program content, to deepen and explore emotions, and to feel that the coach was genuinely interested in how they were feeling. As such, coaches might work to flexibly apply program content to the participant's ongoing experience and context while prioritizing responding with genuineness and empathy. Creators of manualized support models should consider how these results inform the training of their support personnel.

Future research may explore associations among qualities of coaches, amount of synchronous telephone contact, alliance ratings, and participant outcomes. Our findings demonstrate that the participants were capable of discerning beneficial coach qualities through an IPI program in which there was synchronous audio contact. Previous research has suggested that it is possible to form therapeutic alliance within 2 weeks of initiating an intervention and that early alliance associations with outcomes may indicate that guidance may not be necessary for the entire duration of the IPI [10]. This study's data suggest that the users of iCBT supported by phone may desire more contact with support personnel and therefore engagement throughout their use of the IPI. Given that most participants in this sample wanted more coaching, in terms of frequency, length, or duration, program designers will want to carefully consider the balance between participants' wishes, efficient resource allocation, and therapeutic benefit. These considerations may be specific to telephone-based support and will therefore need follow-up research to discern questions of timing and dose effects of telephone-based support.

These results call into question previous research that has indicated that the bond subscale of the model by Bordin [12,13] is less impactful to the alliance-outcome relationship than the task and goal subscales [10,18]. It may be that conceptualization of the alliance in IPIs needs to be further developed and that commonly used measures of the alliance, the WAI [30], need to be further studied for psychometric validity when used to assess alliance in IPIs.

Kaiser et al [16] call for attention to therapeutic alliance in the design of IPIs and specifically the elements that may promote alliance such as mode of communication, individualizing feedback, and elements that contribute to therapist credibility. Pihlaja et al [23], however, commented that detailed descriptions of therapist characteristics for iCBT were challenging and therefore beyond scientific scrutiny because of the range of probable predictors of alliance. The data of this study provide some information for how designers and alliance researchers might consider the support aspect of IPIs.

Although some have suggested that alliance-related factors may be less impactful in IPIs [22], the participant perspectives conveyed in these findings suggest otherwise and are consistent with more recent research. However, a large gap in the IPI literature is the provision of support from a theoretical perspective. The conceptualization of the real relationship within the tripartite model of the alliance by Gelso [44] and further expansion to the tripartite model of the real relationship by Budge and Wampold [45] may be fruitful directions for IPI researchers.

Important to consider in future research is the model from which coaches are trained and supervised, as this is likely to impact treatment outcomes, engagement, adherence, and attrition. Future research with these data is planned that will compare the association between relationship factors, engagement measures, and outcomes under accountability-focused and alliance-focused support models.

The literature on therapist responsiveness also suggests that strict standardization of support may be counterproductive. Swift et al [47] summarized that matching a treatment to client preferences, compared with not matching the treatment to the client preferences, increased success rates, led to more progress, and resulted in fewer dropouts. When a participant is engaging with the IPI content, an appropriate place to tailor treatment to the individual, via attending to preferences, is the support relationship. Our findings support therapist responsiveness to relationship factors as a particularly important component of treatment.

Although these participants engaged with their coaches for an average of 12 minutes per coaching session [35], much less than the typical FTF therapy, some of these findings may be alluding to the distinction between counseling therapy and supportive coaching. Future work is warranted to examine the extent to

which human contact is perceived as adjunctive to IPIs and vice versa and for whom it is adjunctive.

Perhaps the most important feature in the training context was that the coaches for this program were doctoral level trainees in clinical and counseling psychology. Future implementations of iCBT for postpartum depression and other IPIs may use trainees knowing that the therapeutic alliance and responsiveness variables identified as beneficial herein are trainable and regarded highly by participants. However, given the nature of the valued coach characteristics identified herein, properly trained peers may also be a viable and scalable option for future IPIs and research.

One strength of this study is the relatively large sample size (n=184), particularly in terms of typical qualitative studies. Several characteristics of the sample suggest that the results are generalizable to other supported iCBT interventions and supported IPIs. For example, participant demographics were diverse and, through nationwide recruitment, represented diverse areas of the country, including both rural and urban locations. In addition, previous work demonstrated that engagement and outcomes of this program were comparable with those of other IPIs [35].

Given that most codes were not specific to CBT ingredients or to the functioning of this specific program, but instead were intrapersonal and process comments, these results are likely generalizable to non-CBT IPIs. Furthermore, qualitative analysis, as a methodology, is nearly absent from the IPI literature, and as such, these data provide a unique perspective into potential mechanisms of change, which has been an elusive aspect of process and outcomes research [48].

Acknowledgments

This material is based upon the work supported in part by a contract from the Department of Veterans Affairs; Veterans Health Administration; Office of Rural Health; and Veterans Rural Health Resource Center, Iowa City (award 16235). The views expressed in this paper are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. The authors would like to thank Dr Leslie Locke for her expertise and oversight of our qualitative processes for this project. MomMoodBooster [49] was developed by Scott Stuart, MD (University of Iowa), and an international team in the United States (Brian G. Danaher, PhD; Milagra Tyler, MS; and John R. Seeley, PhD; Oregon Research Institute & Influents Innovations) and in Australia (Jeannette Milgrom, PhD; Alan W. Gemmill, PhD; and colleagues; Parent-Infant Research Institute [PIRI]). The authors especially thank Brian Danaher and Milagra Tyler for their work in adapting the program to fit the requirements of their Veterans Affairs trial and for providing support and training. The authors would also like to thank Brian Danaher for reviewing and editing this manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete thematic structure including definitions and participant quotes.
[DOCX File , 34 KB - mental_v10i1e46061_app1.docx]

Multimedia Appendix 2
Relative frequency of codes indicating amount of coaching preferences, relative contributions to change processes by program- and coach-related codes, relative representation of change processes attributed to coach, and relative representation of coach qualities.
[DOCX File , 223 KB - mental_v10i1e46061_app2.docx]

References

1. Titov N, Hadjistavropoulos HD, Nielssen O, Mohr DC, Andersson G, Dear BF. From research to practice: ten lessons in delivering digital mental health services. J Clin Med 2019 Aug 17;8(8):1239 [FREE Full text] [doi: 10.3390/jcm8081239] [Medline: 31426460]
2. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychol Med 2014 Nov;44(15):3137-3150. [doi: 10.1017/S0033291714000245] [Medline: 25065947]
3. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
4. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]

5. Grist R, Cavanagh K. Computerised cognitive behavioural therapy for common mental health disorders, what works, for whom under what circumstances? A systematic review and meta-analysis. J Contemp Psychother 2013 Sep 4;43(4):243-251 [FREE Full text] [doi: 10.1007/s10879-013-9243-y]
6. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
7. Mu TY, Li YH, Xu RX, Chen J, Wang YY, Shen CZ. Internet-based interventions for postpartum depression: a systematic review and meta-analysis. Nurs Open 2021 May;8(3):1125-1134 [FREE Full text] [doi: 10.1002/nop2.724] [Medline: 33373101]
8. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interventions 2014 Oct;1(4):205-215 [FREE Full text] [doi: 10.1016/j.invent.2014.08.003]
9. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
10. Bur OT, Bielinski LL, Krauss S, Häfliger A, Guggisberg J, Krieger T, et al. Working alliance and adherence mediate the effect of guidance in a web-based program for participants with mild to moderate depressive symptoms: a secondary mediation analysis. Internet Interv 2022 Nov 26;30:100593 [FREE Full text] [doi: 10.1016/j.invent.2022.100593] [Medline: 36471704]
11. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
12. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychol Psychother Theory Res Pract 1979;16(3):252-260 [FREE Full text] [doi: 10.1037/h0085885]
13. Bordin ES. Theory and research on the therapeutic working alliance: new directions. In: Horvath AO, Greenberg LS, editors. The Working Alliance: Theory, Research, and Practice. Hoboken, NJ: John Wiley & Sons; 1994:13-37.
14. Nienhuis JB, Owen J, Valentine JC, Winkeljohn Black S, Halford TC, Parazak SE, et al. Therapeutic alliance, empathy, and genuineness in individual adult psychotherapy: a meta-analytic review. Psychother Res 2018 Jul;28(4):593-605. [doi: 10.1080/10503307.2016.1204023] [Medline: 27389666]
15. Wampold BE, Imel ZE. The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work. 2nd edition. Abingdon, UK: Routledge; 2015.
16. Kaiser J, Hanschmidt F, Kersting A. The association between therapeutic alliance and outcome in internet-based psychological interventions: a meta-analysis. Comput Human Behav 2021 Jan;114:106512 [FREE Full text] [doi: 10.1016/j.chb.2020.106512]
17. Probst GH, Berger T, Flückiger C. The alliance-outcome relation in internet-based interventions for psychological disorders: a correlational meta-analysis. Verhaltenstherapie 2019 Oct 10;32(Suppl. 1):135-146 [FREE Full text] [doi: 10.1159/000503432]
18. Berger T. The therapeutic alliance in internet interventions: a narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]
19. Barazzone N, Cavanagh K, Richards DA. Computerized cognitive behavioural therapy and the therapeutic alliance: a qualitative enquiry. Br J Clin Psychol 2012 Nov;51(4):396-417. [doi: 10.1111/j.2044-8260.2012.02035.x] [Medline: 23078210]
20. Hadjistavropoulos HD, Pugh NE, Hesser H, Andersson G. Therapeutic alliance in internet-delivered cognitive behaviour therapy for depression or generalized anxiety. Clin Psychol Psychother 2017 Mar;24(2):451-461. [doi: 10.1002/cpp.2014] [Medline: 27060617]
21. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an internet intervention (Deprexis) on severe depression symptoms: randomized controlled trial. Internet Interv 2015 Mar;2(1):48-59 [FREE Full text] [doi: 10.1016/j.invent.2014.12.003]
22. Andersson G, Paxling B, Wiwe M, Vernmark K, Felix CB, Lundborg L, et al. Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. Behav Res Ther 2012 Sep;50(9):544-550. [doi: 10.1016/j.brat.2012.05.003] [Medline: 22728647]
23. Pihlaja S, Stenberg JH, Joutsenniemi K, Mehik H, Ritola V, Joffe G. Therapeutic alliance in guided internet therapy programs for depression and anxiety disorders - a systematic review. Internet Interv 2017 Nov 17;11:1-10 [FREE Full text] [doi: 10.1016/j.invent.2017.11.005] [Medline: 30135754]
24. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
25. Hilsenroth MJ, Cromer TD, Ackerman SJ. How to make practical use of therapeutic alliance research in your clinical work. In: Levy RA, Ablon JS, Kächele H, editors. Psychodynamic Psychotherapy Research: Evidence-Based Practice and Practice-Based Evidence. Totowa, NJ: Humana Press; 2012:361-380.

26. Paxling B, Lundgren S, Norman A, Almlöv J, Carlbring P, Cuijpers P, et al. Therapist behaviours in internet-delivered cognitive behaviour therapy: analyses of e-mail correspondence in the treatment of generalized anxiety disorder. Behav Cogn Psychother 2013 May;41(3):280-289. [doi: 10.1017/S1352465812000240] [Medline: 22717145]
27. Holländare F, Gustafsson SA, Berglind M, Grape F, Carlbring P, Andersson G, et al. Therapist behaviours in internet-based cognitive behaviour therapy (ICBT) for depressive symptoms. Internet Interv 2015 Nov 19;3:1-7 [FREE Full text] [doi: 10.1016/j.invent.2015.11.002] [Medline: 30135783]
28. Askjer S, Mathiasen K. The working alliance in blended versus face-to-face cognitive therapy for depression: a secondary analysis of a randomized controlled trial. Internet Interv 2021 May 24;25:100404 [FREE Full text] [doi: 10.1016/j.invent.2021.100404] [Medline: 34401364]
29. Lindner P, Olsson EL, Johnsson A, Dahlin M, Andersson G, Carlbring P. The impact of telephone versus e-mail therapist guidance on treatment outcomes, therapeutic alliance and treatment engagement in Internet-delivered CBT for depression: a randomised pilot trial. Internet Interv 2014 Oct;1(4):182-187 [FREE Full text] [doi: 10.1016/j.invent.2014.09.001]
30. Horvath AO, Greenberg LS. Development and validation of the working alliance inventory. J Couns Psychol 1989;36(2):223-233 [FREE Full text] [doi: 10.1037/0022-0167.36.2.223]
31. Gómez Penedo JM, Babl AM, Grosse Holtforth M, Hohagen F, Krieger T, Lutz W, et al. The association of therapeutic alliance with long-term outcome in a guided internet intervention for depression: secondary analysis from a randomized control trial. J Med Internet Res 2020 Mar 24;22(3):e15824 [FREE Full text] [doi: 10.2196/15824] [Medline: 32207689]
32. Danaher BG, Milgrom J, Seeley JR, Stuart S, Schembri C, Tyler MS, et al. MomMoodBooster web-based intervention for postpartum depression: feasibility trial results. J Med Internet Res 2013 Nov 04;15(11):e242 [FREE Full text] [doi: 10.2196/jmir.2876] [Medline: 24191345]
33. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. J Med Internet Res 2016 Mar 07;18(3):e54 [FREE Full text] [doi: 10.2196/jmir.4993] [Medline: 26952645]
34. Milgrom J, Danaher BG, Seeley JR, Holt CJ, Holt C, Ericksen J, et al. Internet and face-to-face cognitive behavioral therapy for postnatal depression compared with treatment as usual: randomized controlled trial of MumMoodBooster. J Med Internet Res 2021 Dec 08;23(12):e17185 [FREE Full text] [doi: 10.2196/17185] [Medline: 34889742]
35. Solness CL, Kroska EB, Holdefer PJ, O'Hara MW. Treating postpartum depression in rural veterans using internet delivered CBT: program evaluation of MomMoodBooster. J Behav Med 2021 Aug;44(4):454-466 [FREE Full text] [doi: 10.1007/s10865-020-00188-5] [Medline: 33052526]
36. Rennie DL. Qualitative research as methodical hermeneutics. Psychol Methods 2012 Sep;17(3):385-398. [doi: 10.1037/a0029250] [Medline: 22823104]
37. Merriam SB, Tisdell EJ. Qualitative Research: A Guide to Design and Implementation. Hoboken, NJ: John Wiley & Sons; 2015.
38. Williams EN, Hill CE. Establishing trustworthiness in consensual qualitative research studies. In: Hill CE, editor. Consensual Qualitative Research: A Practical Resource for Investigating Social Science Phenomena. Washington, DC: American Psychological Association; 2012:175-185.
39. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013 Aug 21;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
40. Hilvert-Bruce Z, Rossouw P, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468. [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]
41. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014 Jul 16;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
42. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
43. Norcross JC, Lambert MJ. Psychotherapy relationships that work: volume 1: evidence-based therapist contributions. In: Norcross JC, Wampold BE, editors. Psychotherapy Relationships that Work: Volume 2: Evidence-Based Therapist Responsiveness. 3rd edition. Oxford, UK: Oxford University Press; 2019:1-14.
44. Gelso C. A tripartite model of the therapeutic relationship: theory, research, and practice. Psychother Res 2014;24(2):117-131. [doi: 10.1080/10503307.2013.845920] [Medline: 24188031]
45. Budge SL, Wampold BE. The relationship: how it works. In: Gelo OC, Pritz A, Rieken B, editors. Psychotherapy Research: Foundations, Process, and Outcome. Cham, Switzerland: Springer; 2014:213-228.
46. Bowlby J, Base A. A Secure Base: Parent-Child Attachment and Healthy Human Development. New York, NY: Basic Books; 1988.

47. Swift J, Callahan J, Cooper M, Parkin S. Preferences. In: Norcross JC, Wampold BE, editors. Psychotherapy relationships that work: Evidence-based therapist responsiveness. Volume 2. 3rd edition. Oxford, UK: Oxford University Press; 2019:157-187.
48. Cuijpers P, Reijnders M, Huibers MJ. The role of common factors in psychotherapy outcomes. Annu Rev Clin Psychol 2019 May 07;15(1):207-231. [doi: 10.1146/annurev-clinpsy-050718-095424] [Medline: 30550721]
49. Helping women with perinatal depression. MomMoodBooster. URL: https://mommoodbooster.com/ [accessed 2023-07-19]

Abbreviations

CBT: cognitive behavioral therapy
FTF: face-to-face
iCBT: internet-delivered cognitive behavioral therapy
IPI: internet-delivered psychological intervention
MMB: MomMoodBooster
WAI: Working Alliance Inventory

Edited by J Torous; submitted 31.01.23; peer-reviewed by R Andriani, K Vernmark; comments to author 02.03.23; revised version received 28.05.23; accepted 29.06.23; published 15.08.23.

Please cite as:
Solness CL, Holdefer PJ, Hsu T, Thomas EBK, O'Hara MW
Relationship Factors in Internet-Delivered Psychological Interventions for Veterans Experiencing Postpartum Depression: Qualitative Analysis
JMIR Ment Health 2023;10:e46061
URL: https://mental.jmir.org/2023/1/e46061
doi:10.2196/46061
PMID:37581917

Short Paper

Identifying Rare Circumstances Preceding Female Firearm Suicides: Validating A Large Language Model Approach

Weipeng Zhou[1], BA; Laura C Prater[2,3], MPH, PhD; Evan V Goldstein[4], MPP, PhD; Stephen J Mooney[5], MS, PhD

[1]Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA, United States

[2]Department of Psychiatry and Behavioral Health, University of Washington, Seattle, WA, United States

[3]Harborview Medical Center, School of Medicine, University of Washington, Seattle, WA, United States

[4]Department of Population Health Sciences, University of Utah, Salt Lake City, UT, United States

[5]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, United States

Corresponding Author:
Stephen J Mooney, MS, PhD
Department of Epidemiology
School of Public Health
University of Washington
Hans Rosling Center for Population Health
3980 15th Ave NE
Seattle, WA, 98195
United States
Phone: 1 206 685 1643
Email: sjm2186@uw.edu

Abstract

Background: Firearm suicide has been more prevalent among males, but age-adjusted female firearm suicide rates increased by 20% from 2010 to 2020, outpacing the rate increase among males by about 8 percentage points, and female firearm suicide may have different contributing circumstances. In the United States, the National Violent Death Reporting System (NVDRS) is a comprehensive source of data on violent deaths and includes unstructured incident narrative reports from coroners or medical examiners and law enforcement. Conventional natural language processing approaches have been used to identify common circumstances preceding female firearm suicide deaths but failed to identify rarer circumstances due to insufficient training data.

Objective: This study aimed to leverage a large language model approach to identify infrequent circumstances preceding female firearm suicide in the unstructured coroners or medical examiners and law enforcement narrative reports available in the NVDRS.

Methods: We used the narrative reports of 1462 female firearm suicide decedents in the NVDRS from 2014 to 2018. The reports were written in English. We coded 9 infrequent circumstances preceding female firearm suicides. We experimented with predicting those circumstances by leveraging a large language model approach in a yes/no question-answer format. We measured the prediction accuracy with F_1-score (ranging from 0 to 1). F_1-score is the harmonic mean of precision (positive predictive value) and recall (true positive rate or sensitivity).

Results: Our large language model outperformed a conventional support vector machine–supervised machine learning approach by a wide margin. Compared to the support vector machine model, which had F_1-scores less than 0.2 for most infrequent circumstances, our large language model approach achieved an F_1-score of over 0.6 for 4 circumstances and 0.8 for 2 circumstances.

Conclusions: The use of a large language model approach shows promise. Researchers interested in using natural language processing to identify infrequent circumstances in narrative report data may benefit from large language models.

(JMIR Ment Health 2023;10:e49359) doi:10.2196/49359

KEYWORDS

female firearm suicide; large language model; document classification; suicide prevention; suicide; firearm suicide; machine learning; mental health for women; violent death; mental health; language models; women; female; depression; suicidal

Introduction

Suicide is a leading cause of death in the United States. Suicide risk factors include physical and mental health disorders, substance use disorders, prior exposure to violence, and having a firearm at home [1-4]. Firearm suicide has been more prevalent among men, but age-adjusted female firearm suicide rates have increased by 20% from 2010 to 2020, outpacing the rate increase among males by about 8 percentage points [5]. However, relatively few studies have focused specifically on female firearm suicide instead focusing on samples in which males are overrepresented (eg, military veterans) [6,7]. More data are needed to identify circumstances surrounding female firearm systems, and a primary source of these data derives from unprocessed narrative reports in the National Violent Death Reporting System (NVDRS).

In a previous study [1], we found that conventional natural language processing (NLP) algorithms could successfully identify some relatively common circumstances preceding female firearm suicide, using coroners' and medical examiners' (CMEs') and law enforcement (LE) narrative reports provided by the NVDRS. However, because reliably training a conventional NLP pipeline requires a sizeable training data set, the approach worked well only for the most common preceding circumstances.

Recently, large language models such as ChatGPT were found to perform well on tasks such as answering yes/no questions and document classification [8-10]. Large language models were developed on the basis of large corpora of data crawled from the web and can be used to solve machine learning tasks in a question-answer format. Moreover, these large language models do not rely on the task's data set size. In this study, we explored the value of a large language model approach by framing our coding task as a binary response for classification. Specifically, we tested a large language model approach to identify infrequent circumstances preceding female firearm suicide and compared this approach's performance to that of traditional machine learning models.

Methods

Overview

ChatGPT is the state-of-the-art large language model, but we could not use it directly in this study. Our data contain protected information; hence, we could not upload them directly to ChatGPT. Instead, we used open-source large language model alternatives. We ran these models locally to protect potentially sensitive information in the data. These models are developed similarly and are competitive in certain areas compared to ChatGPT. In a benchmark evaluation of large language models' ability in problem-solving [11], FLAN-T5 [12] and FLAN-UL2 [13] were found to be less accurate than ChatGPT for world knowledge understanding and programming but competitive in following complex instructions, comprehension and arithmetic, and causal reasoning. In preliminary studies, we experimented with multiple large language models (FLAN-T5 [12], FLAN-UL2 [13], and others) and found that FLAN-UL2 performed the best. We hence used FLAN-UL2 for our subsequent experiments. Developed by Google LLC, FLAN-UL2 is an open-source, 20 billion–parameter encoder-decoder model and is useful for zero-shot learning (ie, the model makes predictions directly without further training).

Data Sets

We used the NVDRS Restricted Access Database of female firearm suicides from 2014 to 2018 [1]. The data set contained unstructured CME and LE narrative reports describing the circumstances leading up to the suicide deaths of 1462 females. The reports were written in English. We manually coded 9 infrequent circumstances (ie, labels) preceding the firearm suicide deaths following the instructions specified by Goldstein et al [1]: sleep problems, abusive relationships, custody issues, sexual violence, isolation or loneliness, substance abuse, dementia, bullying, and caregiver issues. All infrequent labels occurred in <5% of the cases. We have provided details regarding the circumstance distribution in Table S1 in Multimedia Appendix 1. We split the data set into training and test sets with a 0.5:0.5 ratio.

Model Evaluation

A prompt is the input for the large language model and will guide it for generating outputs. For FLAN-UL2, we designed a prompt as a pair of a narrative report and a question. The narrative report is the text we input into a traditional machine learning model. The question varies depending on the circumstances we want to the model to code. For example, for the circumstance "bullying," "Answer the following yes/no question: was the decedent experiencing bullying in-person or online? Answer:" is the question. The model will yield an output of "Yes" if "bullying" is mentioned in the narrative report; if not, "No" will be the output. The question was adapted from the definition of each label developed through a previously reported manual review process with minimal changes [1]. A complete list of questions and definitions for each label is included in Table S2 in Multimedia Appendix 1. As a baseline, we used a series of conventional support vector machine (SVM) models [14] trained to identify each circumstance. FLAN-UL2 was only applied on the test set. SVM models were trained on the training set and applied on the test set. We repeated all experiments 5 times with resampling of the training and test sets. We reported the average F_1-score, which is the harmonic mean of the precision (positive predictive value) and recall (true positive rate or sensitivity). The F_1-score measures the model's accuracy considering the imbalance in the data set and ranges from 0 to 1.

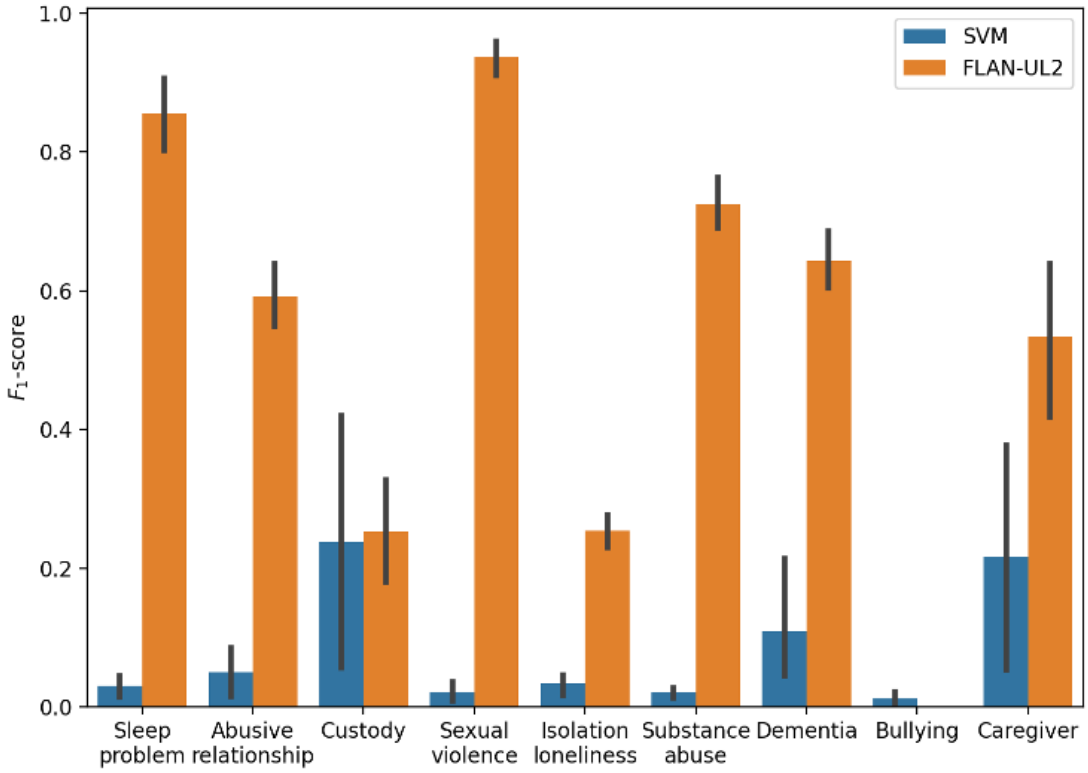
Ethical Considerations

This study did not require approval from the University of Washington institutional research board because these deidentified data on deceased persons were not considered human subjects research. We received ethical approval from the National Violent Death Reporting System Restricted Access Database review committee (request number #410).

Results

FLAN-UL2 performed better than the SVM for most of the female firearm suicide circumstances, sometimes substantially better (Figure 1). The F_1-score of FLAN-UL2 was greater than 0.8 for "sleep problem" (3.4% prevalence) and "sexual violence" (2.6% prevalence). "Bullying" (1.6% prevalence) was the only circumstance where the SVM outperformed FLAN-UL2, and F_1-scores were 0 or nearly 0 for all SVM and FLAN-UL2 runs. See Table S1 in Multimedia Appendix 1 for prevalence in other circumstances.

Figure 1. F_1-scores of the support vector machine (SVM) and FLAN-UL2 model for coding rare circumstances of female firearm suicide from suicide reports when averaged over 5 runs. The height of the bars represents the mean F_1-score, and the line at the tip of the bars represents the SD across 5 runs.



Discussion

Principal Findings

We found that the large language model (FLAN-UL2) outperformed the SVM for 8 of the 9 infrequent circumstances preceding female firearm suicide deaths. These findings suggest that a large language model approach can address a critical gap in identifying infrequent circumstances in unstructured text. Unlocking these data efficiently allows for subsequent analyses of female firearm suicide risk, including relationships among sexual violence, dementia, sleep problems, and firearm suicide.

The characterization of circumstances preceding female firearm suicides is an understudied area. In a previous study, Goldstein et al [1] used traditional NLP methods to predict 5 circumstances from suicide reports, with F_1-scores ranging from 0.6 to 0.8. However, all these circumstances had a prevalence of at least 15%. In our study, all 9 circumstances had a prevalence of less than 5%. We complemented the existing work by providing a method for automatically coding circumstances preceding female firearm suicides at a larger scope.

The failure in identifying the "bullying" circumstance may be due to the fact that bullying is one of the most infrequent circumstances preceding female firearm suicide in the NVDRS. The question we provided to the large language model, "was the decedent experiencing bullying in-person or online?" might not be sufficiently sensitive for the model to understand how to identify these circumstances in the narrative reports. More detailed explanation of "bullying," such as the victim was insulted or hurt at school or at the workplace, might be needed for the model to reason better. Large language models are known to be sensitive to prompt text, and designing an appropriate prompt (also known as prompt engineering) is an essential part of using large language models effectively [10]. Novel prompting techniques, such as few-shot learning (provide problem examples as part of the prompt) [10], have been proposed and may improve large language models' performance. In this study, we used simple and consistent prompts to provide a baseline for using large language models to code infrequent circumstances preceding female suicide. Large language models are also computationally expensive. The experiments in this study were carried out on 2 NVIDIA A100 40 GB graphics processing units. Large language models are also known to be sensitive to "hallucination" [15], meaning that they generate paragraphs of texts that look reasonable but are factually incorrect. In this study, we prompted the model to generate yes/no answers, bypassing the risks of hallucination.

Conclusions

Our large language model successfully identified infrequent circumstances preceding female firearm suicide deaths, having outperformed conventional NLP approaches by a wide margin. This finding suggests that large language models can be used to unlock textual analysis within public health research. More broadly, the success of our relatively simple queries at identifying infrequent circumstances suggests that large language models may be useful in public health surveillance, potentially allowing practitioners to track the prevalence of infrequent conditions that are never explicitly coded into surveillance systems. Future studies should explore the performance of different large language models and variations in the models' underlying mechanisms when applied to coding infrequent circumstances.

Acknowledgments

This work was funded by grants from the National Collaborative on Gun Violence Research (PI: Prater/ Goldstein, 2021) and the National Library of Medicine (4R00LM012868). We thank Dr Meliha Yetisgen for providing the computational resources.

Authors' Contributions

WZ conceptualized the study; carried out the formal analysis, investigation, and validation; designed the methodology; visualized the data; and drafted, edited, and reviewed the manuscript. EVG conceptualized the study, carried out the investigation, acquired funding, designed the methodology, supervised the study, and reviewed and edited the manuscript. LCP conceptualized the study, curated the data, acquired funding, designed the methodology, carried out the investigation, and edited and reviewed the manuscript. SJM conceptualized and supervised the study, acquired funding, designed the methodology, carried out the investigation, and edited and reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prevalence, definitions and derived questions of the rare female firearm suicide circumstances.
[DOCX File , 28 KB - mental_v10i1e49359_app1.docx]

References

1. Goldstein EV, Mooney SJ, Takagi-Stewart J, Agnew BF, Morgan ER, Haviland MJ, et al. Characterizing Female Firearm Suicide Circumstances: A Natural Language Processing and Machine Learning Approach. Am J Prev Med 2023 Aug;65(2):278-285. [doi: 10.1016/j.amepre.2023.01.030] [Medline: 36931986]
2. Anestis MD. Prior suicide attempts are less common in suicide decedents who died by firearms relative to those who died by other means. J Affect Disord 2016 Jan 01;189:106-109. [doi: 10.1016/j.jad.2015.09.007] [Medline: 26432034]
3. Miller M, Zhang Y, Prince L, Swanson SA, Wintemute GJ, Holsinger EE, et al. Suicide Deaths Among Women in California Living With Handgun Owners vs Those Living With Other Adults in Handgun-Free Homes, 2004-2016. JAMA Psychiatry 2022 Jun 01;79(6):582-588 [FREE Full text] [doi: 10.1001/jamapsychiatry.2022.0793] [Medline: 35476016]
4. Kellermann AL, Rivara FP, Somes G, Reay DT, Francisco J, Banton JG, et al. Suicide in the home in relation to gun ownership. N Engl J Med 1992 Aug 13;327(7):467-472. [doi: 10.1056/NEJM199208133270705] [Medline: 1308093]
5. WISQARS™ — Web-based Injury Statistics Query and Reporting System. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/injury/wisqars/index.html [accessed 2023-02-03]
6. Spark TL, Cogan CM, Monteith LL, Simonetti JA. Firearm Lethal Means Counseling Among Women: Clinical and Research Considerations and a Call to Action. Curr Treat Options Psychiatry 2022;9(3):301-311 [FREE Full text] [doi: 10.1007/s40501-022-00273-3] [Medline: 35791313]
7. Rowhani-Rahbar A, Simonetti JA, Rivara FP. Effectiveness of Interventions to Promote Safe Firearm Storage. Epidemiol Rev 2016;38(1):111-124. [doi: 10.1093/epirev/mxv006] [Medline: 26769724]
8. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
9. Xu F, Alon U, Neubig G, Hellendoorn V. A systematic evaluation of large language models of code. 2022 Presented at: MAPS '22: 6th ACM SIGPLAN International Symposium on Machine Programming; June 13, 2022; San Diego, CA. [doi: 10.1145/3520312.3534862]
10. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. arXiv Preprint posted online May 28, 2020. [FREE Full text] [doi: 10.5860/choice.189890]
11. Chia K, Hong P, Bing L, Poria S. INSTRUCTEVAL: towards holistic evaluation of instruction-tuned large language models. arXiv Preprint posted online June 7, 2023. [FREE Full text]

12. Chung HW, Hou L, Longpre S, Zoph B, Tay Y, Fedus W, et al. Scaling instruction-finetuned language models. arXiv Preprint posted online October 20, 2022. [FREE Full text]
13. Tay Y. A New Open Source Flan 20B with UL2. Yi Tay. URL: https://www.yitay.net/blog/flan-ul2-20b [accessed 2023-04-03]
14. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA: MIT Press; 2016.
15. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6) [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]

Abbreviations

CME: coroners or medical examiners
LE: law enforcement
NVDRS: National Violent Death Reporting System
NLP: natural language processing
SVM: support vector machine

Edited by J Torous; submitted 30.05.23; peer-reviewed by R Panirselvam, L Souza; comments to author 26.08.23; revised version received 31.08.23; accepted 02.09.23; published 17.10.23.

Please cite as:
Zhou W, Prater LC, Goldstein EV, Mooney SJ
Identifying Rare Circumstances Preceding Female Firearm Suicides: Validating A Large Language Model Approach
JMIR Ment Health 2023;10:e49359
URL: https://mental.jmir.org/2023/1/e49359
doi:10.2196/49359
PMID:37847549

Original Paper

Patient Health Questionnaire-9 Item Pairing Predictiveness for Prescreening Depressive Symptomatology: Machine Learning Analysis

Darragh Glavin[1,2], BEng; Eoin Martino Grua[1,2], PhD; Carina Akemi Nakamura[3], PhD; Marcia Scazufca[3,4], PhD; Edinilza Ribeiro dos Santos[5], PhD; Gloria H Y Wong[6,7], PhD; William Hollingworth[8], PhD; Tim J Peters[9], PhD; Ricardo Araya[10], PhD; Pepijn Van de Ven[1,2], PhD

[1]Department of Electronic and Computer Engineering, University of Limerick, Limerick, Ireland
[2]Health Research Institute, University of Limerick, Limerick, Ireland
[3]Departamento de Psiquiatria, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil
[4]Instituto de Psiquiatria, Hospital das Clinicas, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil
[5]Health Sciences School, University of the State of Amazonas, Manaus, Brazil
[6]Department of Social Work and Social Administration, University of Hong Kong, Pok Fu Lam, China (Hong Kong)
[7]Sau Po Centre on Ageing, University of Hong Kong, Pok Fu Lam, China (Hong Kong)
[8]Bristol Medical School, University of Bristol, Bristol, United Kingdom
[9]Bristol Dental School, University of Bristol, Bristol, United Kingdom
[10]Centre for Global Mental Health, Institute of Psychiatry, King's College London, London, United Kingdom

Corresponding Author:
Darragh Glavin, BEng
Department of Electronic and Computer Engineering
University of Limerick
Plassey Park Road
Limerick, V94 T9PX
Ireland
Phone: 353 87 260 7623
Email: Darragh.Glavin@ul.ie

Abstract

Background: *Anhedonia* and *depressed mood* are considered the cardinal symptoms of major depressive disorder. These are the first 2 items of the Patient Health Questionnaire (PHQ)–9 and comprise the ultrabrief PHQ-2 used for prescreening depressive symptomatology. The prescreening performance of alternative PHQ-9 item pairings is rarely compared with that of the PHQ-2.

Objective: This study aims to use machine learning (ML) with the PHQ-9 items to identify and validate the most predictive 2-item depressive symptomatology ultrabrief questionnaire and to test the generalizability of the best pairings found on the primary data set, with 6 external data sets from different populations to validate their use as prescreening instruments.

Methods: All 36 possible PHQ-9 item pairings (each yielding scores of 0-6) were investigated using ML-based methods with logistic regression models. Their performances were evaluated based on the classification of depressive symptomatology, defined as PHQ-9 scores ≥10. This gave each pairing an equal opportunity and avoided any bias in item pairing selection.

Results: The ML-based PHQ-9 items 2 and 4 (phq2&4), the *depressed mood* and *low-energy* item pairing, and PHQ-9 items 2 and 8 (phq2&8), the *depressed mood* and *psychomotor retardation or agitation* item pairing, were found to be the best on the primary data set training split. They generalized well on the primary data set test split with area under the curves (AUCs) of 0.954 and 0.946, respectively, compared with an AUC of 0.942 for the PHQ-2. The phq2&4 had a higher AUC than the PHQ-2 on all 6 external data sets, and the phq2&8 had a higher AUC than the PHQ-2 on 3 data sets. The phq2&4 had the highest Youden index (an unweighted average of sensitivity and specificity) on 2 external data sets, and the phq2&8 had the highest Youden index on another 2. The PHQ-2≥2 cutoff also had the highest Youden index on 2 external data sets, joint highest with the phq2&4 on 1, but its performance fluctuated the most. The PHQ-2≥3 cutoff had the highest Youden index on 1 external data set. The sensitivity and specificity achieved by the phq2&4 and phq2&8 were more evenly balanced than the PHQ-2≥2 and ≥3 cutoffs.

Conclusions: The PHQ-2 did not prove to be a more effective prescreening instrument when compared with other PHQ-9 item pairings. Evaluating all item pairings showed that, compared with alternative partner items, the *anhedonia* item underperformed alongside the *depressed mood* item. This suggests that the inclusion of *anhedonia* as a core symptom of depression and its presence in ultrabrief questionnaires may be incompatible with the empirical evidence. The use of the PHQ-2 to prescreen for depressive symptomatology could result in a greater number of misclassifications than alternative item pairings.

(JMIR Ment Health 2023;10:e48444) doi:10.2196/48444

KEYWORDS

Patient Health Questionnaire-2; PHQ-2; Patient Health Questionnaire-9; PHQ-9 items; depressive symptomatology; ultrabrief questionnaires; prescreening; machine learning; cardinal symptoms; low energy; psychomotor dysfunction; depressed mood

Introduction

Background

The Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-V) defines *anhedonia* (inability to find pleasure in what would usually be pleasurable activities) and *depressed mood* as the core criteria of major depressive disorder (MDD) [1]. According to the DSM-V, at least 1 of these cardinal symptoms must be present for a period of at least the past 2 weeks for a positive diagnosis of MDD, along with ≥5 symptoms. The heightened importance of these symptoms is also seen in screening questionnaires derived from the DSM-V's diagnostic criteria for MDD such as the depression module of the Primary Care Evaluation of Mental Disorders [2]; the depression module of the Patient Health Questionnaire (PHQ); the PHQ-9 [3]; and its ultrabrief version, PHQ-2 [4]. This importance stems from a consensus-based approach to the diagnosis of MDD formed from clinical experience [5].

The PHQ-9 has gained widespread popularity since its introduction and is now the most reliable and commonly used screening instrument in primary care and clinical research [6,7]. Each item of the PHQ-9 corresponds to a symptom of MDD: anhedonia, depressed mood, sleep disturbances, fatigue, appetite changes, feelings of worthlessness and excessive guilt, concentration difficulties, psychomotor disturbances, and suicidal ideation. The PHQ-9 assesses symptom frequency over the previous 2 weeks. This is more in line with the criteria required for an MDD diagnosis according to the DSM-V than the Primary Care Evaluation of Mental Disorders' 1-month assessment. Responses for the items are 0 (*Not at all*), 1 (*Several days*), 2 (*More than half the days*), and 3 (*Nearly every day*). The PHQ-9 was originally intended for use as a diagnostic algorithm and a severity-level measure. The PHQ-9 algorithm requires a minimum of 5 items to be scored as ≥2, with at least 1 of the first 2 items: (1) *a loss of interest or pleasure in doing things* and (2) *feeling down, depressed, or hopeless* endorsed as part of the 5 items. Although the algorithm structure closely matches the MDD criterion of the DSM-V, simply applying a ≥10 cutoff to the PHQ-9 severity scores has proven to be a reliable screening instrument for depression with performance equal to or better than the algorithm [3,6,7].

The PHQ-2 contains only the 2 core MDD symptom criteria, *anhedonia* (PHQ-9 item 1) and a *depressed mood* (PHQ-9 item 2 [phq2]). Its performance as a prescreening instrument for depressive symptomatology has been validated against other longer questionnaires [8]. However, in a diagnostic meta-analysis of 21 studies, the PHQ-2 has been shown to have lower accuracy in identifying MDD than initially reported [9]. The predictiveness of individual MDD symptoms has been previously compared [10]. However, symptoms are rarely evaluated when combined as pairings, and when they are, no pairing shows standout superiority [11]. As previously mentioned, the 2 items in the PHQ-2 were chosen as they are deemed the cardinal symptoms of depression, but objective evidence that these are indeed the 2 best items in a 2-item PHQ is lacking. Depressive symptomatology cases may go undetected if the cardinal symptoms are a suboptimal item pairing. Moreover, applying a cutoff to the summation of responses is a rather arbitrary approach, which is likely mainly used to date because of its simplicity. Machine learning (ML) offers a logical approach to tackle both issues. ML algorithms allow a structured, data-driven approach to item selection, thus allowing the selection of the most predictive 2 questions of the PHQ-9. ML algorithms are not limited to the restrictive summation and greater-than-or-equal-to logic used in the PHQ-2 but rather allow for the weighing of individual items as well as nonlinear transformations of the sum score. Consequently, this results in more thresholds to fine-tune instrument performance, leading to more refined classifications of depressive symptomatology.

Prior Work

To our knowledge, our previous analysis is the only one to date that has compared the performance of the PHQ-2 with alternative PHQ-9 item pairings and explored their combination with ML algorithms to predict depressive symptomatology [12]. ML algorithms have previously been combined with PHQ-9 items, but it was used to accurately predict suicidal ideation [13]. Our earlier work provided a data-driven ML analysis of all 36 possible PHQ-9 item pairings to predict depressive symptomatology [12]. Depressive symptomatology was defined as PHQ-9 scores ≥10 and was the reference standard used. This data-driven ML approach investigated the underlying relationship between symptoms and depressive symptomatology, without imposing any preconceptions on symptom importance.

Random oversampling of the screen positive class (ie, PHQ-9 scores ≥10) balanced the output classes to assist ML models in accurately predicting this less-frequent class. For direct comparison, the classification performance of all pairings was compared with the PHQ-2 on the same samples. The item pairings of PHQ-9 items 2 and 4 (phq2&4), the *depressed mood* and low-energy symptoms, and PHQ-9 items 2 and 8 (phq2&8), the *depressed mood* and *psychomotor retardation or agitation* symptoms, achieved the highest area under the receiver

operating characteristic (ROC) curves out of all possible PHQ-9 item pairings, including the PHQ-2 on both cross-validation (CV) and test data. This strong generalization performance achieved by the phq2&4 and phq2&8 on the out-of-sample test data indicated the potential use of these new pairings as prescreening instruments.

Objectives

The first objective was to re-evaluate the ML-based phq2&4 and phq2&8 on the primary data set with oversampling removed. Our previous analysis [12] used random oversampling to balance the output classes (PHQ-9<10 and PHQ-9≥10) and prevent the more frequent class from inflating performance scores, such as accuracy. However, as ROC curve analysis and area under the curve (AUC) are irrespective of class balance, there was no need to oversample; therefore, the analysis of the primary data set was rerun in this study. Next, the main objective of this work was to validate the new phq2&4 and phq2&8 pairings by investigating their generalization performance on 6 external data sets that were not used during the training of the ML models. These external data sets provided new samples with various demographics to test the pairing ability on out-of-sample data.

The length of the ultrabrief questionnaire was limited to 2 items because the performance of the PHQ-2, considered the gold standard ultrabrief version of the PHQ-9, served as the baseline. Maintaining a constant questionnaire length enabled the evaluation of alternative pairings compared with the established PHQ-2. To validate their ability as prescreening instruments, the phq2&4 and phq2&8 must generalize well and outperform the PHQ-2 on these external data sets. The fixed length also facilitated a comparison between a new ML methodology for screening and the traditional sum score psychometric approach used with the PHQ-2. Pairing performance was only evaluated for adult populations, aged ≥18 years, as an alternative questionnaire to the PHQ-9, the PHQ-Adolescents [14], exists for identifying depressive symptomatology in adolescents.

Methods

Data Sources

Overview

This analysis used 7 data sets (Table 1). The ML models were trained on the primary data set training split, and the best pairings were selected based on their CV performance on this split. The generalization performance of the best pairings was first estimated using a test split of the primary data set. To analyze this further, outside of this test split, 6 external data sets with various participant demographics were sourced. Four of the external data sets represented a wider Brazilian population (Pesquisa Nacional de Saúde 2013 [PNS2013], Pesquisa Nacional de Saúde 2019 [PNS2019], Amazonas, and São Paulo-Manaus). The other 2 represented different populations (Mexican Medical Students [MexMedStudents] and Jockey Club [JC] JoyAge). Two of the external data sets represented age demographics similar to the primary data set (São Paulo-Manaus and JC JoyAge). The performance on these 6 external data sets determined if the best pairings generalized well to wider adult populations and would be viable prescreening instruments with performance equal to or better than the PHQ-2.

Table 1. Overview of the characteristics of the 7 data sets presented.

Data set	Country	Year	Sample size, n	Age (y), mean (SD; range)	Sex (female), n (%)	PHQ-9[a] scores, mean (SD)	PHQ-9 scores ≥10, n (%)
PROACTIVE	Brazil	2019	4025	68.4 (6.5; 60-100)	2542 (63.2)	6.9 (7.0)	1216 (30.2)
PNS2013[b]	Brazil	2013	60,202	43.3 (16.7; 18-101)	34,282 (56.9)	2.8 (4.3)	5051 (8.4)
PNS2019[c]	Brazil	2019	88,531	47.2 (17.1; 18-107)	46,869 (52.9)	3.4 (4.7)	9252 (10.5)
Amazonas	Brazil	2013-2014	1631	40.2 (15.4; 20-94)	838 (51.4)	5.3 (5.4)	313 (19.2)
São Paulo-Manaus	Brazil	2010-2011	1377	72.5 (8.5; 60-104)	759 (55.1)	3.2 (4.1)	117 (8.5)
Mexican Medical Students	Mexico	2014	772	20.2 (1.8; 18-31)	399 (51.7)	7.2 (4.4)	192 (24.9)
Jockey Club JoyAge	Hong Kong	2018-2019	4221	77.3 (8.8; 60-118)	3274 (77.6)	6.7 (4.2)	809 (19.2)

[a]PHQ-9: Patient Health Questionnaire–9.

[b]PNS2013: Pesquisa Nacional de Saúde 2013.

[c]PNS2019: Pesquisa Nacional de Saúde 2019.

The data sets were preprocessed independently of one another. PHQ-9 item responses were either missing or within the expected 0-3 Likert scale range. Samples with >2 missing item responses were removed from the data sets. Any remaining missing responses were imputed with the respective item's mode because of the ordinal nature of the item responses.

Primary Data Set

The PROACTIVE study was a cluster randomized controlled trial conducted in socioeconomically deprived areas of Guarulhos, Brazil, where a psychosocial intervention was provided to older adults that aimed to reduce depressive symptoms [15,16]. Individuals registered with primary care clinics were randomly interviewed for recruitment into the randomized controlled trial and were screened for depressive symptomatology using the PHQ-9 in a dedicated application on an Android tablet (n=4034) [17]. The primary data set used for this analysis was the PROACTIVE screening data set, which was used to train, cross-validate, and initially test all ML models for different PHQ-9 item pairings. Nine individuals did not

complete the PHQ-9 questionnaire and were excluded (n=4025). PHQ-9 scores ≥10 were used to classify an individual as having depressive symptomatology and was the principal inclusion criterion for the trial. The data set split was 69.99% (2817/4025) for training and 30.01% (1208/4025) for testing. The training set was further split into 5 folds for CV (approximately n=563 each). In a single CV iteration, 4 of these folds were used to train the ML models and the fifth fold was used for validation. All 5 CV iterations used a different validation fold and subsequently 4 different training folds, until all 5 folds were used for validation. The average performance across the 5 validation folds provided an initial out-of-sample performance estimate.

External Data Sets

Brazilian National Health Survey (Pesquisa Nacional de Saúde)

The PHQ-9 was self-administered in 2 Brazilian national health surveys: one in 2013, PNS2013 (n=222,385), and another in 2019, PNS2019 (n=293,726) [18]. Both data sets provided an evaluation of the pairing generalizability on a broader Brazilian population and outside of a primary care setting. The screening interviews for the primary data set were also conducted in 2019, removing any potential temporal effect on the generalization performance between it and this PNS2019 data set. As this analysis focused on the generalization performance of pairings in adult populations, only PHQ-9 responses from those aged ≥18 years were used (PNS2013: n=145,580 and PNS2019: n=207,845). Individuals who responded to <7 of the PHQ-9 items were excluded (PNS2013: n=60,202 and PNS2019: n=88,531).

Amazonas

Adults registered with primary care clinics in Coari (n=805) and Tefe (n=826), 2 cities in the State of Amazon, Brazil, completed the PHQ-9 during interviews for a cross-sectional study of depressive symptomatology prevalence, defined as PHQ-9 scores ≥10, and depression care [19]. Individuals aged ≥20 years were randomly selected from a database of eligible participants (n=1631). Interviews were held at individuals' homes upon consenting to participation and were conducted between August 2013 and May 2014.

São Paulo-Manaus

This data set comprised PHQ-9 responses from older adults registered with primary care clinics and primary health care professionals (n=1380) in São Paulo (n=703) and Manaus (n=677), Brazil. The study aimed to investigate the public stigma surrounding depression in older adults. The PHQ-9 was administered via interview to assess depressive symptoms in this sample and how these affected stigmatization [20]. Older adults were defined as those aged ≥60 years, the same age demographic as the primary data set, which provided a constant factor within the generalization analysis of this data set. São Paulo city and Guarulhos are both in São Paulo state, another constant factor for a portion of this sample, and Manaus is from the Amazon state, similar to the Amazonas data set. Participants were randomly selected, stratified by age and sex. Three participants with no PHQ-9 responses were excluded (n=1377).

Mexican Medical Students

Medical students registered at a private Mexican university were randomly selected to self-administer a mental health survey, which contained the PHQ-9 among other questionnaires (n=1200) [21]. Of those selected, about two-thirds consented to participate; of these, 2 students aged <18 years were excluded from the analysis because it is concerned with generalization performance in adults, along with 2 adults without PHQ-9 responses (n=772).

JC JoyAge

Jockey Club Holistic Support Project for Elderly Mental Wellness (JC JoyAge) is a community-based mental health service for older adults in Hong Kong [22]. Adults aged ≥60 years completed the PHQ-9 to evaluate depressive symptoms during an assessment interview after referral to the project by peer supporters in the community (n=4267). A total of 46 individuals were excluded owing to missing >2 PHQ-9 item responses (n=4221). There was a higher proportion of mild depressive symptoms (PHQ-9 scores from 5 to 9) in this sample, potentially because of the referral aspect of the study.

Instruments

PHQ-2 Instrument

The PHQ-2 [4] is an ultrabrief questionnaire that contains the first 2 items of the PHQ-9: *little interest or pleasure in doing things* and *feeling down,depressed, or hopeless*. Responses are on the same scale as the parent PHQ-9, "Not at all" to "Nearly every day," but the total scores range from 0 to 6. The optimal PHQ-2 score cutoff for classifying depressive symptomatology has been debated. The original study suggested a cutoff of ≥3 for optimal performance [4]. This was validated as the optimal threshold in other studies [7,8], but some studies required a lower cutoff of ≥2 to maximize sensitivity and specificity [9].

ML-Based Pairings

In our previous analysis, the phq2&4 and phq2&8 were found to best classify individuals into screen positive and screen negative cases of depressive symptomatology [12]. The phq2&4 contains the second and fourth items of the PHQ-9, and the phq2&8 contains the second and eighth items. As both contain 2 PHQ-9 items, their total scores also range from 0 to 6. However, these ML-based instruments do not use greater-than-or-equal-to cutoffs to classify individuals in the screen negative and screen positive groups. Instead, they use some (generally nonlinear) function applied to the inputs to calculate an output. In the case of logistic regression (LR), this output is a linear combination of the inputs, which is then nonlinearly transformed to an output domain of 0 to 1 by the sigmoid function. This output can be interpreted as a probability score for class membership of the presented input. The classification performance of the model can be adjusted by tuning the probability threshold for which a certain input is deemed to belong to 1 of the classes. As the 2 items that formed the input pairings for these models can each take on 4 values, the input space of these models consisted of 4^2=16 different patterns. These 16 patterns in turn led to 16 different probability thresholds that could be chosen to obtain a desired model performance.

ML Analysis

The ML methodology applied to the primary data set in this study was based on the work done in our previous analysis [12]. As in the previous analysis, all 36 unique PHQ-9 item pairings were iterated. Each pairing trained an ML model for a complete performance comparison between all pairings. Random oversampling of the less-frequent screen positive class was removed from this analysis. Random oversampling is typically performed in ML analyses to balance the output classes to prevent strong performance on the more frequent class inflating the accuracy. As the accuracy of the item pairings was not the main metric evaluated in this analysis, there was no need to duplicate samples of the less-frequent class to balance the classes. Instead, ROC curve analysis and AUC, which are irrespective of class balance, investigated pairing performance for multiple thresholds. The primary data set was split into a training and a test set. A 5-fold CV was applied to the training set to obtain initial out-of-sample performance insights during the training process, without exposing the test set. Multiple ML algorithms were evaluated to optimize the classification of depressive symptomatology. The LR, decision tree, extreme gradient boosting, support vector classifier, and multilayer perceptron ML algorithms were evaluated.

All ML algorithms, except for decision trees achieved similar CV performance. To simplify the pipeline, LR models were chosen because (1) these are well-established models in both the statistics and ML domains, (2) ease of training, and (3) interpretability of model coefficients. To make classifications using an LR model, the probability of a given set of inputs (a pairing response combination) belonging to the positive class (PHQ-9≥10) is estimated. The inputs are linearly combined using the following multiple linear regression equation:

$$z = w_1.phq_i + w_2.phq_j + b \quad \textbf{(1)},$$

where w_1 and w_2 represent the weights that multiply the 2 PHQ items phq_i and phq_j, and b represents the bias term. Subsequently, the resulting sum of this equation (z) is nonlinearly transformed using the sigmoid function:

$$\hat{y} = 1 / 1+e^{-z} \quad \textbf{(2)},$$

which bounds the output ($\hat{y}$) to a value between 0 and 1. This output can be interpreted as a probability score for class membership for a given set of inputs. If the output probability is above a set threshold (≥0.5 is the default), a screen positive prediction is made.

Separate LR models were trained for each item pairing on the training data set. The optimal LR regularization hyperparameter value was selected for each model based on the CV AUC performance through Bayesian hyperparameter tuning [23]. Each LR model's weights and bias were learned from the data during the training process. The models learn the association between the 2 item responses and PHQ-9 scores ≥10 and how to best classify individuals as screen positive or screen negative. The weights and bias were updated throughout the training process, and the CV performance was evaluated by comparing the predictions with the real outputs. The phq2&4 and phq2&8 multiple linear regression equations and hyperparameters are provided in Multimedia Appendix 1. By withholding a proportion of the data for validation during CV, the model generalization on unseen data can be estimated. Without performing CV, models may overfit on the training data, that is, they may learn the specifics of the training data too well and may generalize poorly on out-of-sample data.

The performance of both the ≥2 and ≥3 cutoffs was reported for the PHQ-2, as these are the most commonly reported [9]. For the ML models, the threshold was chosen from the 16 candidates to maximize the Youden index (sensitivity + specificity − 1) based on CV performance on the training split of the primary data set. The maximized Youden index is a common selection criterion for an optimal threshold, as it weighs sensitivity and specificity equally [24]. Although it may not correspond to the optimal threshold in all prescreening contexts, it avoids any personal biases on the importance of sensitivity or specificity influencing the reported results.

Pairings were ranked based on their mean AUC score for the ROC across the 5 CV folds. The AUC scores measure the overall performance of a binary classifier irrespective of the chosen threshold. The predictive ability of the 2 best-performing PHQ-9 item pairings was compared with that of the PHQ-2. As the PHQ-2 does not need to be trained (ie, learn how best to classify individuals as screen negative or screen positive), its CV scores were calculated by manually assessing its performance on the same 5 CV folds on which the ML models were evaluated. The generalization performance of the best pairings and their optimal thresholds were then evaluated on the primary data set's test set and the 6 external data sets. These external data sets were used solely for testing, meaning the pairing ML models were not retrained and so did not learn new information from these external data sets. Their purpose was solely to investigate the performance of the new pairings as depressive symptomatology prescreening instruments and to compare it with that of the PHQ-2. The thresholds were not adjusted to suit the external data, and therefore, any data-driven optimal threshold reporting bias was removed [24].

Other metrics, such as positive predictive value (PPV) and negative predictive value (NPV), were reported, but the thresholds were not optimized for these. A complete performance report for all thresholds is provided in Multimedia Appendix 2. All the data preprocessing and ML models were coded using *Python 3.9* (Python Software Foundation). *Pandas 1.4.4* was used for data management. The entire ML pipeline was constructed using *Scikit-Learn 1.1.1*. The *BayesSearchCV* function in *Scikit-optimize 0.9.0* provided the algorithm evaluation and hyperparameter optimization. Graphs were plotted using *Matplotlib 3.5.2*.

Ethical Considerations

The PROACTIVE trial obtained written informed consent from participants before face-to-face interviews, and verbal consent was obtained for telephone interviews. The trial was approved by the Research Ethics Committee of the University of São Paulo Medical School (Comitê de Ética em Pesquisa Faculdade de Medicina da Universidade de São Paulo 2.836.569) and was authorized by the Guarulhos Health Secretary. The Brazilian National Health Ethics Research Committee of the Brazilian National Health Council approved both the PNS2013 and

PNS2019 surveys. All the participants signed an informed consent form. Anonymized versions of both the PNS2013 and PNS2019 surveys are publicly available for download and analysis. The University of São Paulo Medical School Ethical Committee approved the Amazonas study. The Health Secretariat of the municipalities of Coari and Tefe consented to the study, and written informed consent was obtained from the participants.

The São Paulo-Manaus study procedures were approved by the Research Ethics Committee of the Faculty of Medicine at the University of São Paulo, the Municipal Secretary of Health of São Paulo, the University of Amazonas State, and the Municipal Secretary of Health of Manaus. The participants provided written informed consent before data collection began. The MexMedStudents study complied with the ethical considerations stipulated in the Helsinki Treaties, Good Clinical Practices, and Ethics and Epidemiology: International Guidelines. The participants provided written informed consent for voluntary participation in the study and the use of their data. This data set is publicly available on the web. The JC JoyAge study received ethics approval from the Human Research Ethics Committee of the University of Hong Kong (reference EA1709021). All the participants provided informed consent.

None of the data sets used in this analysis contained personally identifiable information. None of the participants received financial compensation for their participation in their respective studies.

Results

Primary Data Set

Probability Thresholds

The input feature space for the phq2&4 (Figure 1) contains each combination of *depressed mood* (phq2) and *lack of energy* (PHQ-9 item 4; phq4) responses. The Likert scale 0 to 3 response options for phq2 are on the x-axis and those for phq4 are on the y-axis. Each response combination has an estimated probability of belonging to the screening-positive class. For example, if an individual responded 0 (“Not at all”) to both phq2 and phq4 (bottom left of the input feature space), the assigned probability of that individual being screen positive is 0.015 according to the phq2&4 ML model. These probabilities represent the 16 candidate probability thresholds of the phq2&4 instrument observed on the ROC curve (Figure 2).

Similar to the psychometric measure thresholds, a greater-than-or-equal-to threshold was applied to these probabilities. The performance of the model could be adjusted by varying the applied threshold. As the sensitivity and specificity can be determined from the ROC curve (Figure 2), a threshold that most closely matches the required sensitivity and specificity performance can be chosen. As reported in the *Methods* section, the probability thresholds applied to the item pairings were chosen to maximize the Youden index. The ≥0.322 probability threshold maximized the Youden index for the phq2&4. An individual will be classified as screen positive by the phq2&4 ML model if the assigned probability of their item response combination is ≥0.322; otherwise, they will be classified as screen negative. The red (screen positive) and blue (screen negative) areas represent the predictions of the phq2&4 ML model with the ≥0.322 probability threshold applied (Figure 1). The white line (the decision boundary) that separates the 2 colored areas represents this probability threshold. The input feature space and probability threshold ROC curve for the phq2&8 are provided in Multimedia Appendix 3.

Figure 1. Input feature space showing all the 16 possible item response combinations for the Patient Health Questionnaire–9 items 2 and 4 (phq2&4) instrument. phq2: Patient Health Questionnaire–9 item 2, phq4: Patient Health Questionnaire–9 item 4.

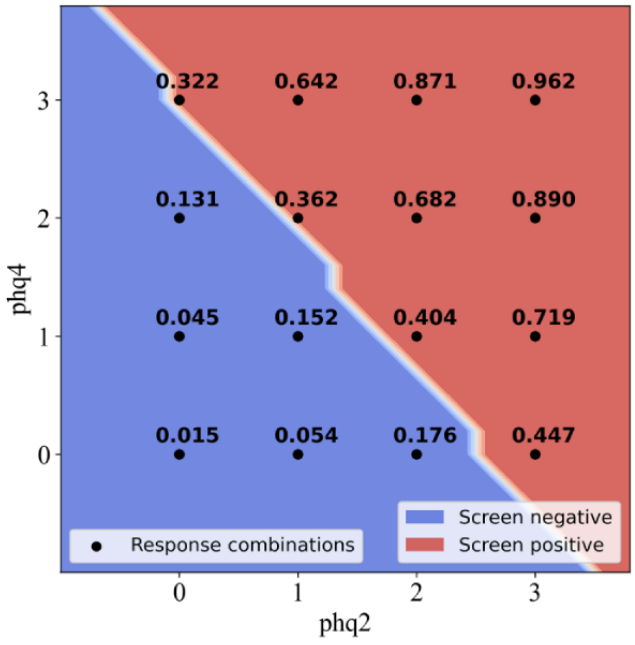
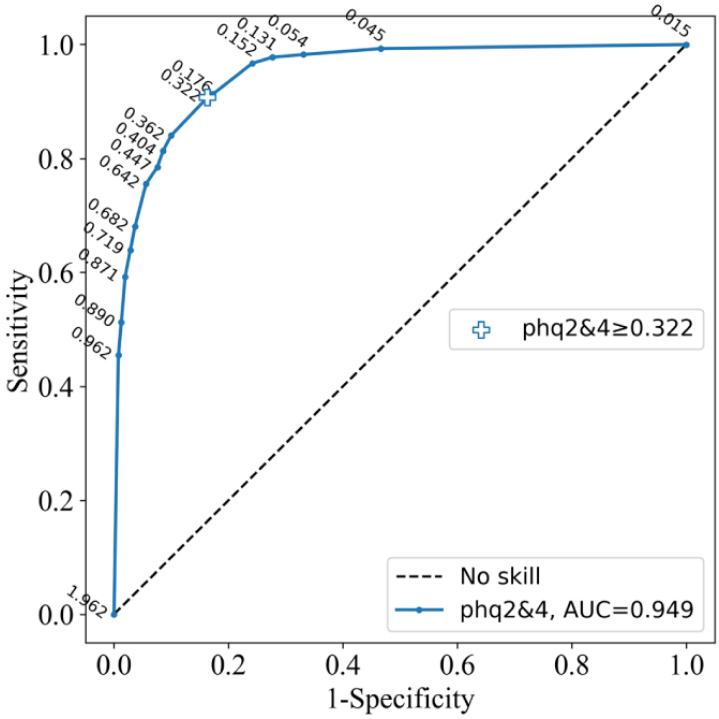


Figure 2. Probability thresholds on the Patient Health Questionnaire–9 items 2 and 4 (phq2&4) machine learning receiver operating characteristic curve on the PROACTIVE training set. Each threshold represents a decision boundary in the input feature space. AUC: area under the curve.



CV Performance

As found in our previous analysis [12], the phq2&4 and phq2&8 were the best performing, irrespective of removing the oversampling. The phq2&4 and phq2&8 had slightly higher AUCs than the PHQ-2 on CV data (0.949, 0.947, and 0.932, respectively; Figure 3). The reported results are for the maximized CV Youden index thresholds of the phq2&4 and phq2&8 along with the most common PHQ-2 cutoffs of ≥2 and ≥3 (Table 2). The phq2&4 threshold of ≥0.322 achieved a Youden index of 0.744, the highest out of all probability thresholds for the instrument. The PHQ-2≥2 achieved the same Youden index, meaning that their overall performance levels were equivalent despite different sensitivity and specificity statistics.

Figure 3. Receiver operating characteristic curves for the Patient Health Questionnaire–9 items 2 and 4 (phq2&4), Patient Health Questionnaire–9 items 2 and 8 (phq2&8), and Patient Health Questionnaire–2 (PHQ-2) instruments on the PROACTIVE training set. AUC: area under the curve.

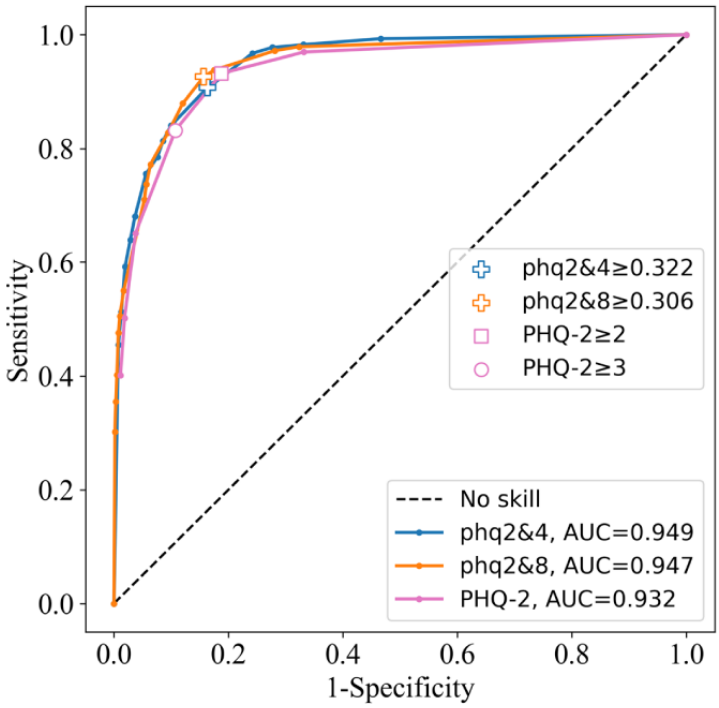


Table 2. Generalization results of Patient Health Questionnaire–2 (PHQ-2), Patient Health Questionnaire–9 items 2 and 4 (phq2&4), and Patient Health Questionnaire–9 items 2 and 8 (phq2&8) instruments on the PROACTIVE data set.

Data set and instrument	Youden index	Sensitivity	Specificity	Positive predictive value	Negative predictive value
PROACTIVE CV[a]					
PHQ-2≥2	0.744	0.932	0.812	0.682	0.965
PHQ-2≥3	0.725	0.832	0.893	0.770	0.925
phq2&4≥0.322	0.744	0.907	0.837	0.706	0.954
phq2&8≥0.306	0.769	0.926	0.843	0.719	0.963
PROACTIVE test					
PHQ-2≥2	0.753	0.934	0.819	0.690	0.966
PHQ-2≥3	0.739	0.849	0.890	0.769	0.932
phq2&4≥0.322	0.749	0.915	0.834	0.705	0.958
phq2&8≥0.306	0.735	0.907	0.828	0.695	0.954

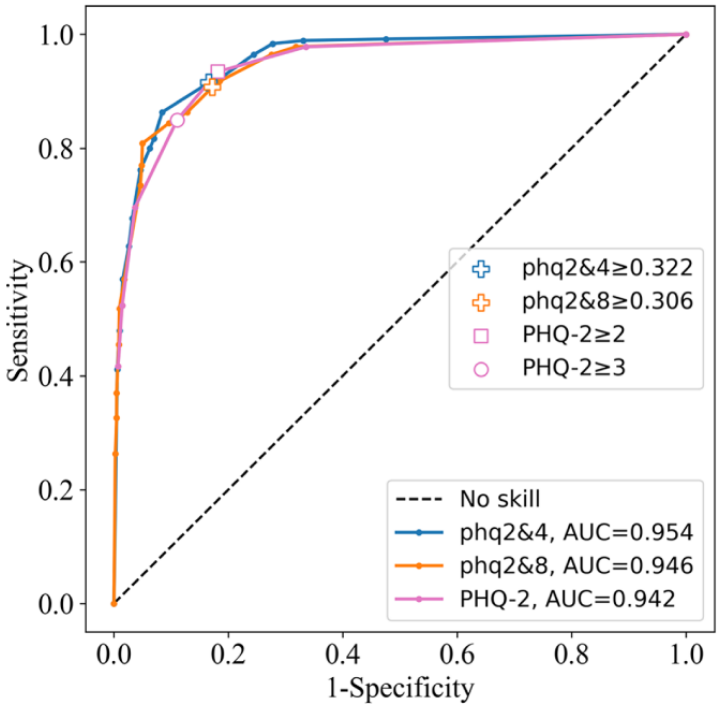
[a]CV: cross-validation.

The optimal threshold for the phq2&8 was ≥0.306 with a Youden index of 0.769, which was higher than the phq2&4 threshold and both PHQ-2 cutoffs. The PHQ-2≥3 cutoff had the lowest Youden index (0.725). As expected, the higher ≥3 cutoff on the PHQ-2 is less sensitive and more specific than the ≥2 cutoff. The PHQ-2≥3 cutoff had the highest PPV at 0.770, followed by phq2&8, phq2&4, the PHQ-2≥2 cutoff. The NPVs were higher than the PPVs for all instruments. At 0.965, the best NPV was from the PHQ-2 ≥2 cutoff, slightly better than the phq2&8, which was followed by the phq2&4 and PHQ-2 ≥3 cutoff with the lowest score (Table 2).

Test Performance

The new pairings again achieved slightly higher area under the ROC curves than the PHQ-2 on the primary data set's test split (Figure 4). The phq2&4 yielded 0.954, the phq2&8 yielded 0.946, and the PHQ-2 yielded 0.942. The PHQ-2≥2 cutoff achieved the highest Youden index of 0.753, closely followed by the phq2&4 with 0.749. The PHQ-2≥3 cutoff scored 0.739, whereas the phq2&8 achieved the lowest with 0.735, the largest decrease in CV performance. This lower Youden index for the phq2&8 was a result of a drop in both sensitivity and specificity. Its PPV and NPV were also lower. Metric scores for the phq2&4 and both PHQ-2 cutoffs were similar to those seen in CV (Table 2).

Figure 4. Receiver operating characteristic curves for the Patient Health Questionnaire–9 items 2 and 4 (phq2&4), Patient Health Questionnaire–9 items 2 and 8 (phq2&8), and Patient Health Questionnaire–2 (PHQ-2) instruments on the PROACTIVE test set. AUC: area under the curve.



External Data Sets

Performance Overview

The performance of the PHQ-2, phq2&4, and phq2&8 instruments was evaluated on 6 external data sets used to test the generalization of these as prescreening instruments (Table 3). The phq2&4 had the highest AUC on 4 of the 6 data sets: PNS2013, PNS2019, Amazonas, and JC JoyAge. The phq2&8 achieved the highest AUCs on the 2 other data sets, São Paulo-Manaus and MexMedStudents, with the phq2&4 slightly lower but higher than the PHQ-2. The phq2&4 threshold had the highest Youden index on the Amazonas external data set and the joint highest on the PNS2019 external data set. The phq2&8's threshold scored highest in terms of Youden index on the São Paulo-Manaus and MexMedStudents external data sets. The PHQ-2≥2 cutoff achieved the highest Youden index on the PNS2013 external data set and the joint highest on the PNS2019 external data set. The PHQ-2≥3 cutoff only had the highest score on the JC JoyAge data set, where the ≥2 cutoff was substantially lower. On the MexMedStudents data set, both of the PHQ-2 cutoffs had considerably lower Youden indexes than those of phq2&4 and phq2&8.

Table 3. Area under the receiver operating characteristic curve performance across all external data sets used to test the Patient Health Questionnaire–2 (PHQ-2), Patient Health Questionnaire–9 items 2 and 4 (phq2&4), and Patient Health Questionnaire–9 items 2 and 8 (phq2&8) generalization performances.

Instrument	PNS2013[a], AUC[b]	PNS2019[c], AUC	Amazonas, AUC	São Paulo-Manaus, AUC	Mexican Medical Students, AUC	Jockey Club JoyAge, AUC
PHQ-2	0.960	0.961	0.899	0.941	0.838	0.869
phq2&4	0.966	0.969	0.921	0.942	0.879	0.886
phq2&8	0.946	0.949	0.912	0.944	0.884	0.851

[a]PNS2013: Pesquisa Nacional de Saúde 2013.

[b]AUC: area under the curve.

[c]PNS2019: Pesquisa Nacional de Saúde 2019.

Brazilian National Health Survey (Pesquisa Nacional de Saúde)

The ROC curves and AUC performance of each of the instruments were similar across both the PNS2013 (Figure 5) and PNS2019 data sets (Figure 6). The phq2&4 achieved the highest AUC of 0.966 on the PNS2013 data set and 0.969 on the PNS2019 data set. The phq2&8 AUC performance was lower, at 0.946 on the PNS2013 data set and 0.949 on the PNS2019 data set. The PHQ-2 scored lower than phq2&4 but higher than phq2&8, with values of 0.961 and 0.960, respectively. On the PNS2013 data set, the PHQ-2≥2 cutoff outperformed the phq2&4 threshold for Youden index, at 0.813, compared with 0.800. The phq2&8 scored 0.769, and the PHQ-2≥3 cutoff had the lowest score of 0.749. The sensitivity and specificity statistics were high across each instrument's threshold, with none being overly sensitive or specific. For Youden indices on the PNS2019 data set, the phq2&4 and the PHQ-2≥2 cutoff achieved the same score of 0.808, indicating that they are equally optimal points on the ROC curve despite having different values of sensitivity and specificity (giving equal weight to false positives and false negatives). The phq2&8 achieved a Youden index of 0.772, and the PHQ-2≥3 cutoff scored the lowest with 0.756 (Table 4).

Figure 5. Receiver operating characteristic curves for the Patient Health Questionnaire–9 items 2 and 4 (phq2&4), Patient Health Questionnaire–9 items 2 and 8 (phq2&8), and Patient Health Questionnaire–2 (PHQ-2) instruments on the Pesquisa Nacional de Saúde 2013 data set. AUC: area under the curve.

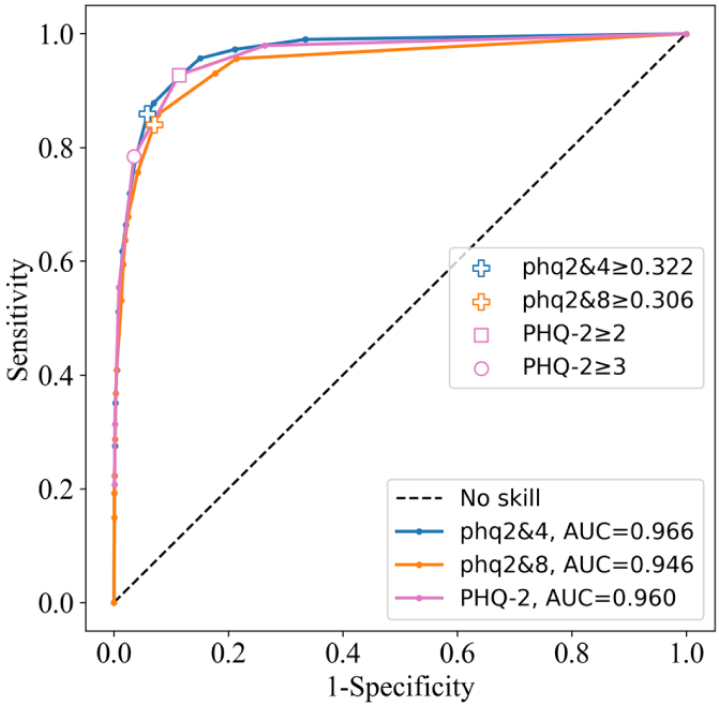


Figure 6. Receiver operating characteristic curves for the Patient Health Questionnaire–9 items 2 and 4 (phq2&4), Patient Health Questionnaire–9 items 2 and 8 (phq2&8), and Patient Health Questionnaire–2 (PHQ-2) instruments on the Pesquisa Nacional de Saúde 2019 data set. AUC: area under the curve.

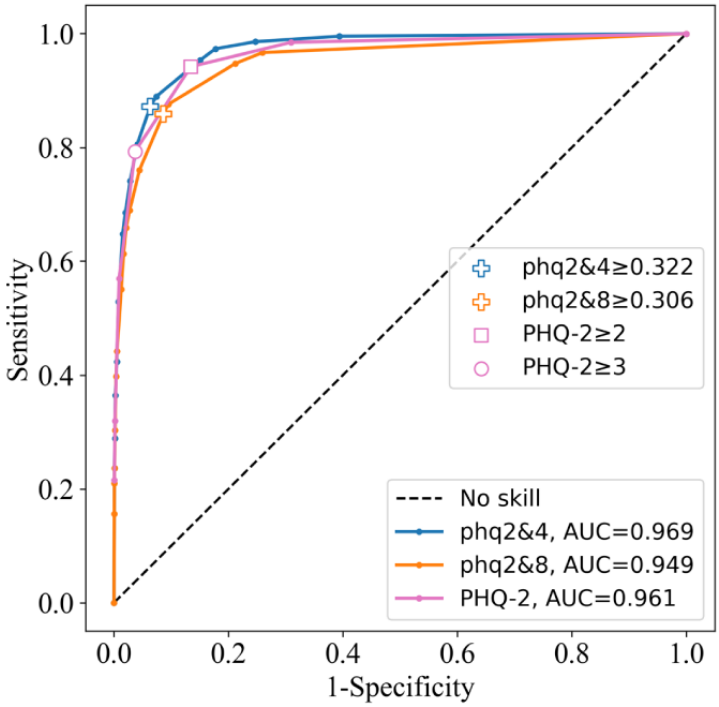


Table 4. Generalization results of Patient Health Questionnaire–2 (PHQ-2), Patient Health Questionnaire–9 items 2 and 4 (phq2&4), and Patient Health Questionnaire–9 items 2 and 8 (phq2&8) instruments on the 6 external data sets.

Data set and instrument	Youden index	Sensitivity	Specificity	Positive predictive value	Negative predictive value
PNS2013[a]					
PHQ-2≥2	0.813	0.927	0.886	0.428	0.993
PHQ-2≥3	0.749	0.784	0.965	0.674	0.980
phq2&4≥0.322	0.800	0.859	0.941	0.572	0.986
phq2&8≥0.306	0.769	0.840	0.930	0.522	0.984
PNS2019[b]					
PHQ-2≥2	0.808	0.942	0.866	0.450	0.992
PHQ-2≥3	0.756	0.793	0.963	0.716	0.976
phq2&4≥0.322	0.808	0.872	0.937	0.616	0.984
phq2&8≥0.306	0.772	0.859	0.914	0.537	0.982
Amazonas					
PHQ-2≥2	0.656	0.891	0.765	0.474	0.967
PHQ-2≥3	0.640	0.751	0.889	0.617	0.938
phq2&4≥0.322	0.684	0.834	0.850	0.569	0.956
phq2&8≥0.306	0.674	0.827	0.847	0.562	0.954
São Paulo-Manaus					
PHQ-2≥2	0.766	0.906	0.860	0.375	0.990
PHQ-2≥3	0.755	0.821	0.934	0.536	0.982
phq2&4≥0.322	0.719	0.829	0.890	0.411	0.982
phq2&8≥0.306	0.776	0.889	0.887	0.423	0.989
Mexican Medical Students					
PHQ-2≥2	0.463	0.906	0.557	0.403	0.947
PHQ-2≥3	0.492	0.557	0.935	0.738	0.865
phq2&4≥0.322	0.623	0.755	0.868	0.653	0.915
phq2&8≥0.306	0.649	0.781	0.868	0.661	0.923
Jockey Club JoyAge					
PHQ-2≥2	0.452	0.958	0.494	0.310	0.980
PHQ-2≥3	0.595	0.822	0.773	0.462	0.948
phq2&4≥0.322	0.590	0.915	0.675	0.400	0.971
phq2&8≥0.306	0.535	0.841	0.695	0.395	0.948

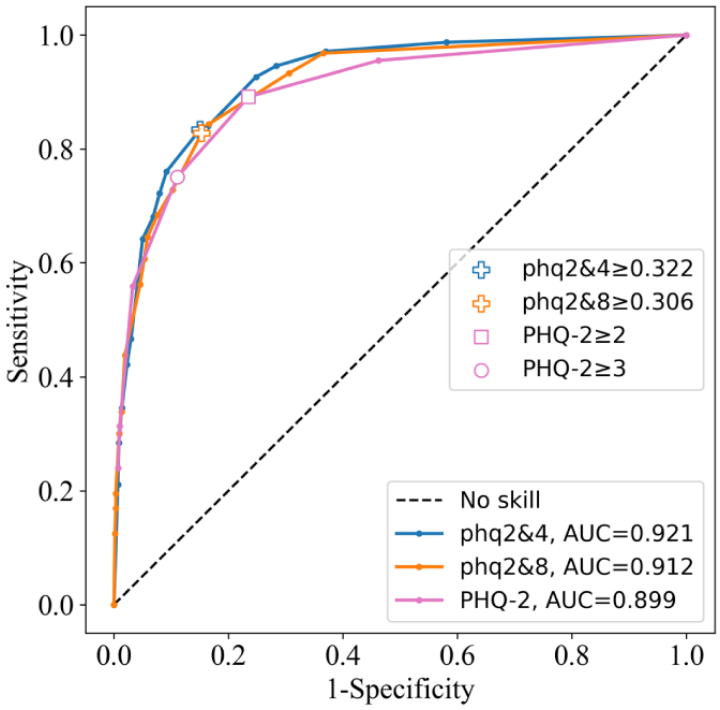
[a]PNS2013: Pesquisa Nacional de Saúde 2013.

[b]PNS2019 Pesquisa Nacional de Saúde 2019.

Amazonas

The phq2&4 and phq2&8 had the highest AUCs: 0.921 and 0.912, respectively. The phq2&4 threshold scored the highest for Youden index (0.684), and the phq2&8 threshold scored 0.674. The PHQ-2 achieved the lowest AUC of 0.899, and both the ≥2 and ≥3 cutoffs' Youden indexes were lower, at 0.656 and 0.640, respectively. Again, the sensitivity and specificity were high across the board for all the instruments. The ROC curves on the Amazonas data set (Figure 7) show superior ROC performance of the phq2&4 and phq2&8 compared with the PHQ-2, having higher AUC values and thresholds located closer to the optimal (top left) point of the graph.

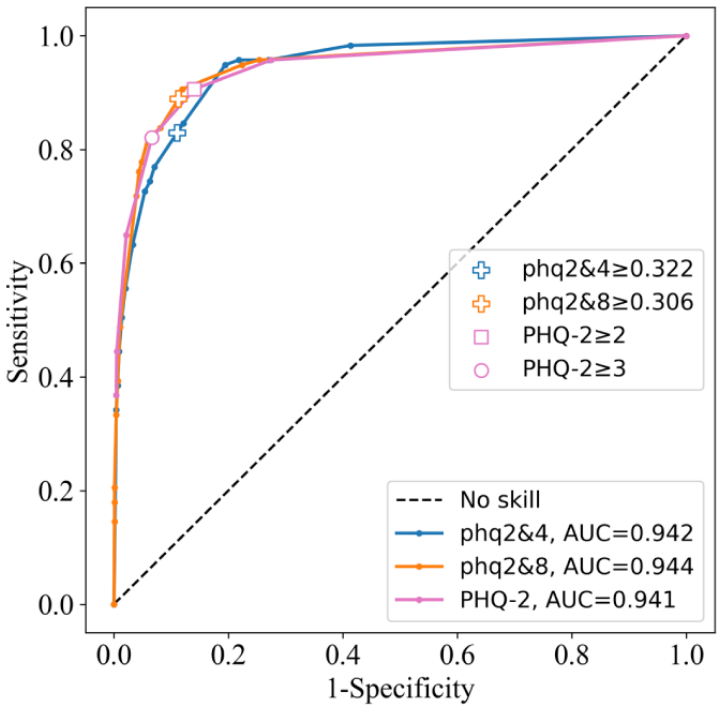
Figure 7. Receiver operating characteristic curves for the Patient Health Questionnaire–9 items 2 and 4 (phq2&4), Patient Health Questionnaire–9 items 2 and 8 (phq2&8), and Patient Health Questionnaire–2 (PHQ-2) instruments on the Amazonas data set. AUC: area under the curve.



São Paulo-Manaus

The phq2&8 was the best instrument in this sample, with an AUC of 0.944 and a Youden index of 0.776 for its threshold. The phq2&4 achieved a slightly lower AUC of 0.942, but its threshold was the poorest in this sample, with a Youden index of 0.719. This is evident when observing the phq2&4 threshold of ≥0.322, which falls below the ROC curves of the other instruments (Figure 8). The PHQ-2 AUC was marginally lower (0.941). The ≥2 cutoff outperformed the ≥3 cutoff with a Youden index of 0.766 compared with 0.755. Again, no threshold or cutoff was overly sensitive or specific in this sample (Table 4).

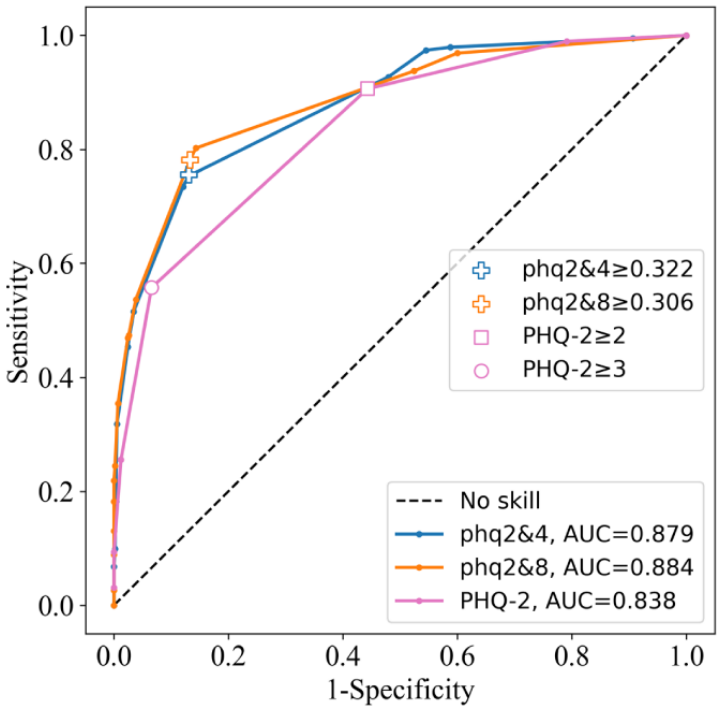
Figure 8. Receiver operating characteristic curves for the Patient Health Questionnaire–9 items 2 and 4 (phq2&4), Patient Health Questionnaire–9 items 2 and 8 (phq2&8), and Patient Health Questionnaire–2 (PHQ-2) instruments on the São Paulo-Manaus data set. AUC: area under the curve.



Mexican Medical Students

The performance gap between the new phq2&4 and phq2&8 instruments and the PHQ-2 was the largest in this data set. The ROC curves displayed this gap in terms of shape, AUC, and threshold and cutoff locations (Figure 9). The phq2&4 and phq2&8 instruments achieved higher AUC values of 0.879 and 0.884, respectively, compared with 0.838 for the PHQ-2. The phq2&4 ≥0.322 and phq2&8 ≥0.322 thresholds generalized well, with Youden indices of 0.623 and 0.649, respectively. Both PHQ-2 cutoffs performed poorly in this sample: the ≥2 cutoff was highly sensitive with low specificity, and ≥3 was highly specific with low sensitivity. This resulted in poor combined performance, as seen with Youden indices of 0.463 and 0.492 for the PHQ-2≥2 and PHQ-2≥3 cutoffs, respectively (Table 4). This was the first data set in which the ≥2 cutoff performed worse than the ≥3 cutoff.

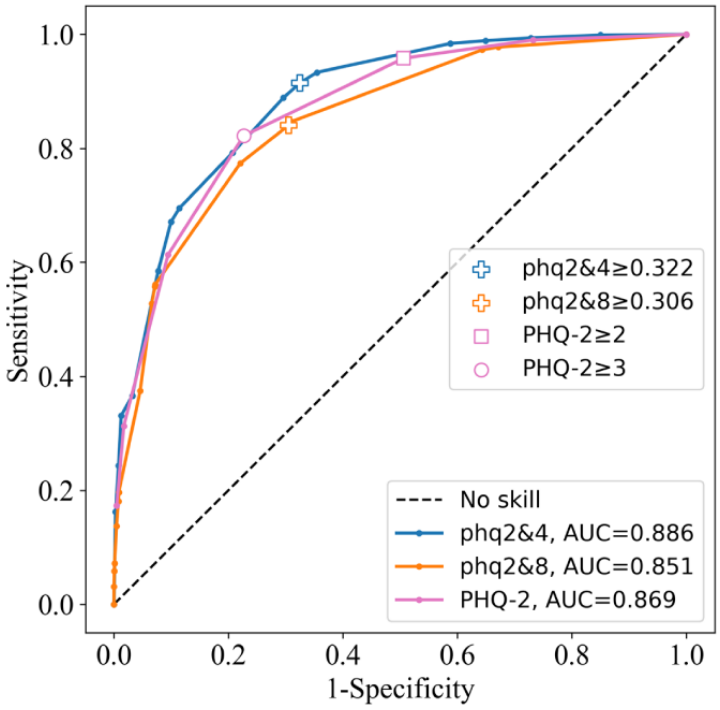
Figure 9. Receiver operating characteristic curves for the Patient Health Questionnaire–9 items 2 and 4 (phq2&4), Patient Health Questionnaire–9 items 2 and 8 (phq2&8), and Patient Health Questionnaire–2 (PHQ-2) instruments on the Mexican Medical Students data set. AUC: area under the curve.



JC JoyAge

The phq2&4 achieved the highest AUC in this sample (0.886). The performance of the phq2&8 dropped with an AUC of 0.851, whereas the PHQ-2 achieved 0.869 (Figure 10). The PHQ-2≥3 cutoff had the highest Youden index (0.595), again outperforming the ≥2 cutoff. The PHQ-2≥2 cutoff performance was highly sensitive but poorly specific, resulting in a low Youden index of 0.452. The phq2&4 and phq2&8 thresholds scored 0.590 and 0.535, respectively, for Youden index (Table 4).

Figure 10. Receiver operating characteristic curves for the Patient Health Questionnaire–9 items 2 and 4 (phq2&4), Patient Health Questionnaire–9 items 2 and 8 (phq2&8), and Patient Health Questionnaire–2 (PHQ-2) instruments on the Jockey Club JoyAge data set. AUC: area under the curve.



Discussion

Principal Findings

To avoid selective outcome reporting in threshold results, the optimal thresholds of the phq2&4 and phq2&8 instruments were based on the highest CV Youden index during the model training process. Only the optimal thresholds of the phq2&4 and phq2&8 instruments, ≥0.322 and ≥0.306, respectively, were reported, but both the common PHQ-2 cutoffs, ≥2 and ≥3, were reported. All cutoffs and threshold performance scores for the PHQ-2 psychometric measure method and phq2&4 and phq2&8 ML method are provided in Multimedia Appendix 2.

The phq2&4 instrument generalized best, having the highest AUC on the PROACTIVE test set and in 4 of the 6 external data sets (range 0.879-0.969). It achieved a higher AUC than the PHQ-2 on all data sets and was only outperformed by the phq2&8 on 2 of the external data sets. The phq2&4 threshold had the highest Youden index on the Amazonas data set and was joint highest on the PNS2019 data set with the PHQ-2. However, it was the most reliable across all sets, with the narrowest Youden index range (range 0.590-0.908). The phq2&4's performance waned most compared with the other instruments on the São Paulo-Manaus data set with a Youden index well below that of the phq2&8 and the 2 PHQ-2 cutoffs. Despite this, the phq2&4 AUC was marginally higher than that of the PHQ-2. The phq2&8 scored highest in terms of AUC and Youden index on the São Paulo-Manaus and MexMedStudents external data sets. Its AUC was lower than that of the PHQ-2 on the PNS2013, PNS2019, and JC JoyAge external data sets. However, overall, the phq2&8 fluctuated less on both AUC (range 0.851-0.949) and Youden index (range 0.535-0.776).

The PHQ-2 did not achieve the highest AUC in any of the data sets evaluated. Its performance also fluctuated more than the phq2&4 and phq2&8 on the external data sets (range 0.838-0.961). The PHQ-2's worst performance was on the MexMedStudents data set with a substantially lower AUC than the phq2&4 and phq2&8. Here, both the ≥2 and ≥3 cutoffs of the PHQ-2 also had considerably lower Youden indices than the phq2&4 and phq2&8. Both PHQ-2 cutoffs showed variable performance across the various external data sets. Their performance levels fluctuated more than the phq2&4 and phq2&8 thresholds seen with the broader range of Youden indexes from the ≥2 (range 0.452-0.813) and ≥3 (range 0.492-0.755) cutoffs.

The PHQ-2≥2 cutoff achieved the highest Youden index on 1 of the external data sets presented, PNS2013, and achieved the joint highest on PNS2019. Interestingly, the PHQ-2≥3 achieved the highest Youden index on the JC JoyAge data set, whereas the ≥2 cutoff scored substantially lower in this sample. The ≥3 cutoff also had a higher Youden index on the MexMedStudents data set. Arrieta et al [25] also found that ≥3 was the optimal cutoff in a Mexican cohort when using PHQ-9 scores ≥10 as the reference standard. This highlights the uncertainty around the optimal cutoff, which may be specific to certain populations and may be unknown before administering it as a prescreening instrument. The higher NPV than PPV seen from all instruments supports the view that ultrabrief questionnaires are better suited as "rule-out" instruments, where a prescreen negative strongly suggests the absence of depressive symptoms [9]. Nonetheless, the findings may differ depending on the probability threshold or the cutoff applied.

The weakest performance for all 3 instruments was observed on the MexMedStudents and JC JoyAge data sets, which were the only non-Brazilian data sets used in this analysis. The JC JoyAge data set represented a different cultural context in Hong Kong but had a similar age demographic to the Brazilian primary

data set, PROACTIVE. Conversely, the MexMedStudents data set consisted of younger and more educated individuals from Mexico. The comparatively lower performance of the phq2&4 and phq2&8 on these 2 non-Brazilian data sets, in contrast to the other Brazilian data sets, may be attributed to the fact that these pairings were identified within the Brazilian PROACTIVE data set. Despite the phq2&4's AUC being lower on these 2 data sets compared with the other 4 external data sets, it still outperformed the PHQ-2 on both data sets. Meanwhile, the phq2&8 performed better than the PHQ-2 on the MexMedStudents data set but worse on the JC JoyAge data set. The PHQ-2 was developed in a Western population, which may explain its lower AUC performance on these data sets collected in different populations. Nevertheless, these results highlight the PHQ-2's limited effectiveness in prescreening for depressive symptoms in the various cultures and demographics analyzed in this study.

The superior AUC performance of the phq2&4, albeit only slightly in some data sets, suggests that the *low-energy* item merits inclusion in ultrabrief questionnaires for the prescreening of depressive symptomatology. In terms of AUC, the phq2&8 did not generalize as well. However, its threshold achieved the highest Youden index on multiple data sets. This suggests that the *psychomotor disturbances* item also merits consideration for inclusion in prescreening instruments. Symptoms of fatigue and psychomotor dysfunction may have been overlooked despite evidence suggesting their increased importance in the diagnosis of MDD [26].

Limitations

A limitation of this study was the use of PHQ-9 scores of ≥10 as the reference standard and not a clinical diagnosis from an interview such as the Composite International Diagnostic Interview, Mini-International Neuropsychiatric Interview, or Structured Clinical Interview for DSM-V. Ultrabrief questionnaires for depressive symptoms are intended for use as part of a 2-stage screening process and should be followed by a more in-depth measure, such as the PHQ-9, following a prescreen positive result [4]. Therefore, it could be argued that the primary role of a depressive symptomatology prescreening instrument is to optimally identify individuals who will score ≥10 on the full PHQ-9, a common entry requirement for depression clinical trials, and not to directly predict depression. Nevertheless, before either the phq2&4 or phq2&8 are administered as prescreening instruments, their performance must be evaluated with an MDD clinical diagnosis used as the reference standard. Alternative optimal ultrabrief questionnaires may emerge when performance is evaluated against a clinical diagnosis.

Potential result biases could arise from the fact that 4 out of the 6 external data sets used in this study were from Brazil, the same country as the primary data set from which the optimal pairings were selected. However, it is important to note that these external data sets encompassed diverse populations within Brazil, including different states and age groups. The remaining 2 external data sets were collected in Mexico and Hong Kong, with the former consisting of younger medical school students and the latter consisting of individuals with an age profile similar to the PROACTIVE data set. The selection of the phq2&4 and phq2&8 was solely based on their performance on the primary data set. Their performance on 6 external data sets, without any model retraining, did not influence their selection as the best pairings. This minimized the instrument selection biases and enhanced the validity of the findings.

In a similar vein, another limitation of the study was that the data sets used were all from non-Western populations. This could also be considered a strength, as Western populations have been overrepresented in previous studies and meta-analyses [27]. Therefore, our findings increase the availability of studies from alternative populations. Further exploration of pairing performance in data sets from different cultures and where the PHQ-9 was administered in languages and dialects other than those evaluated here is required to validate the reliability of the newly found optimal pairings. It also remains essential to assess the generalizability of the newly found pairing in Western populations to evaluate pairing performance in the demographic in which the PHQ-2 was established. To examine reproducibility rather than generalizability, new ML models would need to be trained on each data set evaluated. This may lead to different pairing results, with unique optimal pairings found for each sample. This would suggest each population might require their own depressive symptomatology classification instrument, and a global optimal ultrabrief questionnaire may not exist. However, this approach can lead to overfitting when the instruments are too specialized and only suitable for the particular sample in which it was found.

The strong generalization performance of the phq2&4 and phq2&8 compared with the PHQ-2 on the 6 external data sets, despite being trained on 1 data set, indicated that they are more suitable as global ultrabrief questionnaires. Before administering either the phq2&4 or phq2&8 as prescreening instruments, further investigation is required into the impact item order has on the PHQ-9 and depressive symptomatology screening outcomes. The effect item order has on how a respondent endorses an item, on the neighboring items, and on the PHQ-9 sum scores is unknown. The use of either of these newly proposed prescreening instruments could affect item responses, and the outcome of the full PHQ-9, given items would be skipped. Future research involving a split test (A/B test) with different PHQ-9 item orders or with the PHQ-2 versus the phq2&4 for prescreening would help narrow this knowledge gap.

Conclusions

A re-examination of the cardinal symptoms of depression has been suggested to maintain the high standards required for good clinical practice [26]. Depression symptoms are rarely evaluated when combined as pairings, and when they are, performance levels are so similar that the selection of the best symptom pairing is considered somewhat arbitrary [11]. The objective of this study was to use a data-driven ML approach to identify and validate the most predictive 2-item depressive symptomatology ultrabrief questionnaire from the 9 items that comprise the PHQ-9. ML algorithms have been previously used to develop brief versions of parent questionnaires by identifying the most significant predictors [28].

In this analysis, the PHQ-2 has not demonstrated its superiority as a prescreening instrument when compared with other item pairings within the PHQ-9 item set. Comparing the performance of all 36 pairings gave an equal opportunity to each pairing and avoided any selection bias of the best 2-item ultrabrief questionnaire. Solely looking at individual performance when constructing ultrabrief questionnaires for depressive symptoms may lead to suboptimal performance as context on how items interrelate could be overlooked. Compared with the alternatives, the *anhedonia* item underperformed when paired with the *depressed mood* item. This suggests that it may be an arbitrary choice as a partner for the *depressed mood* item in prescreening instruments. *Anhedonia*, which involves a lack of pleasure rather than overt sadness, has been associated with higher levels of depression severity [29]. The inclusion of an *anhedonia* item in ultrabrief screening questionnaires, instead of potentially superior alternatives, may result in failure to identify individuals with moderate levels of depression severity.

The idea that cognitive symptoms play a more significant role in the diagnosis of depression is primarily a concept of the 20th-century Western world and may not hold true in other cultures to the same extent [29]. Instead, pairing the *depressed mood* item with the *low-energy* item (phq2&4) or the *psychomotor agitation or retardation* item (phq2&8) achieved higher AUC statistics on the primary data set. The phq2&4 and phq2&8 pairings include 1 cognitive symptom in *depressed mood*, 1 of the so-called cardinal symptoms of depression, alongside a neurovegetative symptom *in low energy* or *psychomotor changes*. This combination may be more beneficial in detecting the broad range of symptom profiles instead of asking 2 cognitive symptoms.

To further delve into these new pairings, their performance was investigated on 6 external data sets from 6 separate studies and with different participant demographics. These external data sets were used purely as test sets, meaning the ML models were not retrained and so did not learn any new information on how to best classify screen positive and screen negative samples in these data sets. Therefore, this analysis tested the generalizability of the phq2&4 and phq2&8 rather than their reproducibility across multiple data sets. Prevalence was not controlled for in this analysis, which enabled a more complete evaluation of the phq2&4 and phq2&8 generalization performance. This determined whether these new pairings and their respective ML models trained on an older adult, socioeconomically deprived Brazilian population generalized well and could be used as depressive symptomatology prescreening instruments for clinical trial recruitment in different demographics in place of the PHQ-2.

Combining item pairings with ML models allowed for a more flexible approach to the classification of depressive symptomatology. The summation and greater-than-or-equal-to logic of psychometric measures such as the PHQ-2 place equal weighting on each item. This approach has been questioned both theoretically and empirically [30]. ML models are not limited to this summation and greater-than-or-equal-to logic. This is possible because the LR coefficients allow for different weighting of items depending on their learned importance in the classification of depressive symptomatology. Hence, ML models have more thresholds to choose from (16 compared with 7 of the PHQ-2) to fine-tune performance until a desirable classification threshold is achieved.

Previously, the use of LR in conjunction with a screening questionnaire was considered too complex, as the presence of depressive symptomatology could not be identified quickly or by hand owing to the different weighting of the items [31]. However, given the drastic improvement in technology and the increased use of technical devices to collect data in clinical trials [17], the use of LR for screening is more viable nowadays. ML models can be deployed within screening applications where interviewers input patient responses and receive predictions on the presence of depressive symptomatology (prescreen positive or negative).

During this analysis, it was assumed that the optimal thresholds of the phq2&4 and phq2&8 should be determined using the maximum Youden indices. This may not be a suitable threshold for all prescreening contexts, especially in applications where instruments may be required to be more sensitive or more specific [32]. However, as the most suitable threshold is situational, both sensitivity and specificity were weighted equally using the Youden index. The optimal pairings were ranked on AUC, which measures the performance across all thresholds of an instrument, rather than the performance of the optimal threshold. A higher AUC indicates a minimal performance impact when adjusting the instrument's threshold. Having a greater number of thresholds to tune classification performance is another advantage that combining the PHQ-9 item pairings with ML models has over the summation and greater-than-or-equal-to logic of the psychometric approach.

An instrument's threshold or cutoff must be selected before its use. With the PHQ-2's most common cutoffs showing fluctuating performance, its use as a prescreening instrument could result in a larger number of misclassifications if a suboptimal threshold was chosen. The consistency of the findings for the phq2&4 and phq2&8 across multiple external data sets suggests that the strong performance seen is not by chance. These new symptom pairings warrant further investigation into how well they perform as prescreening instruments for depressive symptomatology in various populations.

Acknowledgments

This publication has emanated from research supported in part by a grant from the Science Foundation Ireland under grant 18/CRT/6049 (DG). The authors gratefully acknowledge funding from the Sao Paulo Research Foundation (Fundação de Amparo à Pesquisa do Estado de São Paulo; FAPESP) 2018/19343-9 and FAPESP 2022/05107-7 (CAN) and the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico;

CNPq-Brazil) 307579/2019-0 (MS). The PROACTIVE study was funded by the FAPESP (2017/50094-2) and the Joint Global Health Trials initiative jointly funded by the Department of Health and Social Care; the Foreign, Commonwealth & Development Office; the Medical Research Council (MRC); and Wellcome Trust (MRC MR/R006229/1). The São Paulo-Manaus study was funded by the CNPq-Brazil (472902/2011-3). The Jockey Club JoyAge study was funded by the Hong Kong Jockey Club Charities Trust. For the purpose of open access, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hyperparameters and multiple linear regression equations of the Patient Health Questionnaire–9 items 2 and 4 and Patient Health Questionnaire–9 items 2 and 8 logistic regression models.
[DOC File , 63 KB - mental_v10i1e48444_app1.doc]

Multimedia Appendix 2
Tables containing all threshold scores for the Patient Health Questionnaire–9 items 2 and 4, Patient Health Questionnaire–9 items 2 and 8, and Patient Health Questionnaire–2 on each data set.
[DOC File , 476 KB - mental_v10i1e48444_app2.doc]

Multimedia Appendix 3
The input feature space and probability threshold receiver operating characteristic curve for the Patient Health Questionnaire–9 items 2 and 8.
[DOC File , 386 KB - mental_v10i1e48444_app3.doc]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Washington, DC: American psychiatric association; 2013.
2. Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV3, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. JAMA 1994 Dec 14;272(22):1749-1756. [Medline: 7966923]
3. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
4. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
5. Lux V, Kendler KS. Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. Psychol Med 2010 Oct;40(10):1679-1690 [FREE Full text] [doi: 10.1017/S0033291709992157] [Medline: 20059797]
6. Levis B, Benedetti A, Thombs BD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019 Apr 12;365:l1781 [FREE Full text] [doi: 10.1136/bmj.l1781] [Medline: 30979729]
7. Mitchell AJ, Yadegarfar M, Gill J, Stubbs B. Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: a diagnostic meta-analysis of 40 studies. BJPsych Open 2016 Mar 09;2(2):127-138 [FREE Full text] [doi: 10.1192/bjpo.bp.115.001685] [Medline: 27703765]
8. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Feb;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
9. Manea L, Gilbody S, Hewitt C, North A, Plummer F, Richardson R, et al. Identifying depression with the PHQ-2: a diagnostic meta-analysis. J Affect Disord 2016 Oct;203:382-395. [doi: 10.1016/j.jad.2016.06.003] [Medline: 27371907]
10. Mitchell AJ, McGlinchey JB, Young D, Chelminski I, Zimmerman M. Accuracy of specific symptoms in the diagnosis of major depressive disorder in psychiatric out-patients: data from the MIDAS project. Psychol Med 2009 Jul;39(7):1107-1116. [doi: 10.1017/S0033291708004674] [Medline: 19000337]
11. Henkel V, Mergl R, Coyne JC, Kohnen R, Möller HJ, Hegerl U. Screening for depression in primary care: will one or two items suffice? Eur Arch Psychiatry Clin Neurosci 2004 Aug;254(4):215-223. [doi: 10.1007/s00406-004-0476-3] [Medline: 15309389]
12. Glavin D, Maekawa E, Grua EM, Nakamura CA, Scazufca M, Araya R, et al. Other PHQ-9 item pairings are better than the PHQ-2: A machine learning analysis. Procedia Computer Science 2022;206:101-110 [FREE Full text] [doi: 10.1016/j.procs.2022.09.089]
13. Kim S, Lee HK, Lee K. Which PHQ-9 items can effectively screen for suicide? Machine learning approaches. Int J Environ Res Public Health 2021 Mar 24;18(7):3339 [FREE Full text] [doi: 10.3390/ijerph18073339] [Medline: 33804879]

14. Johnson JG, Harris ES, Spitzer RL, Williams JB. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolesc Health 2002 Mar;30(3):196-204. [doi: 10.1016/s1054-139x(01)00333-0] [Medline: 11869927]
15. Scazufca M, Nakamura CA, Peters TJ, Henrique MG, Seabra A, La Rotta EG, et al. A collaborative care psychosocial intervention to improve late life depression in socioeconomically deprived areas of Guarulhos, Brazil: the PROACTIVE cluster randomised controlled trial protocol. Trials 2020 Nov 05;21(1):914 [FREE Full text] [doi: 10.1186/s13063-020-04826-w] [Medline: 33153482]
16. Scazufca M, Nakamura CA, Seward N, Moreno-Agostino D, van de Ven P, Hollingworth W, et al. A task-shared, collaborative care psychosocial intervention for improving depressive symptomatology among older adults in a socioeconomically deprived area of Brazil (PROACTIVE): a pragmatic, two-arm, parallel-group, cluster-randomised controlled trial. Lancet Healthy Longev 2022 Oct;3(10):e690-e702 [FREE Full text] [doi: 10.1016/S2666-7568(22)00194-5] [Medline: 36202129]
17. van de Ven P, Araya R, P de Paula Couto MC, Henrique MG, Meere D, Vilela Mendes A, et al. Investigating software requirements for systems supporting task-shifted interventions: usability study. J Med Internet Res 2019 Nov 12;21(11):e11346 [FREE Full text] [doi: 10.2196/11346] [Medline: 31714246]
18. Lopes CD, Gomes NL, Junger WL, Menezes PR. Trend in the prevalence of depressive symptoms in Brazil: results from the Brazilian National Health Survey 2013 and 2019. Cad Saude Publica 2022 May 06;38Suppl 1(Suppl 1):e00123421 [FREE Full text] [doi: 10.1590/0102-311X00123421] [Medline: 35544918]
19. Ribeiro Dos Santos E, Huang H, Menezes PR, Scazufca M. Prevalence of depression and depression care for populations registered in primary care in two remote cities in the Brazilian Amazon. PLoS One 2016 Mar 01;11(3):e0150046 [FREE Full text] [doi: 10.1371/journal.pone.0150046] [Medline: 26930353]
20. Scazufca M, P de Paula Couto MC, Huang H, Kester R, Braga PE, Peluso É, et al. Public stigma towards older adults with depression: findings from the São Paulo-Manaus elderly in primary care study. PLoS One 2016 Jun 28;11(6):e0157719 [FREE Full text] [doi: 10.1371/journal.pone.0157719] [Medline: 27352293]
21. López TJ, García D. Mental health data of 776 Mexican medical students (PHQ-9, GAD-7 and Epworth Sleepiness Scale Scores). Dataset. 2018. URL: https://figshare.com/articles/dataset/PHQ_9_GAD_7_and_Epworth_Sleepiness_Scale_Scores_of_783_Mexican_Medical_Students/1305817 [accessed 2022-06-02]
22. Liu T, Leung DK, Lu S, Kwok WW, Sze LC, Tse SS, et al. Collaborative community mental health and aged care services with peer support to prevent late-life depression: study protocol for a non-randomised controlled trial. Trials 2022 Apr 11;23(1):280 [FREE Full text] [doi: 10.1186/s13063-022-06122-1] [Medline: 35410292]
23. El-Koka A, Cha KH, Kang DK. Regularization parameter tuning optimization approach in logistic regression. In: Proceedings of the 15th International Conference on Advanced Communications Technology. 2013 Presented at: ICACT '13; January 27-30, 2013; PyeongChang, Korea p. 13-18 URL: https://ieeexplore.ieee.org/document/6488130
24. Levis B, Benedetti A, Levis AW, Ioannidis JP, Shrier I, Cuijpers P, et al. Selective cutoff reporting in studies of diagnostic test accuracy: a comparison of conventional and individual-patient-data meta-analyses of the patient health questionnaire-9 depression screening tool. Am J Epidemiol 2017 May 15;185(10):954-964 [FREE Full text] [doi: 10.1093/aje/kww191] [Medline: 28419203]
25. Arrieta J, Aguerrebere M, Raviola G, Flores H, Elliott P, Espinosa A, et al. Validity and utility of the Patient Health Questionnaire (PHQ)-2 and PHQ-9 for screening and diagnosis of depression in rural Chiapas, Mexico: a cross-sectional study. J Clin Psychol 2017 Sep;73(9):1076-1090 [FREE Full text] [doi: 10.1002/jclp.22390] [Medline: 28195649]
26. Kennedy SH. Core symptoms of major depressive disorder: relevance to diagnosis and treatment. Dialogues Clin Neurosci 2008;10(3):271-277 [FREE Full text] [doi: 10.31887/DCNS.2008.10.3/shkennedy] [Medline: 18979940]
27. Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, et al. Accuracy of the PHQ-2 alone and in combination with the PHQ-9 for screening to detect major depression: systematic review and meta-analysis. JAMA 2020 Jun 09;323(22):2290-2300 [FREE Full text] [doi: 10.1001/jama.2020.6504] [Medline: 32515813]
28. Feng L, Hancock R, Watson C, Bogley R, Miller ZA, Gorno-Tempini ML, et al. Development of an abbreviated adult reading history questionnaire (ARHQ-Brief) using a machine learning approach. J Learn Disabil 2022 Sep;55(5):427-442 [FREE Full text] [doi: 10.1177/00222194211047631] [Medline: 34628989]
29. Paykel ES. Basic concepts of depression. Dialogues Clin Neurosci 2008;10(3):279-289 [FREE Full text] [doi: 10.31887/DCNS.2008.10.3/espaykel] [Medline: 18979941]
30. Rosenström T, Elovainio M, Jokela M, Pirkola S, Koskinen S, Lindfors O, et al. Concordance between composite international diagnostic interview and self-reports of depressive symptoms: a re-analysis. Int J Methods Psychiatr Res 2015 Sep;24(3):213-225 [FREE Full text] [doi: 10.1002/mpr.1478] [Medline: 26140369]
31. Burnam MA, Wells KB, Leake B, Landsverk J. Development of a brief screening instrument for detecting depressive disorders. Med Care 1988 Aug;26(8):775-789. [doi: 10.1097/00005650-198808000-00004] [Medline: 3398606]
32. Smits N. A note on Youden's J and its cost ratio. BMC Med Res Methodol 2010 Sep 30;10:89 [FREE Full text] [doi: 10.1186/1471-2288-10-89] [Medline: 20920288]

Abbreviations

AUC: area under the curve
CV: cross-validation
DSM-V: Diagnostic and Statistical Manual of Mental Disorders, fifth edition
JC: Jockey Club
LR: logistic regression
MDD: major depressive disorder
MexMedStudents: Mexican Medical Students
ML: machine learning
NPV: negative predictive value
PHQ: Patient Health Questionnaire
phq2&4: Patient Health Questionnaire–9 items 2 and 4
phq2&8: Patient Health Questionnaire–9 items 2 and 8
phq2: Patient Health Questionnaire–9 item 2
phq4: Patient Health Questionnaire–9 item 4
PNS2013: Pesquisa Nacional de Saúde 2013
PNS2019: Pesquisa Nacional de Saúde 2019
PPV: positive predictive value
ROC: receiver operating characteristic

Edited by J Torous; submitted 24.04.23; peer-reviewed by T Liu, N Mungoli, Y Zheng; comments to author 18.05.23; revised version received 13.07.23; accepted 10.08.23; published 19.10.23.

Please cite as:
Glavin D, Grua EM, Nakamura CA, Scazufca M, Ribeiro dos Santos E, Wong GHY, Hollingworth W, Peters TJ, Araya R, Van de Ven P
Patient Health Questionnaire-9 Item Pairing Predictiveness for Prescreening Depressive Symptomatology: Machine Learning Analysis
JMIR Ment Health 2023;10:e48444
URL: https://mental.jmir.org/2023/1/e48444
doi:10.2196/48444
PMID:37856186

Original Paper

Comparing the Acceptability and Quality of Intervention Modalities for Suicidality in the Emergency Department: Randomized Feasibility Trial

Celine Larkin[1], PhD; Bengisu Tulu[2], PhD; Soussan Djamasbi[2], PhD; Roscoe Garner[1], BS; Fatima Varzgani[2], MS; Mariam Siddique[1], MS; John Pietro[1], MS; Edwin D Boudreaux[1], PhD

[1]Department of Emergency Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[2]The Business School, Worcester Polytechnic Institute, Worcester, MA, United States

Corresponding Author:
Celine Larkin, PhD
Department of Emergency Medicine
University of Massachusetts Chan Medical School
55 Lake Avenue North
Worcester, MA, 01655
United States
Phone: 1 5084211436
Email: celine.larkin@umassmed.edu

Abstract

Background: Emergency departments (EDs) manage many patients with suicide risk, but effective interventions for suicidality are challenging to implement in this setting. ReachCare is a technology-facilitated version of an evidence-based intervention for suicidal ED patients. Here, we present findings on the acceptability and quality of ReachCare in the ED, as well as a comparison of these measures across 3 potential delivery modalities

Objective: Our aim was to test the feasibility of the ReachCare intervention in its entirety through conducting a pilot study with patients presenting with suicidality to the ED. We tested three different ways of receiving the ED-based components of ReachCare: (1) self-administered on the tablet app using a chatbot interface, (2) administered by an in-person clinician, or (3) administered by a telehealth clinician.

Methods: In total, 47 ED patients who screened positive for suicide risk were randomly allocated to receive one of three delivery modalities of ReachCare in the ED: (1) self-administered on the patient-facing tablet app with a chatbot interface, (2) delivered by an in-person clinician, or (3) delivered by a telehealth clinician, with the latter two using a clinician-facing web app. We measured demographic and clinical characteristics, acceptability and appropriateness of the intervention, and quality and completeness of the resulting safety plans.

Results: Patients assigned high ratings for the acceptability (median 4.00/5, IQR 4.00-4.50) and appropriateness (median 4.00/5, IQR 4.00-4.25) of ReachCare's ED components, and there were no substantial differences across the 3 delivery modalities [H(acceptability)=3.90, *P*=.14; H(appropriateness)=1.05, *P*=.59]. The self-administered modality took significantly less time than the 2 clinician modalities (H=27.91, *P*<.001), and the usability of the self-administered version was in the "very high" range (median 93.75/100, IQR 80.00-97.50). The safety plans created across all 3 modalities were high-quality (H=0.60, *P*=.74).

Conclusions: Patients rated ReachCare in the ED as highly acceptable and appropriate regardless of modality. Self-administration may be a feasible way to ensure patients with suicide risk receive an intervention in resource constrained EDs. Limitations include small sample size and demographic differences between those enrolled versus not enrolled. Further research will examine the clinical outcomes of patients receiving both the in-ED and post-ED components of ReachCare.

Trial Registration: ClinicalTrials.gov NCT04720911; https://clinicaltrials.gov/ct2/show/NCT04720911

(JMIR Ment Health 2023;10:e49783) doi:10.2196/49783

KEYWORDS

suicide; self-harm; emergency department; mobile app; intervention; mobile technology; safety planning; safety; suicidal; emergency; mHealth; mobile health; mental health; mobile phone

Introduction

After a brief decline in 2019-2020 [1,2], suicide rates have again begun to rise in the United States [1]. Each year, there are over 1 million suicide-related visits to emergency departments (EDs) in the United States [3], and EDs represent a feasible venue to screen and assess suicide risk [4,5]. However, once risk is detected, many ED patients with suicide risk receive suboptimal care [6].

Several effective interventions have been developed for this patient population [7]. The Emergency Department Safety Assessment and Follow-up Evaluation (ED-SAFE) intervention [8], for example, featured paper-based self-administered safety planning and resource lists in the ED; this was followed by intensive telephone outreach to patients and a chosen significant other in the 12 months after discharge. The telephone counseling involved safety planning, outpatient treatment engagement, values clarification and life planning, and family communication and coordination. The ED-SAFE intervention was associated with a 30% reduction in the total number of suicide attempts in the intervention group compared to treatment-as-usual [8]. Despite its promise and cost-effectiveness [9], the ED-SAFE intervention was resource-intensive and difficult to sustain without research support. Such interventions routinely struggle to "cross the chasm" to become routine care [10].

Clinicians and leadership recognize that ED-based care for suicidality needs to improve but often face substantial constraints in effecting change [11,12]. The dearth of effective care for suicidality in the ED can be attributed to a wide range of factors, including a hectic milieu focused on patient flow [13], a shortage of behavioral health (BH) providers [14], and a shortfall in suicide-related skills and confidence among ED clinicians [15,16]. With these barriers in mind, we set out to adapt the ED-SAFE intervention to become "ReachCare," leveraging technology to make the intervention scalable in resource-constrained EDs, while avoiding a corresponding loss in fidelity.

We previously outlined the user-centered development and testing of the patient-facing technology aspects of the ReachCare intervention [17]. These technologies include a tablet-based app for use within the ED (that allows a patient to build a high-quality safety plan in the absence of a BH clinician) and a patient- and family-facing mobile app (that houses the patient's safety plan and life plan, helpline resources, psychoeducation, distractions, and a BH provider search engine). In that study, user-testing with current ED patients presenting with suicidality demonstrated high usability and acceptability of the technologies [17]. ReachCare is further supported by interoperable clinician-facing software that provides tips, call structure, and note-taking for clinicians, as well as allowing them to update and share patients' therapeutic tools during follow-up phone calls. This clinician-facing technology was designed to be used in-person or via telehealth.

We conducted a feasibility trial with patients presenting with suicidality to the ED to develop and test implementation strategies for the ReachCare intervention that will be used in a future effectiveness trial. We tested three different modalities of delivering the ED-based components of ReachCare, randomizing patients to receive: (1) self-administered on the tablet app using a chatbot interface, (2) administered by an in-person clinician, or (3) administered by a telehealth clinician. Here, we present the findings from the ED-based portion of the feasibility trial, namely the acceptability and appropriateness of ReachCare in the ED overall, the quality and completeness of safety plans created in the ED, intervention duration, and a comparison of these measures across the 3 potential delivery modalities.

Methods

Ethical Considerations

This study was approved by the institutional review board of the University of Massachusetts Chan Medical School on May 19, 2020, under docket number H00020238.

Study Design

This study was a feasibility trial of the ReachCare intervention, in which eligible patients were randomized to receive ReachCare in the ED in one of three ways: (1) self-administered on the tablet app using a chatbot interface, (2) administered by an in-person clinician, or (3) administered by a telehealth clinician. Randomization was on a 1:1:1 ratio based on blocks of 6: randomization tables were prepared by a team member who was not involved in data collection and was uploaded to the Research Electronic Data Capture (Vanderbilt University) data collection system. Allocation was done after the baseline research interview to preserve blinding of the research assistant collecting the baseline demographic and clinical measures. Further, because the research assistant had to know the allocation in order to facilitate the intervention, blinding was not possible for the acceptability measures, which were completed after the intervention. The quality rating of the safety plans was completed after completion of the trial by an assessor blinded to allocation (see more detail under the Measures section). The target sample size of 45 (15 participants per arm) was set based on previous literature recommending sample sizes between 24 and 50 for feasibility studies [18,19].

Recruitment

This study took place in the ED of a large teaching hospital in the northeastern United States. Over a 6-month period, we sought to enroll patients who presented to the ED and screened positive for suicidality on the Patient Safety Screener (PSS) [20]. The PSS is used universally and routinely at the enrollment ED: a positive screen is defined as active ideation in the past 2 weeks or a suicide attempt in the past 6 months. Patients were considered for inclusion if they (1) screened positive for suicidality on the PSS, (2) were aged 18 years or older, (3) were cognitively and medically able to consent and participate, (4) owned a smartphone, and (5) had a stable mailing address to receive compensation. Patients were excluded if they were (1) a prisoner or in police custody, (2) overly agitated or violent, or (3) on precautions for infectious disease, such as COVID-19.

Potentially eligible patients were approached by a research assistant in the ED's medical or psychiatric area and were screened to confirm eligibility. The research assistant then

explained the study: patients were informed as to what the study entailed, procedures to protect confidentiality and privacy, their right to withdraw at any time, and the potential risks and benefits of the study. If the patient was still interested in participating, the research assistant then administered a consent mini-quiz to test the patient's understanding of the information and invited them to read and sign an informed consent form. Participants then completed baseline research measures, were randomly allocated to 1 of 3 arms, participated in the ED portion of the intervention, and then completed acceptability measures. Participants received a US $30 gift-card for completion of the baseline research interview.

Intervention

The ED-based portion of the ReachCare intervention focuses on creating a 6-step safety plan [21]. The safety plan involves working with the patient to identify warning signs of an impending suicidal crisis, internal coping strategies, social supports and places for distraction, social supports for help, professionals for help, and making the environment safer by limiting access to lethal means (Figure 1). In the self-administered arm of the trial, patients engaged with a tablet-based app: they watched an introductory video (with the choice of watching a clinician, community member, or animated character deliver the same information; Figure 2) and then proceeded to create a safety plan by interacting with a chatbot-style interface that presented cues and questions, to which the patient typed in personalized answers [17]. Once the safety plan had been created, the patient had the opportunity to review and edit it, before watching an outro video explaining the next steps in their care. In contrast, in the clinician in-person and telehealth arms, the patient completed safety planning with the clinician. In those arms, the clinician used the ReachCare clinician portal, which provides clinician-facing cues, suggested phrasing, and structure to introduce the intervention to the patient and to create a safety plan with them.

Figure 1. Screenshot of the ReachCare tablet-based app's safety planning functionality.

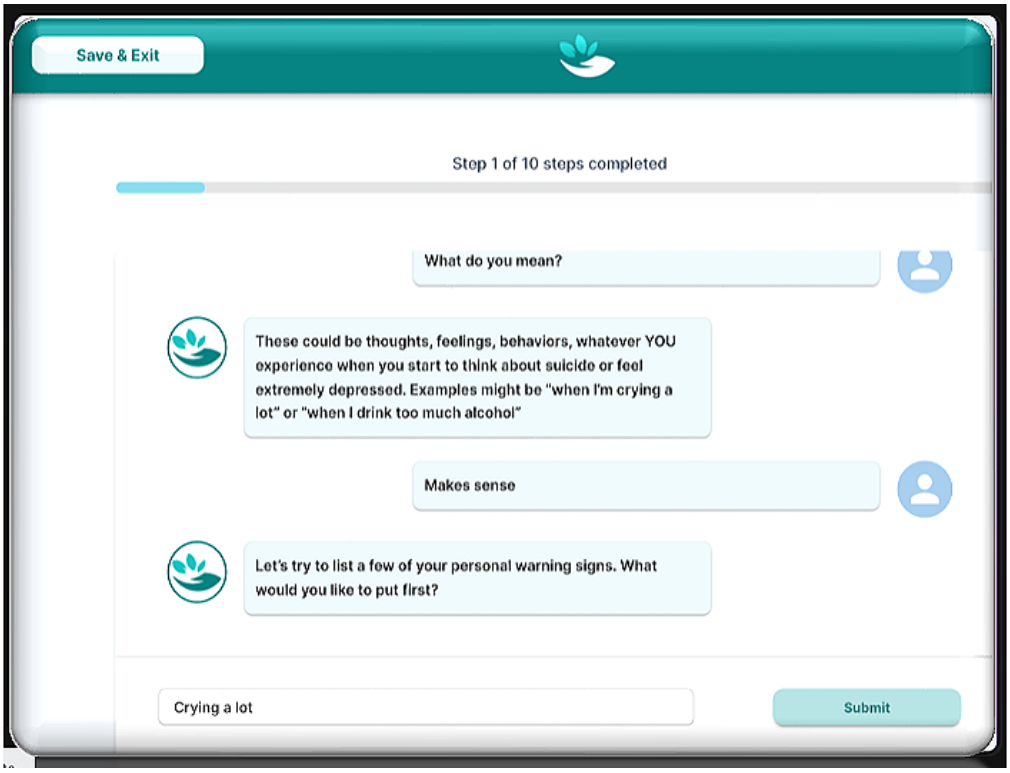
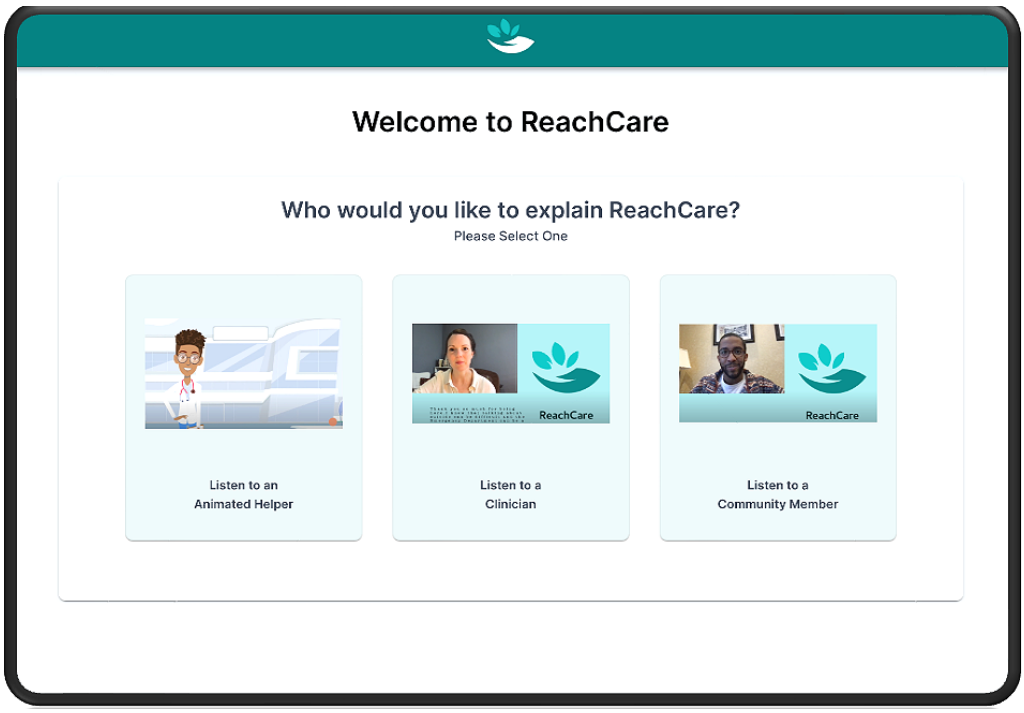


Figure 2. Screenshot of the ReachCare tablet-based app's video functionality.



Measures

Our measures included a variety of modalities, including patient self-report via interview, medical chart review, researcher measurement (of duration of intervention), and researcher rating (of safety plan quality).

We assessed several demographic variables by patient self-report, namely gender, age, race, ethnicity, psychiatric diagnoses, BH treatments received in the past 3 months, access to lethal means, and smartphone usage. We assessed mobile technology self-efficacy by patient self-report using 3 items from Balapour et al [22]. The 15-item version of the Interpersonal Needs Questionnaire [23] was used to assess patient's self-reported thwarted belonginess and perceived burdensomeness, key drivers of suicidality. The Interpersonal Needs Questionnaire includes a 7-point Likert Scale and has good psychometric properties [24]. We used the Columbia Suicide Severity Rating Scale to classify patient's self-reported recent and past suicidal ideation and suicidal behavior [25] and suicide-related impulsivity was assessed using 2 items from Mccullumsmith et al [26].

We used the medical record to extract the patient's insurance status and the circumstances of their current presentation (chief complaint, involuntary hold status, and suicide screening result). We also extracted information regarding the care they received in the ED beyond the research intervention, that is, whether they had received each of the following: a BH evaluation; referral to outpatient BH; a structured suicide risk assessment; constant observation; or hospital admission.

We investigated acceptability and appropriateness, usability (for the self-administered arm only), duration of intervention in minutes, and quality of safety plans. Acceptability and appropriateness were measured using Weiner et al's [27] 4-item Acceptability of Intervention measure and Intervention Appropriateness Measure respectively. Both instruments are on a 4-point Likert-type scale and have good reliability and validity [27]. Usability was assessed using the System Usability Scale [28]: this scale has 10 items, 5 of which are reverse-scored, and is the standard in usability research. The range of possible scores on the System Usability Scale is from 0 to 100. The duration of the intervention was timed by the research assistant and rounded to the closest minute. The quality of the safety plans was rated using the Safety Planning Intervention Scoring Algorithm-Brief [29], which assesses whether the safety plan has content in each line and whether that content is sufficiently specific. Each line receives a rating of 0 (no content), 1 (poor content), or 2 (satisfactory content), which is summed to create a total quality score. This quality rating was completed by a researcher who was blinded to the allocation of the patient. This scale's psychometric properties have not been established but, in our sample, the value for Cronbach α for the 19-item scale was α=.70, indicating acceptable internal consistency.

Analyses

For the first part of the results, we summarized patient and encounter characteristics using descriptive statistics. We then examined differences in implementation measures (median duration, acceptability, appropriateness, and quality) across the 3 arms: these variables were not normally distributed so we used independent samples Kruskal-Wallis tests including post hoc tests as appropriate. Finally, we summarize completeness of the safety plans in each arm using descriptive statistics. All tests were 2-sided with α set at $P<.05$. Analyses were completed in SPSS (version 28; IBM Corp) [30].

Results

Characteristics of Sample

Figure 3 illustrates the patient enrollment flow. A total of 592 patients screened positive for suicide risk during research shifts and their medical charts were reviewed for initial eligibility: 144 of these were approached. Common reasons for nonapproach included the patient being: too ill, cognitively unable, agitated, without a stable mailing address, and being on infection control precautions. Of the 144 patients approached, 78 (55%) agreed to complete verbal eligibility screening. Of these, 64 were confirmed eligible, of whom 47 patients agreed to participate and were randomized: 16 participants to "arm 1 self-administered," 15 participants to "arm 2 clinician in-person," and 16 to "arm 3: telehealth clinician" (Figure 2). In total, 46 participants completed the intervention and all research measures; 1 patient allocated to arm 3 (telehealth clinician) withdrew during the intervention. Comparing those enrolled (N=47) to other patients who were screened for potential participation (n=545), those who were enrolled were more likely to be female (24/47, 51% vs 238/545, 44%), non-Hispanic (40/47, 85% vs 433/545, 79%), White (38/47, 81% vs 388/545, 71%), and younger (median 27.0, IQR 23-37 years vs 39.6, IQR 29-52 years). The characteristics of the enrolled sample are summarized in Tables 1 and 2 and are presented by study arm in Multimedia Appendix 1. Participants demonstrated high prevalence of depressive and anxiety disorders, high levels of smartphone use and self-efficacy, and high levels of perceived burdensomeness and thwarted belongingness.

Features of the participants' index ED visit are summarized in Table 3, and are presented by study arm in Multimedia Appendix 2. Most participants were enrolled in the psychiatric area of the ED, almost all had a psychiatric chief complaint, and many were in the ED on an involuntary hold. Most participants received a BH evaluation during their ED visit and subsequent admission to a psychiatric facility.

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) flowchart showing progress of participants through the trial.

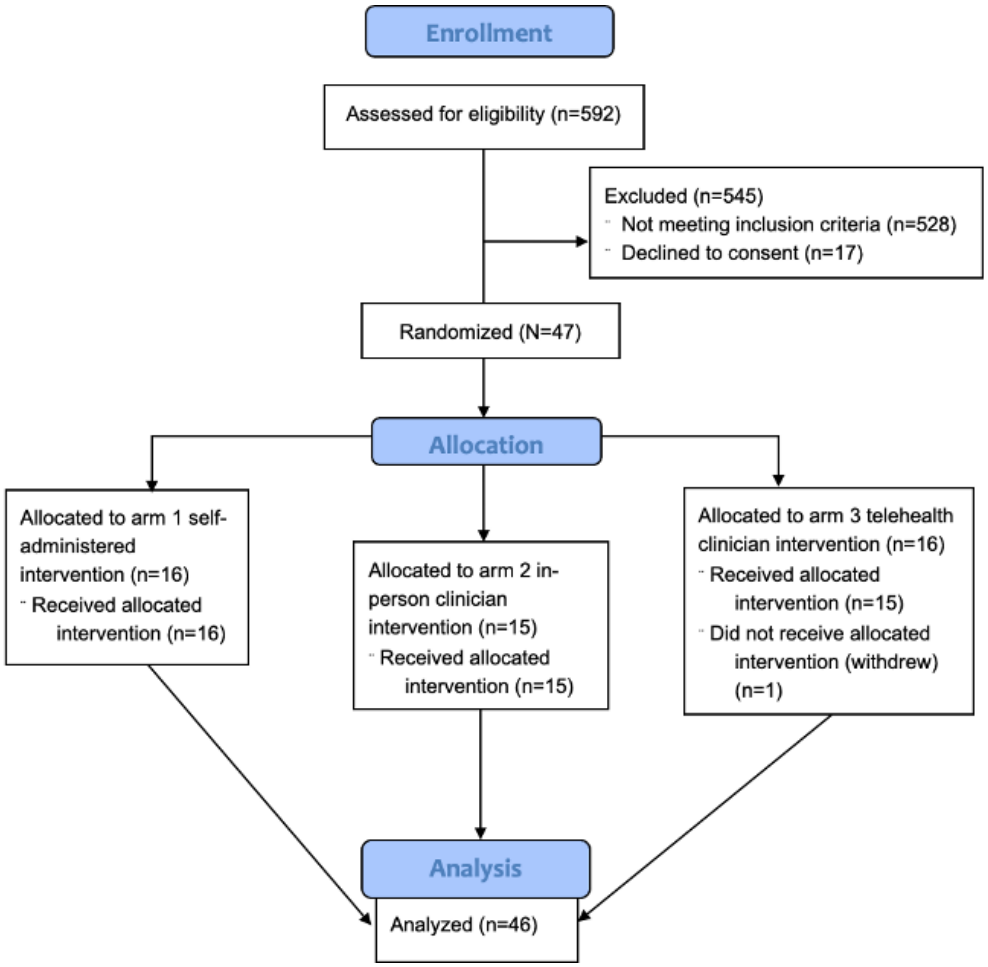


Table 1. Patient characteristics at baseline (N=47).

Categorical variables	Patients, n (%)
Gender	
Female	24 (51.1)
Male	20 (42.6)
Nonbinary	3 (6.4)
Age (years)	
18-24	16 (34.0)
25-34	17 (36.2)
35+	14 (28.8)
Race (check all that apply)	
White	38 (80.9)
Other	5 (10.6)
Black or African American	4 (8.5)
Asian	2 (4.3)
American Indian or Alaskan Native	1 (2.1)
Native Hawaiian or Pacific Islander	0 (0)
Ethnicity	
Non-Hispanic or Latinx	40 (85.1)
Hispanic or Latinx	7 (14.9)
Insurance	
Private	23 (48.9)
Public (eg, Medicare and Medicaid)	19 (40.4)
None or self-pay	5 (10.6)
Psychiatric diagnoses (check all that apply)	
Depressive disorder	40 (85.1)
Anxiety disorder	38 (80.9)
Trauma or stress-related	24 (51.1)
ADHD[a]	21 (44.7)
Substance use disorder	16 (34.0)
Personality disorder	12 (25.5)
Eating disorder	5 (10.6)
Schizophrenia or psychotic disorder	3 (6.4)
Autism spectrum disorder	2 (4.3)
Other	2 (4.3)
Suicidal behavior	
Past-week active ideation	37 (78.7)
Lifetime attempt	30 (63.8)
Current attempt	7 (14.9)
BH[b] treatments received in past 3 months	
Prescription medication	37 (78.7)
Individual therapy	27 (57.4)
Partial hospitalization or intensive outpatient	17 (36.2)

Categorical variables	Patients, n (%)
BH ED[c] visit	12 (25.5)
Group support	11 (23.4)
Inpatient psychiatric hospitalization	9 (19.1)
Helpline support	9 (19.1)
Other	9 (19.1)
Family therapy or couples counseling	4 (8.5)
Lethal means	
Access to medication	40 (85.1)
Access to firearm	0 (0)
Smartphone use in past week per day	
>3 h a day	35 (74.5)
30 min to 3 h	9 (19.1)
<30 min a day	1 (2.1)
Rarely or never	2 (4.3)

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]BH: behavioral health.

[c]ED: emergency department.

Table 2. Patient characteristics at baseline (N=47).

Continuous variables	Mean (SD)
Smartphone self-efficacy	
I can use smartphone technology if there was no one around to tell me what to do (max score 5)	4.62 (0.68)
I can use smartphone technology even if I have never used a similar technology before (max score 5)	4.15 (1.02)
I am confident that I can effectively open and use an app on my smartphone (max score 5)	4.79 (0.46)
Hypothesized mechanisms of action	
INQ[a] perceived burdensomeness subscale	19.57 (10.04)
INQ thwarted belongingness subscale	33.83 (13.24)
BAS[b]-drive subscale	11.17 (3.01)
Suicide-related impulsivity	3.02 (1.35)

[a]INQ: Interpersonal Needs Questionnaire.

[b]BAS: behavioral activation scale.

Table 3. Characteristics of index ED[a] visit (N=47).

	Patients, n (%)
Enrollment location	
Psychiatric ED	30 (63.8)
Medical ED	17 (36.2)
Chief complaint	
Suicidality	35 (74.5)
Nonsuicide psychiatric	11 (23.4)
Nonpsychiatric	1 (2.2)
Involuntary BH[b] hold	
Yes	41 (87.2)
No	6 (12.8)
Patient Safety Screener result (check all that apply)	
Positive for active ideation in past 2 wk	43 (91.5)
Positive for suicide attempt in past 6 mo	15 (31.9)
BH Evaluation	
Yes	40 (85.1)
No	7 (14.9)
Referral to outpatient BH	
Yes	1 (2.1)
No	46 (97.9)
Structured suicide assessment	
Yes	26 (55.3)
No	21 (44.7)
Under observation	
Yes	45 (95.7)
No	2 (4.3)
Disposition	
Psychiatric admission or transfer	29 (61.7)
Home	16 (34.0)
Other or not documented	2 (4.3)

[a]ED: emergency department.

[b]BH: behavioral health.

Implementation Outcomes of ReachCare in the ED

Table 4 summarizes the response of participants to ReachCare in the ED. Participants rated the acceptability and appropriateness of ReachCare in the ED highly, with no significant difference across the 3 delivery modalities. The usability of the self-administered (chatbot) modality was also rated very highly by participants, and post hoc tests showed that the self-administered arm took significantly less time to complete than the 2 clinician arms. The overall quality of the safety plans generated were not significantly different across the 3 arms (Table 4).

Table 5 summarizes the completeness of each step of the safety plans, compared across the 3 study arms. There were no statistical differences in the completeness of each step across arms, but Step 2 tended to be more complete in the self-administered arm (H=5.14, *P*=.08) and Step 5 tended to be more complete in the 2 clinician arms (H=4.96, *P*=.08).

Table 4. Implementation outcomes postintervention (n=46) by treatment arm.

	Median (IQR)	Kruskal-Wallis H	*P* value
Acceptability of intervention (patient report, range 1-5)		3.90	.14
Overall	4.00 (4.00-4.50)		
Arm 1 self	4.00 (3.81-4.56)		
Arm 2 in-person	4.25 (4.00-4.50)		
Arm 3 telehealth	4.00 (3.75-4.50)		
Appropriateness of intervention (patient report, range 1-5)		1.05	.59
Overall	4.00 (4.00-4.25)		
Arm 1 self	4.25 (3.75-4.25)		
Arm 2 in-person	4.00 (4.00-4.50)		
Arm 3 telehealth	4.00 (4.00-4.25)		
System Usability Scale (patient report, range 0-100)		N/A[a]	N/A
Arm 1 only	93.75 (80.00-97.50)		
Duration of intervention (researcher-measured, min)		27.91	<.001[b]
Overall	34.50 (16.50-41.75)		
Arm 1 self	15.00 (14.75-17.75)		
Arm 2 in-person	38.00 (33.50-41.75)		
Arm 3 telehealth	42.50 (34.50-53.38)		
Quality of safety plan (researcher-rated, range 0-38)		0.60	.74
Overall	30.00 (26.00-33.00)		
Arm 1 self	30.00 (26.00-32.75)		
Arm 2 in-person	30.00 (24.00-33.00)		
Arm 3 telehealth	28.00 (26.00-34.00)		

[a]N/A: not applicable.

[b]Post hoc tests showed arm 1 had significantly shorter duration than arms 2 and 3 ($P<.001$).

Table 5. Safety plan completeness by step overall and by treatment arms (n=46).

Safety plan step	Number of lines with content, median (IQR)				H	*P* value
	Overall (all 3 arms)	Arm 1 self-administered	Arm 2 clinician in-person	Arm 3 clinician telehealth		
Step 1 warning signs (out of 3 lines)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	1.28	.53
Step 2 coping strategies (out of 3 lines)	3.0 (2.0-3.0)	3.0 (3.0-3.0)	3.0 (2.0-3.0)	2.0 (2.0-3.0)	5.14	.08
Step 3 people or places for distraction (out of 4 lines)	3.0 (2.8-4.0)	4.0 (2.0-4.0)	3.0 (3.0-4.0)	3.0 (2.0-4.0)	1.02	.60
Step 4 people for help (out of 3 lines)	2.0 (1.0-2.3)	2.0 (1.3-3.0)	2.0 (1.0-3.0)	2.0 (1.0-2.0)	1.12	.57
Step 5 professionals for help (out of 4 lines)	3.0 (3.0-4.0)	3.0 (2.0-3.0)	3.0 (3.0-4.0)	3.0 (3.0-4.0)	4.96	.08
Step 6 making environment safe (out of 2 lines)	2.0 (1.0-2.0)	2.0 (2.0-2.0)	2.0 (1.0-2.0)	2.0 (1.0-2.0)	2.84	.24

Discussion

Many patients who present to the ED with suicidality do not receive effective intervention [6]. EDs struggle with shortages in BH clinician availability and suicide-specific skills [15,16] and patients often describe their experiences of suicide-related care in the ED as negative [31]. ReachCare was developed to initiate an acceptable and feasible form of evidence-based care in the ED and support the patient in the high-risk period after discharge [17]. In this paper, we summarized initial implementation outcomes of the in-ED components of ReachCare across a variety of modalities. We found that patients rated the acceptability and appropriateness of the intervention highly in the context of other studies that have used the same measures [32-34], and there were no substantial differences across the various modalities. At just over 15 minutes, the self-administered, chatbot-style modality took less than half the time of the clinician modalities and did not require BH specialist

input. The self-administered safety plans tended to have more content and the usability of the patient-facing tablet app was in the "very high" range. The safety plans that were created with ReachCare, whether self-administered with a chatbot or delivered by a clinician, were of high quality and completeness. Given that higher safety plan quality is associated with reduced risk of prospective suicidal behavior [35,36] and of subsequent psychiatric hospitalization [37], ReachCare has the potential to positively impact patient outcomes in a fully powered trial, even when it is implemented using self-administered modality.

Technology is becoming a key implementation strategy for BH interventions in health care settings [38]. Brief interventions for suicidality in the ED are recommended by the Joint Commission [39] and the American College of Emergency Physicians [40] but in practice have proven difficult to implement. In EDs, where there is a dearth of BH clinicians and the focus is on evaluation and disposition, engaging methods of self-administration may be one of the few ways to ensure that patients receive any intervention. Self-guided digital interventions for suicidality have been found to have a significant, if small, effect [41]. This study contributes to growing evidence that, when delivered through engaging digital platforms, self-guided versions of effective interventions can be efficient, feasible, and acceptable. For example, the tablet-based intervention "Jaspr" was associated with better care process and decreased distress in patients presenting with suicidality [42]. Similarly, a self-administered tablet-based safety planning website designed for use in the ED was associated with good usability and reduced suicidal intensity [43]. Finally, "Lock to Live," is a promising tablet-based intervention for counseling on access to lethal means in the ED [44,45]. These encouraging results support the idea that self-administration could offer unique advantages in terms of timely, tailored, and scalable suicide intervention in resource-constrained EDs.

Like the broader population of people presenting with suicidality, our sample displayed a high prevalence of depression, anxiety disorder, trauma, and attention-deficit/hyperactivity disorder, as well as elevated levels of thwarted belongness and perceived burdensomeness. Given that those who are experiencing suicidality may be feeling particularly disconnected [46], there are ethical implications around simply handing a patient a tablet computer as a panacea to suicide-related care in the ED. The development of such technology-facilitated intervention should start with careful mapping of user needs from a variety of stakeholders and apply an iterative approach to testing and refinement [17,47]. Although self-administration holds promise as a way of scaling evidence-based interventions, programs that propagate loneliness and disempowerment in suicidal patients have the potential to do more harm than good. Any self-administered software for this population should be engaging but easy to navigate, foster a sense of connection as much as possible (for example by using videos and interaction), and anticipate and seek to ameliorate feelings of burdensomeness and disconnection. When done well, it is possible that technology could increase empowerment by introducing flexibility and choice to fit with patients' timeline and needs [48,49]. Self-administered intervention in this population may be thoughtfully introduced by a clinician, such as a physician, nurse, or a medical assistant: the patient could be offered a choice, to complete a safety plan with a BH clinician (in-person or via telehealth if necessary) or for themselves if they prefer. A third option is a hybrid approach, where a patient initiates the safety plan using software and a BH clinician reviews it with them, ideally before discharge or as part of telephone follow-up. Intervention is ideally not limited to the ED but carries on out into the community in the form of caring contacts or outpatient engagement that can help to address the underlying drivers of the individual's suicidality [50]. It is also important to note that technology is not a sufficient implementation strategy in itself: digital interventions require careful implementation planning, training, and support if they are to result in improved reach and sustainability [51].

Finally, our study led to some interesting observations around digitally supported clinician delivery. There appeared to be a slight trend toward lower acceptability of the telehealth clinician modality compared to the in-person clinician modality in our small trial. Larger trials that have compared telehealth to in-person delivery for a variety of psychosocial interventions have found little difference in satisfaction or effectiveness between the 2 modalities [52-54]. Moreover, telehealth evaluations can help to improve access and timeliness of mental health care in ED settings [55,56]. Further research is required to examine whether telemental health interventions specifically are less acceptable in an emergency medicine context.

This study has several limitations. Although the broader trial involved 3-month follow-up of clinical outcomes, the current analyses were cross-sectional only and our sample size was relatively small. In addition, the patients who were successfully enrolled tended to be younger, female, White, and non-Hispanic than those who were not enrolled. This may reflect the fact that the intervention was only available in English and trial enrollment required the patient to own a smartphone: additional user-centered research is needed to identify the needs of underserved populations and design systems that can adapt to their needs and preferences. For example, future research should explore Spanish-language adaptation of the intervention, as well as adaptations of the post-ED components that can be delivered without a smartphone. Notably, the majority of patients who screened positive for suicide risk during the trial were not eligible to participate because of medical or cognitive issues, which interfered with their ability to give informed consent: it remains to be seen what the reach of this intervention would be under naturalistic clinical conditions, where the capacity threshold for participation may be more flexible.

Our findings suggest that patients considered the ReachCare intervention in the ED to be acceptable and appropriate across a variety of modalities. Technology facilitation of evidence-based interventions in the ED holds much promise, with the caveat that we must take account of the particular and diverse needs of those in suicidal crisis in this setting. There are several potential directions for future research. These include possible adaptations for older patients, those whose preferred language is not English, and patients who do not own a smartphone. Other potential research questions include: which intervention modality patients or clinicians tend to select when

given a choice; whether that choice is affected by patient characteristics such as age and medical condition; whether the selected modality affects longitudinal clinical outcomes; and rates of uptake under more naturalistic implementation conditions.

Acknowledgments

This research was supported by the National Institute of Mental Health of the National Institutes of Health (award R34MH123578). The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We gratefully acknowledge the contribution of our participants and consultants.

Data Availability

Data are available in the National Institute of Mental Health Data Archive.

Conflicts of Interest

The authors of this paper include the developers of the ReachCare app.

Multimedia Appendix 1
Patient characteristics at baseline (N=47) by study arm.
[DOCX File , 21 KB - mental_v10i1e49783_app1.docx]

Multimedia Appendix 2
Characteristics of index ED visit (N=47) by arm.
[DOCX File , 16 KB - mental_v10i1e49783_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1264 KB - mental_v10i1e49783_app3.pdf]

References

1. Stone DM, Jones CM, Mack KA. Changes in suicide rates—United States, 2018-2019. MMWR Morb Mortal Wkly Rep 2021;70(8):261-268 [FREE Full text] [doi: 10.15585/mmwr.mm7008a1] [Medline: 33630824]
2. Ehlman DC, Yard E, Stone DM, Jones CM, Mack KA. Changes in suicide rates—United States, 2019 and 2020. MMWR Morb Mortal Wkly Rep 2022;71(8):306-312 [FREE Full text] [doi: 10.15585/mmwr.mm7108a5] [Medline: 35202357]
3. Owens PL, McDermott KW, Lipari RN, Hambrick MM. Emergency Department Visits Related to Suicidal Ideation or Suicide Attempt, 2008-2017. Rockville (MD): Agency for Healthcare Research and Quality (US); 2020.
4. Ballard ED, Cwik M, Van Eck K, Goldstein M, Alfes C, Wilson ME, et al. Identification of at-risk youth by suicide screening in a pediatric emergency department. Prev Sci 2017;18(2):174-182 [FREE Full text] [doi: 10.1007/s11121-016-0717-5] [Medline: 27678381]
5. Grumet JG, Boudreaux ED. Universal suicide screening is feasible and necessary to reduce suicide. Psychiatr Serv 2023;74(1):81-83. [doi: 10.1176/appi.ps.202100625] [Medline: 35811561]
6. Olfson M, Marcus SC, Bridge JA. Emergency treatment of deliberate self-harm. Arch Gen Psychiatry 2012;69(1):80-88 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2011.108] [Medline: 21893643]
7. Inagaki M, Kawashima Y, Kawanishi C, Yonemoto N, Sugimoto T, Furuno T, et al. Interventions to prevent repeat suicidal behavior in patients admitted to an emergency department for a suicide attempt: a meta-analysis. J Affect Disord 2015;175:66-78 [FREE Full text] [doi: 10.1016/j.jad.2014.12.048] [Medline: 25594513]
8. Miller IW, Camargo CA, Arias SA, Sullivan AF, Allen MH, Goldstein AB, et al. Suicide prevention in an emergency department population: the ED-SAFE study. JAMA Psychiatry 2017;74(6):563-570 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0678] [Medline: 28456130]
9. Dunlap LJ, Orme S, Zarkin GA, Arias SA, Miller IW, Camargo CA, et al. Screening and intervention for suicide prevention: a cost-effectiveness analysis of the ED-SAFE interventions. Psychiatr Serv 2019;70(12):1082-1087 [FREE Full text] [doi: 10.1176/appi.ps.201800445] [Medline: 31451063]
10. Ruffolo MC, Capobianco J. Moving an evidence-based intervention into routine mental health care: a multifaceted case example. Soc Work Health Care 2012;51(1):77-87. [doi: 10.1080/00981389.2011.622674] [Medline: 22251392]
11. True G, Pollock M, Bowden CF, Cullen SW, Ross AM, Doupnik SK, et al. Strategies to care for patients being treated in the emergency department after self-harm: perspectives of frontline staff. J Emerg Nurs 2021;47(3):426-436 [FREE Full text] [doi: 10.1016/j.jen.2020.12.016] [Medline: 33610311]

12. Cullen SW, Diana A, Olfson M, Marcus SC. If you could change 1 thing to improve the quality of emergency care for deliberate self-harm patients, what would it be? A national survey of nursing leadership. J Emerg Nurs 2019;45(6):661-669 [FREE Full text] [doi: 10.1016/j.jen.2019.06.007] [Medline: 31495507]
13. Kirk JW, Nilsen P. Implementing evidence-based practices in an emergency department: contradictions exposed when prioritising a flow culture. J Clin Nurs 2016;25(3-4):555-565 [FREE Full text] [doi: 10.1111/jocn.13092] [Medline: 26818380]
14. Richmond JS, Dragatsi D, Stiebel V, Rozel JS, Rasimas JJ. American association for emergency psychiatry recommendations to address psychiatric staff shortages in emergency settings. Psychiatr Serv 2021;72(4):437-443 [FREE Full text] [doi: 10.1176/appi.ps.201900501] [Medline: 33467869]
15. Betz ME, Sullivan AF, Manton AP, Espinola JA, Miller I, Camargo CA, et al. Knowledge, attitudes, and practices of emergency department providers in the care of suicidal patients. Depress Anxiety 2013;30(10):1005-1012 [FREE Full text] [doi: 10.1002/da.22071] [Medline: 23426881]
16. Rayner G, Blackburn J, Edward KL, Stephenson J, Ousey K. Emergency department nurse's attitudes towards patients who self-harm: a meta-analysis. Int J Ment Health Nurs 2019;28(1):40-53. [doi: 10.1111/inm.12550] [Medline: 30387232]
17. Larkin C, Djamasbi S, Boudreaux ED, Varzgani F, Garner R, Siddique M, et al. ReachCare mobile apps for patients experiencing suicidality in the emergency department: development and usability testing using mixed methods. JMIR Form Res 2023;7:e41422 [FREE Full text] [doi: 10.2196/41422] [Medline: 36705961]
18. Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. J Clin Epidemiol 2012;65(3):301-308. [doi: 10.1016/j.jclinepi.2011.07.011] [Medline: 22169081]
19. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceut Statist 2005;4(4):287-291. [doi: 10.1002/pst.185]
20. Boudreaux ED, Jaques ML, Brady KM, Matson A, Allen MH. The patient safety screener: validation of a brief suicide risk screener for emergency department settings. Arch Suicide Res 2015;19(2):151-160. [doi: 10.1080/13811118.2015.1034604] [Medline: 25826715]
21. Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. Cogn Behav Pract 2012;19(2):256-264. [doi: 10.1016/j.cbpra.2011.01.001]
22. Balapour A, Reychav I, Sabherwal R, Azuri J. Mobile technology identity and self-efficacy: implications for the adoption of clinically supported mobile health apps. Int J Inf Manag 2019;49:58-68 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2019.03.005]
23. Van Orden KA, Cukrowicz KC, Witte TK, Joiner TE. Thwarted belongingness and perceived burdensomeness: construct validity and psychometric properties of the interpersonal needs questionnaire. Psychol Assess 2012;24(1):197-215 [FREE Full text] [doi: 10.1037/a0025358] [Medline: 21928908]
24. Hill RM, Rey Y, Marin CE, Sharp C, Green KL, Pettit JW. Evaluating the interpersonal needs questionnaire: comparison of the reliability, factor structure, and predictive validity across five versions. Suicide Life Threat Behav 2015;45(3):302-314. [doi: 10.1111/sltb.12129] [Medline: 25308815]
25. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-suicide severity rating scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
26. Mccullumsmith CB, Williamson DJ, May RS, Bruer EH, Sheehan DV, Alphs LD. Simple measures of hopelessness and impulsivity are associated with acute suicidal ideation and attempts in patients in psychiatric crisis. Innov Clin Neurosci 2014;11(9-10):47-53 [FREE Full text] [Medline: 25520888]
27. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017;12(1):108 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
28. Brooke J. SUS-a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. Milton Park, Abingdon-on-Thames, Oxfordshire United Kingdom: Taylor & Francis; 1996:4-7.
29. Goodman M, Brown GK, Galfalvy HC, Spears AP, Sullivan SR, Kapil-Pair KN, et al. Group ("Project Life Force") versus individual suicide safety planning: a randomized clinical trial. Contemp Clin Trials Commun 2020;17:100520 [FREE Full text] [doi: 10.1016/j.conctc.2020.100520] [Medline: 32043013]
30. IBM Corp. Released 2021. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp URL: https://www.ibm.com/support/pages/how-cite-ibm-spss-statistics-or-earlier-versions-spss [accessed 2023-09-21]
31. MacDonald S, Sampson C, Turley R, Biddle L, Ring N, Begley R, et al. Patients' experiences of emergency hospital care following self-harm: systematic review and thematic synthesis of qualitative research. Qual Health Res 2020;30(3):471-485 [FREE Full text] [doi: 10.1177/1049732319886566] [Medline: 31933427]
32. Hofmann G, Hampanda K, Harrison MS, Fasano M, Nacht A, Yeoman M. Virtual prenatal and postpartum care acceptability among maternity care providers. Matern Child Health J 2022;26(7):1401-1408 [FREE Full text] [doi: 10.1007/s10995-022-03412-7] [Medline: 35292887]

33. Rosenberg M, Kianersi S, Luetke M, Jozkowski K, Guerra-Reyes L, Shih PC, et al. Wearable alcohol monitors for alcohol use data collection among college students: feasibility and acceptability. Alcohol 2023;111:75-83 [FREE Full text] [doi: 10.1016/j.alcohol.2023.05.007] [Medline: 37295566]
34. Lai-Kwon J, Kelly B, Lane S, Biviano R, Bartula I, Brennan F, et al. Feasibility, acceptability, and utility of a nurse-led survivorship program for people with metastatic melanoma (MELCARE). Support Care Cancer 2022;30(11):9587-9596 [FREE Full text] [doi: 10.1007/s00520-022-07360-4] [Medline: 36136246]
35. Green JD, Kearns JC, Rosen RC, Keane TM, Marx BP. Evaluating the effectiveness of safety plans for military veterans: do safety plans tailored to veteran characteristics decrease suicide risk? Behav Ther 2018;49(6):931-938. [doi: 10.1016/j.beth.2017.11.005] [Medline: 30316491]
36. Chalker SA, Parrish EM, Ceren CSM, Depp CA, Goodman M, Doran N. Predictive importance of social contacts on U.S. veteran suicide safety plans. Psychiatr Serv 2023;74(3):244-249. [doi: 10.1176/appi.ps.202100699] [Medline: 36039556]
37. Gamarra JM, Luciano MT, Gradus JL, Stirman SW. Assessing variability and implementation fidelity of suicide prevention safety planning in a regional VA healthcare system. Crisis 2015;36(6):433-439 [FREE Full text] [doi: 10.1027/0227-5910/a000345] [Medline: 26648231]
38. Hasselberg MJ. The digital revolution in behavioral health. J Am Psychiatr Nurses Assoc 2020;26(1):102-111 [FREE Full text] [doi: 10.1177/1078390319879750] [Medline: 31587624]
39. Joint Commission. Suicide prevention resources to support Joint Commission accredited organizations implementation of NPSG 15.01. 01, revised July 2020. URL: https://www.jointcommission.org/-/media/tjc/documents/resources/patient-safety-topics/suicide-prevention/suicide_prevention_compendium_5_11_20_updated-july2020.pdf [accessed 2023-04-19]
40. Wilson MP, Moutier C, Wolf L, Nordstrom K, Schulz T, Betz ME. ED recommendations for suicide prevention in adults: The ICAR2E mnemonic and a systematic review of the literature. Am J Emerg Med 2020;38(3):571-581 [FREE Full text] [doi: 10.1016/j.ajem.2019.06.031] [Medline: 31493978]
41. Torok M, Han J, Baker S, Werner-Seidler A, Wong I, Larsen ME, et al. Suicide prevention using self-guided digital interventions: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2020;2(1):e25-e36 [FREE Full text] [doi: 10.1016/S2589-7500(19)30199-2] [Medline: 33328037]
42. Dimeff LA, Jobes DA, Koerner K, Kako N, Jerome T, Kelley-Brimer A, et al. Using a tablet-based app to deliver evidence-based practices for suicidal patients in the emergency department: pilot randomized controlled trial. JMIR Ment Health 2021;8(3):e23022 [FREE Full text] [doi: 10.2196/23022] [Medline: 33646129]
43. Boudreaux ED, Brown GK, Stanley B, Sadasivam RS, Camargo CA, Miller IW. Computer administered safety planning for individuals at risk for suicide: development and usability testing. J Med Internet Res 2017;19(5):e149 [FREE Full text] [doi: 10.2196/jmir.6816] [Medline: 28506957]
44. Betz ME, Knoepke CE, Simpson S, Siry BJ, Clement A, Saunders T, et al. An interactive web-based lethal means safety decision aid for suicidal adults (lock to live): pilot randomized controlled trial. J Med Internet Res 2020;22(1):e16253 [FREE Full text] [doi: 10.2196/16253] [Medline: 32012056]
45. Betz ME, Knoepke CE, Siry B, Clement A, Azrael D, Ernestus S, et al. 'Lock to live': development of a firearm storage decision aid to enhance lethal means counselling and prevent suicide. Inj Prev 2019;25(Suppl 1):i18-i24 [FREE Full text] [doi: 10.1136/injuryprev-2018-042944] [Medline: 30317220]
46. Joiner T. Why People Die by Suicide. Cambridge, MA: Harvard University Press; 2005.
47. Akeel AU, Mundy D. Re-thinking technology and its growing role in enabling patient empowerment. Health Informatics J 2019;25(4):1278-1289 [FREE Full text] [doi: 10.1177/1460458217751013] [Medline: 29366367]
48. Lau F. Patient empowerment: the role of technology. In: Lau F, Bartle-Clar JA, Bliss G, Borycki EM, Courtney KL, Kuo AMH, et al, editors. Improving Usability, Safety and Patient Outcomes with Health Information Technology: From Research to Practice. Amsterdam, Berlin, Washington, DC: IOS Press; 2019.
49. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2015;18(5):643-652 [FREE Full text] [doi: 10.1111/hex.12089] [Medline: 23711169]
50. Jobes DA, Gregorian MJ, Colborn VA. A stepped care approach to clinical suicide prevention. Psychol Serv 2018;15(3):243-250. [doi: 10.1037/ser0000229] [Medline: 30080081]
51. Graham AK, Lattie EG, Powell BJ, Lyon AR, Smith JD, Schueller SM, et al. Implementation strategies for digital mental health interventions in health care settings. Am Psychol 2020;75(8):1080-1092 [FREE Full text] [doi: 10.1037/amp0000686] [Medline: 33252946]
52. Herbert MS, Afari N, Liu L, Heppner P, Rutledge T, Williams K, et al. Telehealth versus in-person acceptance and commitment therapy for chronic pain: a randomized noninferiority trial. J Pain 2017;18(2):200-211 [FREE Full text] [doi: 10.1016/j.jpain.2016.10.014] [Medline: 27838498]
53. Morland LA, Mackintosh MA, Greene CJ, Rosen CS, Chard KM, Resick P, et al. Cognitive processing therapy for posttraumatic stress disorder delivered to rural veterans via telemental health: a randomized noninferiority clinical trial. J Clin Psychiatry 2014;75(5):470-476. [doi: 10.4088/JCP.13m08842] [Medline: 24922484]

54. Choi NG, Hegel MT, Marti N, Marinucci ML, Sirrianni L, Bruce ML. Telehealth problem-solving therapy for depressed low-income homebound older adults. Am J Geriatr Psychiatry 2014;22(3):263-271 [FREE Full text] [doi: 10.1097/JGP.0b013e318266b356] [Medline: 23567376]
55. Fairchild RM, Ferng-Kuo SF, Laws S, Rahmouni H, Hardesty D. Telehealth decreases rural emergency department wait times for behavioral health patients in a group of critical access hospitals. Telemed E Health 2019;25(12):1154-1164 [FREE Full text] [doi: 10.1089/tmj.2018.0227] [Medline: 30735100]
56. Southard EP, Neufeld JD, Laws S. Telemental health evaluations enhance access and efficiency in a critical access hospital emergency department. Telemed E Health 2014;20(7):664-668. [doi: 10.1089/tmj.2013.0257] [Medline: 24811858]

Abbreviations

BH: behavioral health
ED: emergency department
ED-SAFE: Emergency Department Safety Assessment and Follow-up Evaluation
PSS: Patient Safety Screener

Edited by J Torous; submitted 09.06.23; peer-reviewed by C King, C Chung; comments to author 29.06.23; revised version received 25.07.23; accepted 27.07.23; published 24.10.23.

Please cite as:
Larkin C, Tulu B, Djamasbi S, Garner R, Varzgani F, Siddique M, Pietro J, Boudreaux ED
Comparing the Acceptability and Quality of Intervention Modalities for Suicidality in the Emergency Department: Randomized Feasibility Trial
JMIR Ment Health 2023;10:e49783
URL: https://mental.jmir.org/2023/1/e49783
doi:10.2196/49783
PMID:37874619

Original Paper

Circadian Reinforcement Therapy in Combination With Electronic Self-Monitoring to Facilitate a Safe Postdischarge Period for Patients With Major Depression: Randomized Controlled Trial

Anne Sofie Aggestrup[1], MSc; Signe Dunker Svendsen[2], MSc; Anne Præstegaard[1], BS; Philip Løventoft[1], MSc; Lasse Nørregaard[1], MSc; Ulla Knorr[1], MD, PhD; Henrik Dam[1], MD; Erik Frøkjær[3], MSc; Konstantin Danilenko[4†], MD, PhD; Ida Hageman[5], MD; Maria Faurholt-Jepsen[1], MD, DMSc; Lars Vedel Kessing[1], MD, DMSc; Klaus Martiny[1], MD, PhD, DMSc

[1]Mental Health Centre Copenhagen, Copenhagen University Hospital, Frederiksberg, Denmark

[2]Mental Health Centre Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark

[3]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

[4]Institute of Neurosciences and Medicine, Novosibirsk, Russian Federation

[5]Mental Health Services, Copenhagen University Hospital, Copenhagen, Denmark

[†]deceased

Corresponding Author:
Klaus Martiny, MD, PhD, DMSc
Mental Health Centre Copenhagen
Copenhagen University Hospital
Frederiksberg Hospital, Nordre Fasanvej 57, Hovedvejen 17, 1st floor
Frederiksberg, 2000
Denmark
Phone: 45 38647102
Email: klaus.martiny@regionh.dk

Abstract

Background: Patients with major depression exhibit circadian disturbance of sleep and mood, and when they are discharged from inpatient wards, this disturbance poses a risk of relapse. We developed a circadian reinforcement therapy (CRT) intervention to facilitate the transition from the inpatient ward to the home for these patients. CRT focuses on increasing the zeitgeber strength for the circadian clock through social contact, physical activity, diet, daylight exposure, and sleep timing.

Objective: In this study, we aimed to prevent the worsening of depression after discharge by using CRT, supported by an electronic self-monitoring system, to advance and stabilize sleep and improve mood. The primary outcome, which was assessed by a blinded rater, was the change in the Hamilton Depression Rating Scale scores from baseline to the end point.

Methods: Participants were contacted while in the inpatient ward and randomized 1:1 to the CRT or the treatment-as-usual (TAU) group. For 4 weeks, participants in both groups electronically self-monitored their daily mood, physical activity, sleep, and medication using the Monsenso Daybuilder (MDB) system. The MDB allowed investigators and participants to simultaneously view a graphical display of registrations. An investigator phoned all participants weekly to coinspect data entry. In the CRT group, participants were additionally phoned between the scheduled calls if specific predefined trigger points for mood and sleep were observed during the daily inspection. Participants in the CRT group were provided with specialized CRT psychoeducation sessions immediately after inclusion, focusing on increasing the zeitgeber input to the circadian system; a PowerPoint presentation was presented; paper-based informative materials and leaflets were reviewed with the participants; and the CRT principles were used during all telephone consultations. In the TAU group, phone calls focused on data entry in the MDB system. When discharged, all patients were treated at a specialized affective disorders service.

Results: Overall, 103 participants were included. Participants in the CRT group had a significantly larger reduction in Hamilton Depression Scale score (P=.04) than those in the TAU group. The self-monitored MDB data showed significantly improved evening mood (P=.02) and sleep quality (P=.04), earlier sleep onset (P=.009), and longer sleep duration (P=.005) in the CRT group than in the TAU group. The day-to-day variability of the daily and evening mood, sleep offset, sleep onset, and sleep quality

were significantly lower in the CRT group (all P<.001) than in the TAU group. The user evaluation was positive for the CRT method and the MDB system.

Conclusions: We found significantly lower depression levels and improved sleep quality in the CRT group than in the TAU group. We also found significantly lower day-to-day variability in daily sleep, mood parameters, and activity parameters in the CRT group than in the TAU group. The delivery of the CRT intervention should be further refined and tested.

Trial Registration: ClinicalTrials.gov NCT02679768; https://clinicaltrials.gov/study/NCT02679768

International Registered Report Identifier (IRRID): RR2-10.1186/s12888-019-2101-z

(JMIR Ment Health 2023;10:e50072) doi:10.2196/50072

KEYWORDS

major depression; internet interventions; self-monitoring; sleep; circadian; chronobiology; chronotherapy; clinician assisted

Introduction

Patients with major depression exhibit various circadian rhythmic abnormalities. These include day-to-day instability, drifting of the sleep-wake cycle [1-3], diurnal change in mood [4], alteration in cortisol rhythm [5], and changes in the circadian pattern of gene expression and metabolism in the brain [6,7]. Furthermore, mood symptoms are closely linked to the timing of sleep, and sleep delay often causes mood deterioration and sleep phase advancement, generating an antidepressant effect [1,8,9]. The circadian system depends on the external zeitgebers (time cues) to maintain rhythmicity and entrainment with the external light-dark cycle. A low zeitgeber input can cause drifting of sleep and irregularities in the sleep-wake cycle and mood [1,10,11], and adequate zeitgeber input can reset and stabilize rhythms [12]. These mechanisms are described in detail in the published protocol of this study [13].

For patients with a major depressive episode, the transition from an inpatient psychiatric ward to an outpatient status poses a risk of relapse, readmission, suicide attempts, and suicide [1,14,15]. Therefore, various outpatient mental health services that receive patients shortly after discharge have been developed to manage this challenge [16].

To deliver timely help, researchers have made attempts to monitor patients' daily condition throughout this transition to catch signs of deterioration and provide immediate intervention. We previously tested an electronic system (Daybuilder [17]) using daily entering of mood, sleep, and activity data in a one-arm feasibility trial [1] in patients with a major depressive episode. In that study, participants logged data daily, covering a 4-week transition phase from inpatient to outpatient status. The logged data automatically generated a graphic display of all the variables to help participants understand the temporal patterns of the data. Clinicians telephoned the participants weekly to discuss compliance with the data entry and patterns in the graphical display. The results showed a readmission rate of 13% and that the sleep midpoint drifted 39 (SE 10) minutes later over the 4-week period (P<.001), and this drift was associated with a significant worsening of mood (P=.03). Another evident result of that study was that both mood and sleep were very unstable on a day-to-day basis. The usability of Daybuilder was very high, with a System Usability Score (SUS) of 86.2 (SD 9.7).

A newer randomized controlled trial in patients with major depression used the Monsenso system, a smartphone-based electronic monitoring system [17]. Using this system, participants and clinicians could monitor their daily mood and sleep after discharge. No effect on readmission was found, but participants in the active group experienced better recovery than those in the control group, as measured by the Recovery Assessment Scale [15].

It is debatable whether electronic monitoring per se has a positive or negative impact on patients with affective disorders [18,19]. For example, in patients with bipolar disorder, monitoring of negative mood might sustain the level of depressive symptoms; on the other hand, continuous symptom monitoring might also enable patients to acknowledge their condition in due time and receive timely help [15,20,21].

Relevant literature showed that technology-based self-help and minimal contact therapies have been proposed as effective and low-cost interventions for anxiety and mood disorders [22], and the Daybuilder electronic system is created with clinician support facilities built into the system.

This study aimed to combine a newly developed intervention, Circadian Reinforcement Therapy (CRT) [13], with data logging in an electronic system and clinician feedback. CRT intervention is described in detail in the published protocol of this study [13]. CRT strengthens zeitgeber inputs to the circadian system through social contact, physical activity, diet, and daylight exposure and provides sleep guidance, aiming to stabilize mood and prevent drifting of sleep timing. The intervention was delivered through specialized psychoeducation supported by an electronic monitoring system. This is the first study to test CRT intervention.

The electronic system used in this study is a slightly updated version of the Daybuilder electronic system used in our previous study [1]. The updated Daybuilder system was acquired by Monsenso in 2014 [17] and is therefore named Monsenso Daybuilder (MDB). There are only minor differences between the Daybuilder and MDB systems, including a graphical visualization of day-to-day variability and a more detailed description of depression parameters.

This study aimed to prevent the worsening of depression in patients with a major depressive episode after discharge from inpatient psychiatric wards [13]. We hypothesized that the CRT intervention would help patients recover faster and would reduce

the likelihood of relapse in the period after discharge. We did not expect the electronic system to act as a treatment modality, but rather as a medium to facilitate the new CRT intervention.

Methods

Study Design

This study was designed as a 4-week, randomized, controlled, single-blinded, parallel-group trial, with a balanced allocation rate (1:1) to either the CRT or the treatment-as-usual (TAU) group.

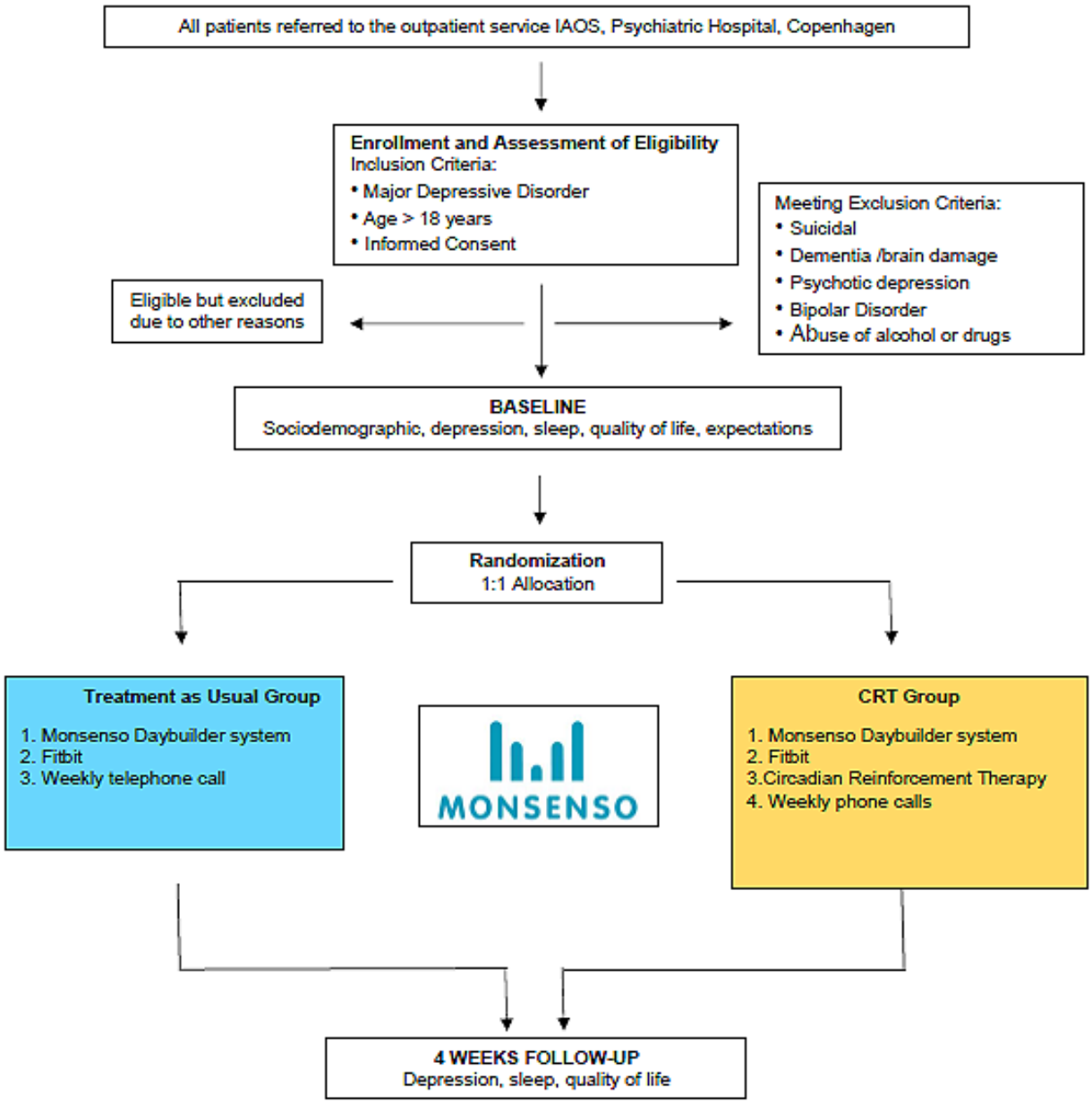
The procedure of this study was described in a published study protocol [13]. We developed a CRT based on the knowledge from chronobiology and chronotherapy of zeitgebers, light pathways in the brain and their relationship to mood [23], the phase response curve of light [12], and sleep regulation [24]. There is evidence that not only are rhythms such as sleep and melatonin and temperature curves distorted in major depression, but there are also changes in circadian gene expression in the brain [7]. The elements in CRT can be considered as an extension of the social metrics therapy developed by Frank et al [25] but with a focus on zeitgeber inputs. Thus, CRT focuses on the extent and timing of the inputs to the circadian system. Associated psychoeducation and written information aim to help patients understand the importance of the circadian system, including the need for zeitgeber inputs to sustain sleep and mood. The purpose of CRT is to stabilize sleep and mood, prevent sleep drift, and induce a slight sleep phase advance to help induce an antidepressant response [26-28].

Participants

Participants who aged ≥18 years and had a diagnosis of a major depressive episode, according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [29], were eligible for inclusion after signing informed consent. Exclusion criteria were a score of ≥2 on the Hamilton suicidal item 3, current abuse of alcohol or illegal substances, comorbid dementia or other brain disorders that would make the use of the MDB difficult, bipolar disorder, and psychotic depression.

All the patients who were referred from the psychiatric inpatient wards at the Mental Health Center Copenhagen for follow-up at the Intensive Affective Outpatient Service (IAOS) were consecutively asked to participate in the study. The IAOS exclusively receives patients with depression immediately after discharge from inpatient psychiatric wards (Figure 1). We aimed to include the patients when they were still in the inpatient ward or at the latest 3 weeks after commencing treatment at the IAOS.

Figure 1. Flowdiagram. IAOS: Intensive Affective Outpatient Service.



Interventions

Using a computer-generated random list without stratification, we randomized the patients to either the group using the CRT intervention as an add-on to the ongoing IAOS treatment or the group receiving only the IAOS treatment (TAU). Treatment at the IAOS included individual consultations with psychiatrists, supporting sessions with specially trained psychiatric nurses, group psychoeducation including some sleep hygiene advice (but no information on circadian-related components), physiotherapy, and pharmacological treatment.

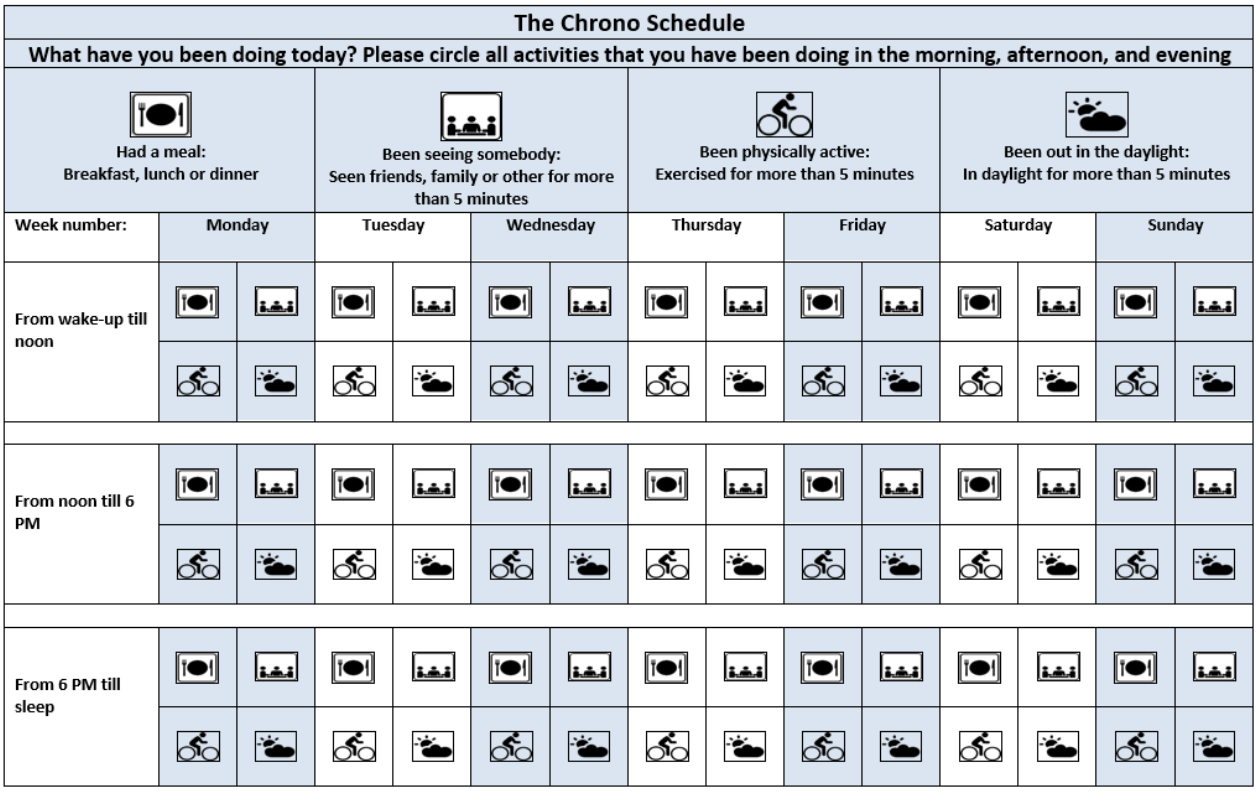
All participants in both the groups used the MDB self-monitoring system to assess daily mood, sleep, activity, and medication compliance [1,30]. In addition, all participants were given a wearable Fitbit activity tracker (model charge 2) to measure the number of steps and heart rate.

CRT Group

Participants allocated to the CRT group were given specialized psychoeducational sessions at the beginning of the trial period to facilitate the implementation of the CRT principles. These sessions were often given for 2 days to facilitate learning, with booster sessions at weekly phone calls. The sessions were scheduled as a 1-hour individual presentation, based on PowerPoint (Microsoft Corp) slides, where the principles of CRT were explained. The presentation was supplemented with paper-based informative material and leaflets. This specialized psychoeducation was intended to provide participants with knowledge on ways to strengthen time signals to their circadian system using temporally structured and higher levels of exposure to zeitgeber elements, such as daylight, exercise, regular meals, and social contact [11,31-33]. CRT psychoeducation also includes sleep hygiene guidance, such as advice on trying to avoid naps that is often associated with deterioration of mood, a phenomenon named “nap-mood-drop” [34]. If napping could

not be avoided, the advice was to nap <30 minutes to help increase the sleep pressure. Advice was also given regarding earlier sleep timing to induce an antidepressant effect. In addition, a paper *chrono diary* (Figure 2) was used to help the participants follow their own implementation of the instructed zeitgeber elements. The participants were directed to check the boxes in the chrono diary daily, indicating when they had engaged in various activities, such as consuming a meal, spending time with someone for >5 minutes, engaging in physical activity for >5 minutes, and being in sunlight for >5 minutes. This was performed at 3 distinct intervals during the day: from the time of awakening until noon, from noon until 6 PM, and from 6 PM until sleep onset [13]. Data from the chrono diary were not entered into the MDB system but were used in weekly phone calls. An individualized plan for future weeks was made for each participant in the CRT group, focusing on 3 self-selected elements (eg, awakening at 7 AM, having exposure to daylight, and exercising or walking). An action plan covering how to handle deterioration was also developed.

Figure 2. Chrono Schedule.



Once a week, an investigator phoned or met participants at the IAOS, and together, they examined the self-assessment data from the last week through the graphical display in the MDB system and discussed possible solutions using the elements in the CRT method. In addition, MDB data entries from participants in the CRT group were monitored by the investigators on all weekdays, and predetermined triggers elicited additional phone calls to participants. The trigger points were sudden mood drops, changes in sleep timing, reduced medication compliance, low data entry compliance, and negative text messages (Multimedia Appendix 1). CRT principles are described in more detail in the protocol [13], and the usability of the MDB system is described by Lauritsen et al [1].

CRT methods were used in all phone calls in the CRT group to help stabilize mood, sleep, and activity by recalling CRT elements that could strengthen zeitgeber signals. For example, an observed drift to late sleep onset would elicit advice on morning daylight exposure, morning exercise, avoiding late-night light exposure, late-night exercise, and late meals. Likewise, unstable sleep patterns would elicit advice on more regular sleep schedules and examine the underlying reasons for deviations (being out late at night, eating late, and ruminating when going to bed).

During the planned weekly phone calls, the investigator also evaluated and discussed (1) experience with the MDB system, (2) suicidal ideation, (3) side effects of medications, (4) worsening of depression and plan for action, and (5) compliance with data entry.

TAU Group

Patients in the TAU group also received weekly telephone calls, and these talks only covered the items (1) to (5) that are mentioned in the previous paragraph for the CRT group. See the published study protocol for more details [13].

Outcomes

Questionnaires and Self-Monitoring

The Danish version of the Mini International Neuropsychiatric Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition was used for diagnosis [35]. The

investigators were certified using the Mini International Neuropsychiatric Interview. Sociodemographic data were collected at the baseline. All participants were interviewed using the Hamilton Depression Rating Scale (HAM-D_{17}) [36] by a blinded investigator at baseline and end points. The following self-assessment scales were performed by the patient at baseline and the end point: Major Depression Inventory (MDI) [37] (score range 0-50, 50=severest depression), the World Health Organization Well-Being Index (WHO-5) [38] (score range 0-100, 100=best quality of life), and the Morningness-Eveningness Questionnaire (MEQ) [39] that assesses preferred time points for defined tasks as a measure of chronotype (with scores <42 indicating *evening type*, 42-58 indicating *intermediate type*, and >58 indicating *morning type*). The Pittsburgh Sleep Quality Index (PSQI) assesses sleep quality during the last month (a global score of ≥5 signifies poor sleep quality) [40]. The SUS scale (10 questions rated based on agreement with statements from 1 to 5 and converted to a score range from 0 to 100; 100=best) [1,41] was used at the end point to evaluate the MDB system. Participants in both the groups logged self-assessment data into the MDB system daily for a 4-week period. The procedure has been described in more detail in the protocol [13]. The following parameters were logged in the MDB system: depression levels assessed at the time of awakening "depression_morning," depression levels assessed just before bedtime "depression_evening," and depression levels assessed as a mean for the last day and entered before bedtime "depression_daily," activity as the number of steps in the last 24 hours from the Fitbit bracelet, daily medication adherence with categories of yes or no, sleep onset, sleep offset, number of awakenings, sleep quality, and daytime naps with time points for nap start and end. Mood and sleep quality were monitored using a scale, with response categories ranging from 0 to 10, where 10 indicates the best quality. Text comments, viewable to participants and investigators, could be made next to the data entries, for example, as additional information on sleep or mood problems. In the CRT group, a text comment option "sleep-wake cycle" was specifically designed so that the participants could describe what measures they used each day to stabilize the sleep-wake cycle. This was used on each phone call to evaluate the participants' use of CRT principles in the previous week.

The wearable Fitbit bracelet is a wrist-worn data logger with a display that counts the number of steps. The number of steps was entered into the MDB system each night by the participant, covering the last 24 hours. Pulse data from the Fitbit bracelet were read through a computer interface at the end of the visit.

At baseline, we administered an Expectation of Treatment Outcome scale, with a score range from 0-10, where 0 signifies unchanged levels of depression and 10 no depression). In addition, a short questionnaire concerning diurnal patterns of depression symptoms was administered. At the end point, all participants were asked questions to evaluate their general experience in the study period, and participants in the CRT group were also asked CRT-specific questions.

Biochemical Outcomes

A subset of participants performed repeated evening saliva sampling to assess the Dim Light Melatonin Onset (DLMO) at baseline and end point. Oral and written instructions on how to perform saliva sampling were provided. This included securing dim light exposure during the sampling period and avoiding tea, caffeine, cola, bananas, chocolate, smoking, and fluids with colorants from 1 PM to the end of the sampling period. In addition, participants were advised not to brush their teeth during the sampling period and not to use drugs containing aspirin or ibuprofen for the entire sampling day. Saliva samples were taken at home beginning 5 hours before expected bedtime and thereafter every half hour, with the last sample collected 1 hour after expected bedtime, summing up to 13 samples. Instructions were provided by KD. Samples were frozen at −20 °C until analysis. The sensitivity of melatonin assay is highly variable. The IBL RE54041 enzyme-linked immunosorbent assay used in this study was considered valid [42].

DLMO was determined using the hockey-stick method [43]. The phase angle difference (PAD) [44] was calculated as the difference between the mean of the last 3 days of midpoint sleep and DLMO for each participant. The sleep midpoint was calculated from the entered sleep onset and offset data in the MDB system.

Ranking of Outcomes

The primary outcome was defined as the difference between the 2 groups in terms of changes in HAM-D_{17} scores from baseline to end point. The secondary outcome was defined as the difference between the 2 groups in terms of changes in the MDI scores from baseline to end point. The tertiary outcome was defined as the difference in the PAD magnitude between the groups at the end point. Explorative outcomes were defined as the difference between the estimated end point scores of the self-monitored data from the MDB system, activity data from the Fitbit, and the WHO-5 scale scores of the CRT group and those of the TAU group. The MEQ and PSQI scores were used as descriptive data and covariates in the models.

Statistics

The intention-to-treat method was applied using all data for all included participants in all analyses. Interviewer-based and self-reported paper-and-pencil data with baseline and end point assessments were analyzed using a general linear model (GLM), with end point scores as the outcome. Data from the MDB system were analyzed using a mixed model for repeated measures, using all available data without imputation. In the GLM and mixed model, covariates were removed by backward selection, depending on the magnitude of the parameter estimate and significance. The level of significance was set at a P value of .05. Continuous baseline characteristics were compared using a 2-tailed t test or Wilcoxon 2-sample test, depending on the data distribution. Categorical data were analyzed using the chi-square test.

For analyses of day-to-day variability of sleep data, the data were truncated, excluding periods of sleep onset before 6 PM and sleep offset after 6 PM.

In the GLM model and the mixed model for repeated measures, the following covariates were used: baseline values of the outcome, the treatment group, interaction term between day and the 2 treatment groups, time from discharge to visit 1 (TDV1),

interaction term between TDV1 and the treatment groups, electroconvulsive treatment (ECT; yes or no), interaction term between ECT and the treatment groups, medication adherence (yes or no), and MEQ scores.

The covariates were chosen based on the assumption of the influence on the respective outcomes in the models: ECT was selected because the resultant cognitive disturbances might make CRT psychoeducation less effective [45] and TDV1 as it was the primary assumption in the study that CRT should be implemented as a preventive measure as early as possible after discharge to influence the depression course. Medication compliance was included because of its possible effect on depression levels [46], while MEQ was included, as this is assumed to be linked to sleep timing and thus mood [47].

Day-to-day variability of the MDB data was estimated as residuals (covariance parameter estimates) for each group in a repeated-measures model. Statistical significance was calculated from the difference in log-likelihood values in a random versus repeated model as a chi-square value with 1 degree of freedom.

Ethical Considerations

The study was approved by the Regional Committee on Health Research Ethics (H-15013943) and Danish Data Protection Agency (2012-58-0004/RHP-2016-015 I-suite number 04734). All participants provided written informed consent before enrollment in the study. Care was taken to evaluate whether the participants found the study procedures stressful. Daily self-monitoring can cause stress and inconvenience, and participants could withdraw from the study at any time.

Results

Sociodemographic and Baseline

A total of 342 patients were screened for participation, and 103 gave informed consent to participate in the study from September 2016 to November 2020, with 51 participants randomized to the CRT group and 52 participants in the TAU group. One participant withdrew informed consent before enrollment, and one participant's inclusion was canceled because of the decision to extend the inpatient stay. Both the participants were randomized into the CRT group. No outcome data were collected for these 2 participants. Discharge date was missing for 3 participants. Overall, 29% (29/98) of the participants were included on or before the day of discharge, 56% (55/98) on or before day 8 after discharge, 60% (59/98) on or before day 15 after discharge, 71% (70/98) on or before day 22 after discharge, and 78% (76/98) on or before day 29 after discharge. Sociodemographic data in Table 1 show that the parameters were evenly distributed between the 2 groups except for the Expectation of Treatment Outcome scale score that was significantly higher in the CRT group than in the TAU group (CRT: mean 5.9, SD 2.5; TAU: mean 4.7, SD 2.5; *P*=.03).

Of the remaining 101 participants, 15 (14.9%) did not attend the end point assessment visit (9 participants in the CRT group and 6 in the TAU group) and could therefore not be included in the analyses based only on the questionnaires administered at baseline and end point visits. However, their MDB data were used for exploratory analyses. The reasons for missing the end point assessment in the CRT group were given as "feeling participation too stressful or burdensome" (n=4) and "deterioration of depression with follow-up not possible" (n=4), and 1 participant was lost to follow-up; and the reasons in the TAU group include "deterioration of depression with follow-up not possible" (n=5) and "end point assessment delayed for practical reasons" (4 months; n=1).

A total of 8 (8%) participants were readmitted to the inpatient psychiatric ward during the study period, with 3 (6%) in the CRT group and 5 (10%) in the TAU group. These participants continued their study procedures when possible and were assessed at the end point. Two participants, both from the TAU group, took a nonfatal overdose due to suicidal ideation (one took venlafaxine plus an antihypertensive drug and one oxazepam). One of these participants dropped out of the study, and the other completed the study. Both the incidents were reported to the regulatory authorities. Score data from 7 Hamilton interviews were lost when they were sent between the 2 departments. Thus, for the analysis of the primary HAM-D_{17} outcome, data from 79 participants were available, of whom 35 were allocated to the CRT group and 44 to the TAU group. For the analyses of the secondary outcome on the MDI scale, data from 81 participants were available for analyses, of whom 39 were in the CRT group and 42 in the TAU group. For the explorative self-monitored data in the MDB system, 45 participants in the CRT group and 50 in the TAU group logged the data for one or more days. Compliance with the data logged into the MDB system was 90.7% (2565/2828) for the depression_daily measure (of the 101 eligible participants), 82.49% (2333/2828) for activity data, and 97.81% (2766/2828) for sleep onset and offset. The questionnaire examining the diurnal patterns of depression symptoms showed that 89.8% (44/49) in the CRT group and 73.1% (38/52) in the TAU group experienced a characteristic change in depression severity during the day. "Best-in-the-morning" pattern was experienced in 14.3% (6/42) and 21.2% (7/33), "best-at-noon" in 31% (13/42) and 30.3% (10/33), and "best-at-evening" in 54.8% (23/42) and 48.5% (16/33) in the CRT and TAU groups, respectively (between groups, *P*=.72).

Table 1. Sociodemographics.

Parameter	CRT[a] group (n=50)	TAU[b] group (n=52)
Age (years), mean (SD; range)	41.8 (14.8; 18-70)	40.6 (14.8; 18-74)
Gender (women), n (%)	31 (62)	27 (51)
Duration of current episode (month)		
Values, mean (SD)	11.3 (15.7)	15.1 (19.9)
Values, median (IQR; range)	6.0 (4.0-11.0; 2-78)	7.0 (3.0-16.5; 1-89)
Recurrent episodes, n (%)	26 (52)	30 (58)
Number of previous episodes		
Values, mean (SD)	2.6 (5.1)	1.5 (1.8)
Values, median (IQR)	1.0 (0-3.0)	1.0 (0-2.5)
Suicide attempt in connection with actual admittance to inpatient ward, n (%)	10 (20)	10 (19)
One or more previous suicide attempt, n (%)	15 (30)	11 (21)
Patients experiencing an eliciting factor for actual episode, n (%)	45 (90)	47 (90)
Smoking, n (%)	16 (32)	21 (40)
Number of cigarettes per day, mean (SD)	12.1 (6.2)	14.2 (11.3)
ECT[c] in current episode, n (%)	17 (34)	10 (19)
Number of ECT treatments in current episode, mean (SD; range)	14.6 (6.5; 4-26)	11.8 (3.2; 8-17)
Drinking alcohol, n (%)	29 (58)	28 (54)
Alcohol intake more than once per week, n (%)	6 (12)	10 (19)
Expectation of Treatment Outcome (0-10, with 0=unchanged and 10=no depression), mean (SD)	5.9 (2.5)[d]	4.7 (2.5)

[a]CRT: circadian reinforcement therapy.

[b]TAU: treatment as usual.

[c]ECT: electroconvulsive treatment.

[d]*P*=.03.

Analysis of the Primary Outcome

The mean HAM-D_{17} scores at inclusion were 15.8 (SE 0.9) and 15.5 (SE 0.8) for the CRT and TAU groups, respectively, ranging from 2 to 31. Analysis of the HAM-D_{17} end point as the primary outcome with the described covariates (see the *Statistics* section) showed no significant effects of ECT or MEQ scores, and these covariates were removed. The final model showed a significantly larger reduction in HAM-D_{17} scores in the CRT group than in the TAU group (F_1-score=4.2; parameter estimate=2.6; *P*=.04). Furthermore, there was a statistically significant interaction between TDV1 and the groups (F_1-score=8.5; parameter estimate=−0.08; *P*=.005). For participants included at discharge (TDV1=0), the parameter output thus showed an estimated difference in end point HAM-D_{17} scores of 2.6 points, with a gradually diminishing effect of −0.08 HAM-D_{17} points per day. This shows that earlier implementation of CRT intervention was associated with a greater antidepressant end point difference between the groups. The estimated end point HAM-D_{17} values for the CRT and TAU groups were 12.8 (SE 0.7) and 13.6 (SE 0.6; Table 2), respectively. The HAM-D_{17} baseline used in the estimation model is the whole sample mean 15.6 (SE 0.6).

The model fit for the primary outcome showed 2 observations with a high Cook distance, suggesting influential data (participants 35 and 45). Sensitivity analyses with the removal of these 2 observations showed a continued statistically significant effect (*P*=.04), and the observations were included in the analysis. The baseline HAM-D_{17} score for the 15 participants who did not attend the end point assessment was 16.8 (SD 6.5) compared with 15.4 (SD 6.0) for those who did.

Table 2. Results from interviewer and self-assessment scales.

Instrument and visit	CRT[a] group	TAU[b] group
Outcome measures		
HAM-D_{17}[c] (n=79), mean (SE)		
Whole sample baseline	15.6 (0.6)	15.6 (0.6)
Estimated end point	12.8 (0.7)[d]	13.6 (0.6)
MDI[e] (n=81), mean (SE)		
Whole sample baseline	21.5 (1.2)	21.5 (1.2)
Estimated end point	18.4 (1.1)	20.6 (1.1)
WHO-5[f] (n=81), mean (SE)		
Whole sample baseline	40.9 (2.5)	40.9 (2.5)
Estimated end point	46.3 (2.6)	43.4 (2.2)
Chronotype and sleep quality		
MEQ[g] (n=81), mean (SD)		
Baseline	49.8 (10.9)	51.2 (11.9)
End point	52.3 (9.8)	52.1 (11.3)
PSQI[h] (n=84), mean (SD)		
Baseline	8.8 (3.6)	10.2 (4.7)
End point	7.7 (3.0)	9.3 (4.3)

[a]CRT: circadian reinforcement therapy.

[b]TAU: treatment as usual.

[c]HAM-D_{17}: Hamilton Depression Rating Scale.

[d]*P*=.04.

[e]MDI: Major Depression Inventory.

[f]WHO-5: World Health Organization Well-Being Index

[g]MEQ: Morningness-Eveningness Questionnaire.

[h]PSQI: Pittsburgh Sleep Quality Index.

Analyses of the Secondary Outcome

The mean MDI scores at inclusion were 21.7 (SE 1.7) and 21.3 (SE 1.6) for the CRT and TAU groups, respectively. Analyses showed no statistically significant effect of any covariate on MDI scores. The estimated end point MDI values for the CRT and TAU groups were 18.4 (SE 1.1) and 20.6 (SE 1.1; Table 3), respectively. The MDI baseline used in the estimation model is the whole sample mean 21.5 (SE 1.2).

Table 3. Results from the self-monitored parameters in the Monsenso DayBuilder system.

Parameter and visit	CRT[a] group	TAU[b] group
Depression_daily (all data), mean (SE)		
Whole sample baseline	5.6 (0.2)	5.6 (0.2)
End point	7.7 (0.4)	7.6 (0.4)
Depression_morning (all data), mean (SE)		
Whole sample baseline	5.6 (0-3)	5.6 (0-3)
Estimated end point	6.6 (0.4)	6.8 (0.4)
Depression_evening (all data; medication adherence: yes), mean (SE)		
Whole sample baseline	6.1 (0.3)	6.1 (0.3)
Estimated end point	7.2 (0.2)[c]	6.6 (0.2)
Depression_evening (all data; medication adherence: no), mean (SE)		
Whole sample baseline	6.1 (0.3)	6.1 (0.3)
Estimated end point	8.3 (0.4)[d]	7.7 (0.4)
Sleep quality (all data), mean (SE)		
Whole sample baseline	6.2 (0.2)	6.2 (0.2)
Estimated end point	7.0 (0.2)[e]	6.5 (0.2)
Sleep onset (all data), mean hour:minutes (SE)		
Whole sample baseline	23:26 (0:09)	23:26 (0:09)
Estimated end point	23:17 (0:09)[f]	23:42 (0:09)
Sleep offset (all data), mean hour:minutes (SE)		
Whole sample baseline	7:23 (0:08)	7:23 (0:08)
Estimated end point	7:50 (0:08)	7:47 (0:08)
Sleep duration (all data), mean decimal hours (SE)		
Whole sample baseline	7.9 (0.1)	7.9 (0.1)
Estimated end point	8.4 (0.1)[g]	7.9 (0.1)
Sleep wakeups (all data), numbers (SE)		
Whole sample baseline	1.5 (0.2)	1.5 (0.2)
Estimated end point	2.3 (0.5)	2.5 (0.6)
Steps from Fitbit (all data), number (SE)		
Whole sample baseline	5857 (479)	5857 (479)
Estimated end point	9195 (764)	9151 (737)

[a]CRT: circadian reinforcement therapy.

[b]TAU: treatment as usual.

[c]P=.02.

[d]P=.04.

[e]P=.04.

[f]P<.001.

[g]P=.005.

Analyses of the Tertiary Outcome

Melatonin assessments were performed only in 12 participants (5 in the CRT group and 7 in the TAU group, with 1 missed assessment at the end point in each group) due to COVID-19–related restrictions. For participants with both baseline and end point DLMO assessments, the mean DLMO at baseline and end point was 19:40 (SD 1:10) and 19:27 (SD 0:46) hour:minutes in the CRT group (n=4) and 19:38 (SD=1:06) and 20:01 (SD=0:56) hour:minutes in the TAU group (n=6) (P=.30).

For participants with DLMO assessment at both baseline and end points, we analyzed the first 3 days of sleep midpoint during baseline assessment and the last 3 days of sleep midpoint from end point assessment. The analyzes showed a baseline and end point PAD of 7.40 (SD 0.28) and 7.33 (SD 0.53) hours, respectively, in the CRT group and 7.50 (SD 1.20) and 7.03 (SD 1.13) hours, respectively, in the TAU group (*P*=.63).

Analyses of Explorative Outcomes (Pen-and-Paper Data)

The mean WHO-5 score at inclusion was 40.6 (SE 4.1) and 37.4 (SE 3.3) in the CRT and TAU groups, respectively. We found no statistically significant effect of any covariate on the WHO-5 score. The baseline WHO-5 score used in the estimation model is the whole sample mean of 40.9 (SE 2.5).

The distribution of baseline MEQ chronotypes was 20% (20/98) evening, 57% (56/98) intermediary, and 22% (22/98) morning types for the 2 groups combined. The mean baseline PSQI score was high, indicating poor sleep quality (Table 3).

Analyses of Explorative Outcomes With Daily Self-Monitored Parameters From the MDB System

Results from the MDB system are shown in Table 3.

Analyses of the depression_daily scores (baseline values, CRT group: 5.7, SE 0.4; TAU group: 5.5, SE 0.4) showed that the only significant covariate was medication adherence (F_1-score=5.5; parameter estimate 0.83; *P*=.02), indicating an association between higher depression_daily scores and participants who reported not taking their medication.

For depression_morning depression scores (baseline values, CRT group: 5.5, SE 0.4; TAU group: 5.7, SE 0.3), the only significant covariate was MEQ (F_1-score=7.6; parameter estimate=0.03; *P*=.007), indicating an association between higher (better) depression_morning scores and the morning chronotype.

For the depression_evening scores (baseline values, CRT group: 6.0, SE 0.4; TAU group: 6.3, SE 0.4), we found a significantly higher estimated end point score in the CRT group than in the TAU group (F_1-score=5.7; parameter estimate=0.62; *P*=.02) and a significant effect of medication (F_1-score=9.6; parameter estimate 1.1; *P*=.002), indicating a better mood for participants who reported that they did not take their medication.

Analyses of the difference between depression_evening and depression_morning scores showed significantly higher estimated scores (better) in the evening (F_1-score=14.3; parameter estimate=0.36; *P*=.003) as an indication of a diurnal variation.

Analysis of sleep quality showed a significantly larger estimated end point score (baseline values, CRT group: 6.1, SE 0.4; TAU group: 6.3, SE 0.3) in the CRT group than in the TAU group (F_1-score=4.6; parameter estimate=0.51; *P*=.04) and a significant effect of MEQ (F_1-score=9.9; parameter estimate 0.03; *P*=.002), indicating better sleep quality for patients with a higher MEQ (morning type).

Analysis of sleep onset showed that the estimated end point sleep onset (baseline values, CRT group: 23:28, SE 0:12; TAU group: 23:25, SE 0:12 hours:minutes) was significantly earlier in the CRT group than in the TAU group (F_1-score=7.2; parameter estimate=26.6 min; *P*=.009). The MEQ covariate was significant (F_1-score=21.9; parameter estimate=−0.03; *P*<.001), indicating an association between the morning type and earlier sleep onset.

For sleep offset (baseline values, CRT group: 7:23, SE 0:13; TAU group: 7:23, SE 0:11 hours:minutes), the only significant covariate was MEQ (F_1-score=48.8; parameter estimate=−0.05; *P*<.001), indicating an earlier sleep offset for morning types.

The estimated end point sleep duration (baseline values, CRT group: 7.9, SE 0.18; TAU group: 8.0, SE 0.22) was significantly longer in the CRT group than in the TAU group (F_1-score=8.4; parameter estimate=0.48 [decimal hours]; *P*=.005).

There was no significant difference in the estimated end points between the groups for the number of awakenings (baseline values, CRT group: 1.5, SE 0.2; TAU group: 1.4, SE 0.2).

The estimated mean frequency of naps per day was 15.9% (confidence level [CL] 13.8%-18%) in the CRT group and 21.3% (CL 19.1%-23.4%) in the TAU group (χ^2_1=12.3; *P*<.001). Estimated mean nap duration was 1.60 (CL 1.0-2.6) hours in the CRT group and 1.77 (CL 1.2-2.7) hours in the TAU group (*P*=.75). The estimated mean timing of naps was 15:08 (CL 14:03-15:13) in the CRT group and 13:18 (CL 12:22-14:13) hours:minutes in the TAU group (*P*=.01).

The baseline values of the steps registered from the Fitbit bracelet were 5221 (SE 630) and 6514 (SE 715) steps in the CRT and TAU groups, respectively. We found no difference in the estimated mean end point number of daily reported steps between the 2 groups, with 9195 (SE 764) and 9151 (SE 737) steps in the CRT and TAU groups, respectively, but showed a significant increase in activity from baseline. In the model, the MEQ was significant (F_1-score=8.2; parameter estimate=97.8; *P*=.005). This indicates a higher level of activity in the morning type (higher MEQ scores).

When comparing the day-to-day variability of MDB data between the CRT and the TAU groups, we found significantly lower values for depression_daily scores with 1.22/1.56 (*P*<.001), no difference between groups for depression_morning scores with 1.42/1.49, significantly lower values for depression_evening scores with 1.45/1.94 (*P*<.001), significantly lower values for sleep onset with 0.85/1.24 (*P*<.001), significantly lower values for sleep offset with 1.03/1.45 (*P*<.001), significantly lower values for sleep quality scores with 1.69/2.15, and significantly lower values for the number of steps with 10,674,014/14,054,624 (*P*<.001).

The mean SUS score was 62.8 (SD 15.2) and 63.3 (SD 16.6) in the CRT and TAU groups, respectively (*P*=.95). The items with the lowest scores were item 4 (“I think that I would need the support of a technical person to be able to use this system”), with a mean score of 3.8 (SD 1.5), and item 7 (“I would imagine that most people would learn to use this system very quickly”), with a score of only 1.9 (SD 1.6).

Triggered extra phone calls were performed in 46% (23/50) of the participants in the CRT group, with a mean number of 1.8 (SD 1.1) calls in the 4-week period per participant. The rules for entries in the MDB that triggered a phone call and the percentage of participants eliciting a specific trigger by MDB entry are shown in Multimedia Appendix 1.

In total, 852 text messages were logged by the participants in the CRT group in the "sleep-wake cycle" text option in the MDB, corresponding to 70.18% (852/1214) of the days. The content of these comments included the CRT elements that the participants had used to strengthen zeitgebers and stabilize the sleep-wake cycle: "went for a walk in the sun," "doing training," "visiting friends," and "keep regular eating hours." These comments were discussed during the weekly phone calls. The daily compliance with medication was 99.8% (1369/1372) in the CRT group and 96% (1398/1456) in the TAU group. The results of the end point evaluation questions are shown in Multimedia Appendix 2. The text answers were categorized as positive, negative, neutral, or inconclusive. The results showed a significantly higher percentage of positive evaluation to the question, "What has your participation in the project meant for handling your depression after discharge?" in the CRT group than in the TAU group (CRT group: 36/39, 93% vs TAU group: 27/43, 63%; $P<.001$). With regard to the CRT-specific questions, there was a high percentage of positive statements for all questions, with 92% (36/39) recommending the use of the MDB system.

Discussion

Principal Findings

We found a significantly larger reduction in the end point Hamilton score, better evening mood, earlier sleep onset, longer sleep duration, better sleep quality, fewer naps, and lower day-to-day variation in sleep and mood in the CRT group than in the TAU group. Furthermore, a statistically significant interaction was found between TDV1 and the groups (CRT and TAU): the longer the TDV1, the lesser the difference between the Hamilton scores in the 2 groups. This supports the intention of this study to include participants as early as possible. Finally, the MEQ scores showed that the morning types had better sleep quality, earlier sleep onset, and sleep offset.

The observed phase advance of sleep in the CRT group compared with the TAU group and the reduced day-to-day variation of sleep were treatment goals and expected through the known impact of zeitgebers on the circadian system, including the sleep-wake cycle. Fewer naps in the CRT group were expected because of psychoeducational advice to limit naps. The significantly earlier timing of naps in the TAU group could be caused by lower sleep quality and shorter sleep duration in this group, resulting in more daytime sleepiness and thus an earlier need for naps.

MEQ associations with sleep timing and quality were also expected, whereas the influence of the morning type on activity is probably a new finding.

Regarding day-to-day variability, our results are in line with our earlier results using sleep time stabilization and bright light, both of which also enhance zeitgeber strength [2].

The intended early inclusion of patients was difficult, with only 56% (55/98) of patients included 8 days after discharge. This delayed inclusion was caused not only by the COVID-19–related restrictions, which made visits to the inpatient wards impossible or difficult but also by the earlier discharge dates than expected in many cases. Thus, in future studies in this field, the implementation of CRT should preferably use a research coordinator located permanently in the inpatient ward to include patients as early as possible and provide psychoeducation sessions before discharge. This might be done in a group session to optimize the use of resources and facilitate group dynamics in learning. If implemented, CRT could also be incorporated into standard psychoeducation in the ward and thus be provided before discharge.

Another finding in the study was that the participants at discharge still had a significant level of depression, as well as cognitive dysfunction, probably making it difficult for them to acquire new information [48]. This was also reflected in the low-to-moderate SUS score, with low scores on the 2 items evaluating learning and support.

The presence of cognitive dysfunction makes it important to refine both psychoeducation and electronic monitoring systems. Compliance with the MDB system was acceptable but might still have caused some frustration.

The structured questions in the end point evaluation showed that the patients felt that the CRT intervention was helpful; thus, the results from the study show that CRT could be used in the management of depression. However, a subset of patients also gave a negative statement regarding the management of the MDB system, and 4 patients dropped out because they felt that it was stressful or burdensome to participate in the CRT group.

Overall, 45% (22/49) of the participants in the CRT group were phoned once or more between weekly scheduled calls due to trigger alerts in the MDB system. This shows that the trigger systems worked, and the evaluation questions regarding phone calls showed that participants appreciated this feature. This might have had a positive influence on the outcomes of the CRT groups and should be included when implementing the system.

Refinement of CRT methods could include video presentations, podcasts, group sessions, peer involvement, and an electronic monitoring system designed as a smartphone app. CRT should be co-designed with users within the framework of a complex intervention [49,50].

Knowledge of how nonphotic zeitgebers influence the circadian system, sleep, and mood work and how they should be administered is still very rudimentary and should be incorporated into CRT refinement. For example, different types of exercise (eg, aerobic or strength training) and their timing and the nutritional components in diet and timing of diet will most likely have different impacts on the circadian system, sleep, and mood [51].

One of the challenges in conducting randomized controlled trials in this patient group is obtaining sufficient, preferably daily, and precise data from the participants without fatiguing them too much. The development of wearable devices is a possible solution for collecting certain data types. We still await an easy-to-use and dependable method to measure sleep and circadian rhythms. The current candidates are ear electroencephalography that are currently being tested [52], 24-hour thermometer measurements [53], and actigraphy with associated light measurement [54].

Limitations

The sample size was smaller than planned, resulting in less robust findings and limited generalizability. Participants could not be blinded to their allocation to the CRT and the TAU groups, which might have induced an unknown bias in the outcomes. Information on physical comorbidity and types and dosages of medications was not collected, and we cannot rule out an imbalance of these characteristics between groups that could have induced an unknown effect on some of the outcomes. The COVID-19–related restrictions caused difficulty in performing the procedures, and we believe that this might have impaired the implementation of the CRT method. The MDB system might have caused some inconvenience for some patients; nevertheless, their end point evaluation of the system was positive. The 15 patients who did not attend the end point assessment did not deviate substantially in their baseline HAM-D_{17} scores compared with those that were evaluated at the end point, indicating that the sample available for analyses was probably not biased because of these dropouts. We did not have any data on daylight exposure, which would have provided an indication of the signal to the circadian system in the 2 groups, and whether it differed. In future studies, an activity tracker with a light sensor can be used.

Conclusions

CRT delivered in combination with the MDB system showed a moderate add-on antidepressant effect, the magnitude of which was dependent on the time of inclusion after discharge. Furthermore, we found a sleep phase advancing effect, improvement in sleep quality and evening mood, and a significant reduction in day-to-day variability of mood and sleep. The structured questions from the end point evaluations supported these findings. However, the usability of the MDB system was low for 2 specific items and should be improved. There is thus room for improvement with better electronic systems, timely inclusion, and updated CRT content.

Acknowledgments

The study was funded by TrygFonden (grant number 107961) and Grosserer LF Foghts Fond (grant number 21.597).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Rules for entries in the Monsenso-Daybuilder system that triggered a phone call to participants in the circadian reinforcement therapy group, with frequency.
[PDF File (Adobe PDF File), 413 KB - mental_v10i1e50072_app1.pdf]

Multimedia Appendix 2
User evaluation from semistructured questions at end point.
[PDF File (Adobe PDF File), 523 KB - mental_v10i1e50072_app2.pdf]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 8673 KB - mental_v10i1e50072_app3.pdf]

References

1. Lauritsen L, Andersen L, Olsson E, Søndergaard SR, Nørregaard LB, Løventoft PK, et al. Usability, acceptability, and adherence to an electronic self-monitoring system in patients with major depression discharged from inpatient wards. J Med Internet Res 2017 Apr 21;19(4):e123 [FREE Full text] [doi: 10.2196/jmir.6673] [Medline: 28432040]
2. Martiny K, Refsgaard E, Lund V, Lunde M, Sørensen L, Thougaard B, et al. A 9-week randomized trial comparing a chronotherapeutic intervention (wake and light therapy) to exercise in major depressive disorder patients treated with duloxetine. J Clin Psychiatry 2012 Sep;73(9):1234-1242. [doi: 10.4088/JCP.11m07625] [Medline: 23059149]
3. Sato S, Bunney B, Mendoza-Viveros L, Bunney W, Borrelli E, Sassone-Corsi P, et al. Rapid-acting antidepressants and the circadian clock. Neuropsychopharmacology 2022 Mar;47(4):805-816 [FREE Full text] [doi: 10.1038/s41386-021-01241-w] [Medline: 34837078]

4. Matheson GJ, Schain M, Almeida R, Lundberg J, Cselényi Z, Borg J, et al. Diurnal and seasonal variation of the brain serotonin system in healthy male subjects. Neuroimage 2015 May 15;112:225-231. [doi: 10.1016/j.neuroimage.2015.03.007] [Medline: 25772667]
5. Refsgaard E, Schmedes AV, Martiny K. Salivary cortisol awakening response as a predictor for depression severity in adult patients with a major depressive episode performing a daily exercise program. Neuropsychobiology 2022;81(3):246-256. [doi: 10.1159/000521234] [Medline: 35016170]
6. Germain A, Nofzinger EA, Meltzer CC, Wood A, Kupfer DJ, Moore RY, et al. Diurnal variation in regional brain glucose metabolism in depression. Biol Psychiatry 2007 Sep 01;62(5):438-445 [FREE Full text] [doi: 10.1016/j.biopsych.2006.09.043] [Medline: 17217926]
7. Li JZ, Bunney BG, Meng F, Hagenauer MH, Walsh DM, Vawter MP, et al. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. Proc Natl Acad Sci U S A 2013 Jun 11;110(24):9950-9955 [FREE Full text] [doi: 10.1073/pnas.1305814110] [Medline: 23671070]
8. D'Agostino A, Ferrara P, Terzoni S, Ostinelli EG, Carrara C, Prunas C, et al. Efficacy of triple chronotherapy in unipolar and bipolar depression: a systematic review of the available evidence. J Affect Disord 2020 Nov 01;276:297-304 [FREE Full text] [doi: 10.1016/j.jad.2020.07.026] [Medline: 32697712]
9. Wirz-Justice A, Benedetti F. Perspectives in affective disorders: clocks and sleep. Eur J Neurosci 2020 Jan;51(1):346-365. [doi: 10.1111/ejn.14362] [Medline: 30702783]
10. Czeisler CA, Duffy JF, Shanahan TL, Brown EN, Mitchell JF, Rimmer DW, et al. Stability, precision, and near-24-hour period of the human circadian pacemaker. Science 1999 Jun 25;284(5423):2177-2181. [doi: 10.1126/science.284.5423.2177] [Medline: 10381883]
11. Czeisler CA, Gooley JJ. Sleep and circadian rhythms in humans. Cold Spring Harb Symp Quant Biol 2007;72:579-597. [doi: 10.1101/sqb.2007.72.064] [Medline: 18419318]
12. Khalsa SB, Jewett ME, Cajochen C, Czeisler CA. A phase response curve to single bright light pulses in human subjects. J Physiol 2003 Jun 15;549(Pt 3):945-952 [FREE Full text] [doi: 10.1113/jphysiol.2003.040477] [Medline: 12717008]
13. Dunker Svendsen S, Aggestrup AS, Nørregaard LB, Løventoft P, Præstegaard A, Danilenko KV, et al. Circadian reinforcement therapy in combination with electronic self-monitoring to facilitate a safe post-discharge period of patients with depression by stabilizing sleep: protocol of a randomized controlled trial. BMC Psychiatry 2019 Apr 25;19(1):124 [FREE Full text] [doi: 10.1186/s12888-019-2101-z] [Medline: 31023274]
14. Qin P, Nordentoft M. Suicide risk in relation to psychiatric hospitalization: evidence based on longitudinal registers. Arch Gen Psychiatry 2005 Apr;62(4):427-432. [doi: 10.1001/archpsyc.62.4.427] [Medline: 15809410]
15. Tønning ML, Faurholt-Jepsen M, Frost M, Martiny K, Tuxen N, Rosenberg N, et al. The effect of smartphone-based monitoring and treatment on the rate and duration of psychiatric readmission in patients with unipolar depressive disorder: the RADMIS randomized controlled trial. J Affect Disord 2021 Mar 01;282:354-363. [doi: 10.1016/j.jad.2020.12.141] [Medline: 33421863]
16. Nielsen CM, Hjorthøj C, Nordentoft M, Christensen U. A qualitative study on the implementation of flexible assertive community treatment - an integrated community-based treatment model for patients with severe mental illness. Int J Integr Care 2021 Apr 29;21(2):13 [FREE Full text] [doi: 10.5334/ijic.5540] [Medline: 33981190]
17. Monsenso A/S Carlsberg Vej 80. Monsenso. 2023. URL: https://www.monsenso.com/solution-privacy-policy/ [accessed 2023-11-07]
18. Tønning ML, Kessing LV, Bardram JE, Faurholt-Jepsen M. Methodological challenges in randomized controlled trials on smartphone-based treatment in psychiatry: systematic review. J Med Internet Res 2019 Oct 27;21(10):e15362 [FREE Full text] [doi: 10.2196/15362] [Medline: 31663859]
19. Anmella G, Faurholt-Jepsen M, Hidalgo-Mazzei D, Radua J, Passos IC, Kapczinski F, et al. Smartphone-based interventions in bipolar disorder: systematic review and meta-analyses of efficacy. A position paper from the International Society for Bipolar Disorders (ISBD) Big Data Task Force. Bipolar Disord 2022 Sep;24(6):580-614 [FREE Full text] [doi: 10.1111/bdi.13243] [Medline: 35839276]
20. Faurholt-Jepsen M, Frost M, Christensen EM, Bardram JE, Vinberg M, Kessing LV. The effect of smartphone-based monitoring on illness activity in bipolar disorder: the MONARCA II randomized controlled single-blinded trial. Psychol Med 2020 Apr;50(5):838-848. [doi: 10.1017/S0033291719000710] [Medline: 30944054]
21. Faurholt-Jepsen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, et al. Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. Psychol Med 2015 Oct;45(13):2691-2704. [doi: 10.1017/S0033291715000410] [Medline: 26220802]
22. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):89-103. [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
23. Fernandez DC, Fogerson PM, Lazzerini Ospri L, Thomsen MB, Layne RM, Severin D, et al. Light affects mood and learning through distinct retina-brain pathways. Cell 2018 Sep 20;175(1):71-84.e18 [FREE Full text] [doi: 10.1016/j.cell.2018.08.004] [Medline: 30173913]

24. Borbély A. The two-process model of sleep regulation: beginnings and outlook. J Sleep Res 2022 Aug;31(4):e13598 [FREE Full text] [doi: 10.1111/jsr.13598] [Medline: 35502706]
25. Frank E, Hlastala S, Ritenour A, Houck P, Tu XM, Monk TH, et al. Inducing lifestyle regularity in recovering bipolar disorder patients: results from the maintenance therapies in bipolar disorder protocol. Biol Psychiatry 1997 Jun 15;41(12):1165-1173. [doi: 10.1016/s0006-3223(96)00241-7] [Medline: 9171907]
26. Lee HJ. Is advancing circadian rhythm the mechanism of antidepressants? Psychiatry Investig 2019 Jul;16(7):479-483 [FREE Full text] [doi: 10.30773/pi.2019.06.20] [Medline: 31352729]
27. Voderholzer U, Valerius G, Schaerer L, Riemann D, Giedke H, Schwärzler F, et al. Is the antidepressive effect of sleep deprivation stabilized by a three day phase advance of the sleep period? A pilot study. Eur Arch Psychiatry Clin Neurosci 2003 Apr;253(2):68-72. [doi: 10.1007/s00406-003-0408-7] [Medline: 12799743]
28. Kurczewska E, Ferensztajn-Rochowiak E, Jasińska-Mikołajczyk A, Chłopocka-Woźniak M, Rybakowski JK. Augmentation of pharmacotherapy by sleep deprivation with sleep phase advance in treatment-resistant depression. Pharmacopsychiatry 2019 Jul;52(4):186-192. [doi: 10.1055/a-0695-9138] [Medline: 30199904]
29. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition DSM-IV-TR. Washington, DC: American Psychiatric Association; 2000.
30. Løventoft PK, Nørregaard LB, Frøkjær E. Designing daybuilder: an experimental app to support people with depression. In: Proceedings of the 12th Participatory Design Conference: Exploratory Papers, Workshop Descriptions, Industry Cases - Volume 2. 2012 Aug Presented at: PDC '12: 12th Participatory Design Conference; August 12-16, 2012; Roskilde, Denmark. [doi: 10.1145/2348144.2348146]
31. Tahara Y, Shibata S. Chronobiology and nutrition. Neuroscience 2013 Dec 03;253:78-88. [doi: 10.1016/j.neuroscience.2013.08.049] [Medline: 24007937]
32. Yamanaka Y, Hashimoto S, Takasu NN, Tanahashi Y, Nishide SY, Honma S, et al. Morning and evening physical exercise differentially regulate the autonomic nervous system during nocturnal sleep in humans. Am J Physiol Regul Integr Comp Physiol 2015 Nov 01;309(9):R1112-R1121 [FREE Full text] [doi: 10.1152/ajpregu.00127.2015] [Medline: 26333783]
33. Grandin LD, Alloy LB, Abramson LY. The social zeitgeber theory, circadian rhythms, and mood disorders: review and evaluation. Clin Psychol Rev 2006 Oct;26(6):679-694. [doi: 10.1016/j.cpr.2006.07.001] [Medline: 16904251]
34. Martiny K, Refsgaard E, Lund V, Lunde M, Sørensen L, Thougaard B, et al. The day-to-day acute effect of wake therapy in patients with major depression using the HAM-D6 as primary outcome measure: results from a randomised controlled trial. PLoS One 2013 Jun 28;8(6):e67264 [FREE Full text] [doi: 10.1371/journal.pone.0067264] [Medline: 23840645]
35. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
36. Licht RW, Qvitzau S, Allerup P, Bech P. Validation of the Bech-Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression; is the total score a valid measure of illness severity? Acta Psychiatr Scand 2005 Feb;111(2):144-149. [doi: 10.1111/j.1600-0447.2004.00440.x] [Medline: 15667434]
37. Bech P, Timmerby N, Martiny K, Lunde M, Soendergaard S. Psychometric evaluation of the Major Depression Inventory (MDI) as depression severity scale using the LEAD (Longitudinal Expert Assessment of All Data) as index of validity. BMC Psychiatry 2015 Aug 05;15:190 [FREE Full text] [doi: 10.1186/s12888-015-0529-3] [Medline: 26242577]
38. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
39. Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 1976;4(2):97-110. [Medline: 1027738]
40. Buysse DJ, Reynolds CF3, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
41. Tullis T, Albert B, Tullis T, Albert W. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Amsterdam, The Netherlands: Elsevier Science; 2010.
42. Kennaway DJ. A critical review of melatonin assays: past and present. J Pineal Res 2019 Aug;67(1):e12572. [doi: 10.1111/jpi.12572] [Medline: 30919486]
43. Danilenko KV, Verevkin EG, Antyufeev VS, Wirz-Justice A, Cajochen C. The hockey-stick method to estimate evening dim light melatonin onset (DLMO) in humans. Chronobiol Int 2014 Apr;31(3):349-355. [doi: 10.3109/07420528.2013.855226] [Medline: 24224578]
44. Lewy AJ. Melatonin and human chronobiology. Cold Spring Harb Symp Quant Biol 2007;72:623-636. [doi: 10.1101/sqb.2007.72.055] [Medline: 18419322]
45. Abdel Latif A, Nasreldin M, Abdel Kader A, Fathy H, Moussa S, Basheer M, et al. A randomized study comparing the short-term neurocognitive outcome of electroconvulsive therapy versus repetitive transcranial magnetic stimulation in the treatment of patients with depression. J Psychiatr Pract 2020 Jan;26(1):23-36. [doi: 10.1097/PRA.0000000000000436] [Medline: 31913967]

46. Ho SC, Chong HY, Chaiyakunapruk N, Tangiisuran B, Jacob SA. Clinical and economic impact of non-adherence to antidepressants in major depressive disorder: a systematic review. J Affect Disord 2016 Mar 15;193:1-10. [doi: 10.1016/j.jad.2015.12.029] [Medline: 26748881]
47. Au J, Reece J. The relationship between chronotype and depressive symptoms: a meta-analysis. J Affect Disord 2017 Aug 15;218:93-104. [doi: 10.1016/j.jad.2017.04.021] [Medline: 28463712]
48. Bortolato B, Carvalho AF, McIntyre RS. Cognitive dysfunction in major depressive disorder: a state-of-the-art clinical review. CNS Neurol Disord Drug Targets 2014;13(10):1804-1818. [doi: 10.2174/1871527313666141130203823] [Medline: 25470396]
49. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
50. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
51. Kent BA, Rahman SA, St Hilaire MA, Grant LK, Rüger M, Czeisler CA, et al. Circadian lipid and hepatic protein rhythms shift with a phase response curve different than melatonin. Nat Commun 2022 Mar 03;13(1):681 [FREE Full text] [doi: 10.1038/s41467-022-28308-6] [Medline: 35115537]
52. Mikkelsen KB, Phan H, Rank ML, Hemmsen MC, de Vos M, Kidmose P. Sleep monitoring using ear-centered setups: investigating the influence from electrode configurations. IEEE Trans Biomed Eng 2022 May;69(5):1564-1572. [doi: 10.1109/TBME.2021.3116274] [Medline: 34587000]
53. Huang Q, Komarzynski S, Bolborea M, Finkenstädt B, Lévi FA. Telemonitored human circadian temperature dynamics during daily routine. Front Physiol 2021 May 10;12:659973 [FREE Full text] [doi: 10.3389/fphys.2021.659973] [Medline: 34040543]
54. Skeldon AC, Dijk DJ, Meyer N, Wulff K. Extracting circadian and sleep parameters from longitudinal data in schizophrenia for the design of pragmatic light interventions. Schizophr Bull 2022 Mar 01;48(2):447-456 [FREE Full text] [doi: 10.1093/schbul/sbab124] [Medline: 34757401]

Abbreviations

CL: confidence level
CRT: circadian reinforcement therapy
DLMO: Dim Light Melatonin Onset
GLM: general linear model
IAOS: Intensive Affective Outpatient Service
MDI: Major Depression Inventory
MEQ: Morningness-Eveningness Questionnaire
PAD: phase angle difference
PSQI: Pittsburgh Sleep Quality Index
SUS: System Usability Score
TAU: treatment as usual
TDV1: time from discharge to visit 1
WHO-5: World Health Organization Well-Being Index

Edited by J Torous; submitted 20.06.23; peer-reviewed by L Yang, M Shin; comments to author 19.07.23; revised version received 10.09.23; accepted 03.10.23; published 27.11.23.

Please cite as:
Aggestrup AS, Svendsen SD, Præstegaard A, Løventoft P, Nørregaard L, Knorr U, Dam H, Frøkjær E, Danilenko K, Hageman I, Faurholt-Jepsen M, Kessing LV, Martiny K
Circadian Reinforcement Therapy in Combination With Electronic Self-Monitoring to Facilitate a Safe Postdischarge Period for Patients With Major Depression: Randomized Controlled Trial
JMIR Ment Health 2023;10:e50072
URL: https://mental.jmir.org/2023/1/e50072
doi:10.2196/50072
PMID:37800194

Original Paper

Mobile Acceptance and Commitment Therapy in Bipolar Disorder: Microrandomized Trial

Amy Cochran[1,2], PhD; Jacob M Maronge[3], PhD; Amanda Victory[4], BSc; Sydney Hoel[5], BSc; Melvin G McInnis[4], MD; Emily BK Thomas[6], PhD

[1]Department of Population Health Sciences, University of Wisconsin Madison, Madison, WI, United States
[2]Department of Mathematics, University of Wisconsin, Madison, WI, United States
[3]Department of Biostatistics, Monroe Dunaway Anderson Cancer Center, Houston, TX, United States
[4]Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States
[5]Department of Medicine, University of Wisconsin Madison, Madison, WI, United States
[6]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, United States

Corresponding Author:
Amy Cochran, PhD
Department of Population Health Sciences
University of Wisconsin Madison
610 Walnut Street
Madison, WI, 53726
United States
Phone: 1 608 262 0772
Email: cochran4@wisc.edu

Abstract

Background: Mobile interventions promise to fill in gaps in care with their broad reach and flexible delivery.

Objective: Our goal was to investigate delivery of a mobile version of acceptance and commitment therapy (ACT) for individuals with bipolar disorder (BP).

Methods: Individuals with BP (n=30) participated in a 6-week microrandomized trial. Twice daily, participants logged symptoms in the app and were repeatedly randomized (or not) to receive an ACT intervention. Self-reported behavior and mood were measured as the energy devoted to moving toward valued domains or away from difficult emotions and with depressive d and manic m scores from the digital survey of mood in BP survey (digiBP).

Results: Participants completed an average of 66% of in-app assessments. Interventions did not significantly impact the average toward energy or away energy but did significantly increase the average manic score m (P=.008) and depressive score d (P=.02). This was driven by increased fidgeting and irritability and interventions focused on increasing awareness of internal experiences.

Conclusions: The findings of the study do not support a larger study on the mobile ACT in BP but have significant implications for future studies seeking mobile therapy for individuals with BP.

Trial Registration: ClinicalTrials.gov NCT04098497; https://clinicaltrials.gov/ct2/show/NCT04098497

(JMIR Ment Health 2023;10:e43164) doi:10.2196/43164

KEYWORDS

acceptance and commitment therapy; bipolar disorder; mobile applications; randomized controlled trials; micro-randomized trial; precision medicine; mindfulness

Introduction

Only 43% of Americans with a mental illness received treatment in 2017, and even fewer among non-Hispanic Black (31%) and Hispanic (33%) individuals [1]. These numbers are partly explained by limited access to therapy, an issue worse for minority groups [2]. Limited access arises due to cost [3], distance [4], and waiting times [5]. Mobile versions of therapy are a promising solution, as they can deliver care at low costs to most people on a schedule that works for them and when they need it the most. In this study, we investigate a mobile version of acceptance and commitment therapy (ACT) for individuals with bipolar disorder (BP).

A mobile version of ACT has many benefits. First, ACT was intended to be effective in general, rather than for specific diagnoses. Its success is attributed to improving a cognitive-behavioral process known as psychological flexibility, which is the ability to pursue things that matter *independent* of negative thoughts and emotions. To emphasize, the goal is not to get rid of unwanted thoughts and emotions, but rather pursue one's values despite them. Psychological flexibility is theorized to require *awareness* of emotions, thoughts, and behaviors; *openness* to negative thoughts and emotions; and *engagement* with values. Second, ACT is effective in low-dose settings. For example, an online-guided ACT improved depressive symptoms in college students [6], and a mobile ACT improved psychological flexibility [7], suggesting improvements can be achieved with mobile technology [8-11]. Third, ACT teaches specific skills, such as mindfulness, that can be used outside a clinic. Thus, a mobile version of ACT could potentially fill gaps in care for people who would benefit from increased psychological flexibility.

Establishing mobile ACT requires confronting several issues. It is unclear *when* mobile ACT could improve a person's mood and behavior. One could argue that mobile ACT is more effective when a person is struggling with unwanted negative thoughts and emotions, as this is when psychological flexibility is most warranted. However, mobile ACT may be less effective at these times, given that psychological flexibility is difficult to enact at any moment, let alone when a person is struggling. It is also unclear *who* would benefit the most from mobile ACT. Here again, one could make opposing arguments: mobile ACT is more effective for a person severely impaired by negative thoughts and mood, given that psychological flexibility is likely more important for these individuals, or less effective, given that psychological flexibility is difficult to achieve especially for those impaired. Finally, it is unclear *which* of the 3 subprocesses (ie, awareness, openness, and engagement) would be more effective target. It is important to clarify who should be delivered which ACT intervention and when.

A microrandomized trial (MRT) is a special type of randomized control trial (RCT) for learning to personalize delivery of mobile interventions according to momentary information [12,13]. In a traditional RCT, individuals are randomized once to an intervention condition. Researchers then evaluate the intervention effect on long-term outcomes. An MRT, by contrast, repeatedly randomizes individuals to intervention conditions. Researchers then evaluate the intervention effect on short-term outcomes. Moreover, because interventions are delivered in diverse settings, researchers can also learn how the intervention effect changes according to momentary information.

We conducted 6-week pilot MRTs in 2 populations susceptible to different levels of impairment: individuals with BP and distressed first-generation college students [14]. This paper focuses on the BP group. The overarching goal is to establish mobile ACT as an effective and personalized option for individuals with BP. Primary goals of the present study were safety and feasibility. Secondary goals were effectiveness and personalization. For safety, we examined changes in depressive and manic symptoms over the study. For feasibility, we investigated how often people logged symptoms in the app. For effectiveness and personalization, we examined if delivery of ACT interventions had a short-term effect on mood and behavior and whether this effect changed depending on the person, current mood, and the type of intervention. Results have significant implications for future studies seeking to deliver and personalize mobile therapy for individuals with BP.

Methods

Overview

Protocols for this study and the parallel study on distressed first-generation college students were previously published [14]. Briefly, individuals with BP (n=30) participated in a 6-week MRT. Participants were randomized to either receive an ACT intervention or not up to twice each day.

Ethics Approval

This study was approved by the Institutional Review Boards at the University of Michigan (HUM126732) and the University of Wisconsin (2017-1322) and is registered at clinicaltrials.gov (NCT04098497). The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Participants

Individuals with BP were recruited from the Prechter Longitudinal Study of Bipolar Disorder by a research technician [15]. Inclusion criteria were a diagnosis of bipolar disorder (type I, II, not otherwise specified), agreement to be contacted for future research, and access to a smartphone. Each participant received their diagnosis based on the Diagnostic Interview for Genetic Studies [16]. Participants gave their consent over the phone and the consent document was electronically signed. Details on how a target sample size of 30 was determined are found in the published protocol [14].

Study Components

Setup

Participants were mailed an activity tracker (Fitbit Alta HR) and were asked to download a mobile app called Lorevimo (Log, Review, and Visualize your Mood). Lorevimo was designed by the team and is available at Apple App and Google Play stores. Upon first opening the app, participants set typical wake-up and bed times, defining windows for logging symptoms. A morning window was defined as 2-7 hours after the typical wake time. An evening window was defined as 3 hours before and 2 hours after the typical bedtime. Next, they watched a 20-minute video that introduces the ACT Matrix, depicting ACT concepts in four quadrants: identifying and sorting values, internal experiences, avoidance behaviors, and values-based behaviors [17].

Initial and Exit Phone Interview

At start and end of the study, mood and health were assessed over the phone by a trained interviewer. Assessments included the Young Mania Rating Scale (YMRS) [18], the Structured Interview Guide for the Hamilton Depression Rating Scale

(HRSD) [19], and the 36-Item Short Form Survey [20]. Although self-administration may be more common for the 36-Item Short Form Survey, telephone administration is also considered a valid mode of delivery [21-24].

Activity Tracking

Participants wore the Fitbit except during a shower or when charging the Fitbit. The Fitbit tracked sleep, activity, and heart rate.

In-App Assessment

Participants logged mood and behavior in the app in the morning and evening. Push notifications were sent as reminders at 2-hour intervals. Mood was self-reported using the 6-item digital survey for mood in bipolar disorder (digiPB) [25,26]. This survey has 3 items (depressed mood, fatigue, and fidgeting) measuring depressive symptoms, 2 items (increased energy and rapid speech) measuring manic symptoms, and 1 item (irritability) measuring both types of symptoms. Each item is rated on a 0-3 ordinal scale. Two scores, *d* and *m*, are computed to measure the severity of depressive and manic symptoms. Participants also answered a 4-item ACT activity survey about current behavior, which are as follows: (1) "In a few words, what behavior are you engaged in right now?" (2) "Does this behavior move you toward who or what matters or away from internal experiences?" (3) "Since [lunchtime or waking up], how much energy was consumed by trying to get rid of unwanted feelings, thoughts, and other internal experiences?" and (4) "Since around lunchtime [lunchtime or waking up], how much energy was consumed by pursuing your values?" The first question was open-ended, the second was binary, and the third and fourth questions were rated on a 0 to 6 ordinal scale.

Intervention Delivery

Participants were repeatedly randomized to receive an intervention or not. Participants were *available* for randomization every time they completed an in-app assessment. Each time a participant was randomized, they had 50-50 chance of receiving an intervention. Over the study, a participant could complete up to 84 in-app assessments (=2 per day × 42 days), which means a participant could have been randomized up to 84 different times. When a participant was randomized to *not* receive an intervention, they were navigated back to the home page. When they were randomized to receive an intervention, the participant was navigated to an intervention prompt. The intervention was selected at random from one of the 84 prompts so that each prompt was equal regardless of whether the prompt had previously been delivered or not.

A total of 84 intervention prompts were designed by the team to build ACT skills, organized into 3 ACT concepts (openness, awareness, and engagement) with 28 questions per concept. Openness questions encouraged participants to accept internal experiences rather than engage in avoidance to suppress such experiences. Awareness questions encouraged participants to pay attention to internal experiences and external context and to be present in the moment. Engagement questions encouraged participants to consider their values and the people, things, and qualities that are important to them. These questions also encouraged participants to examine alignment between values and current behavior.

Outcomes

Primary (Feasibility and Safety)

Feasibility was evaluated based on completion of in-app assessments. Safety was evaluated based on changes in YMRS and HRSD scores from baseline to exit, providing low-level evidence (ie, not causal evidence) that the study impacted mood symptoms.

Secondary and Exploratory (Effectiveness)

Effectiveness was evaluated based on the effect of intervention delivery on toward and away energy, as measured by the ACT activity survey. We also evaluated intervention effects on the *m* and *d* from the digiBP survey. We also explored intervention effects on individual symptoms and moderation by intervention type, age, sex, diagnosis, and current depressive and manic symptoms prior to randomization.

Statistical Analysis

For feasibility, we used a 1-sample *z*-test to assess whether participants responded to over half of the assessments per day for over 60% of the days of the intervention period on average. For safety, we used a 1-sample *z*-test to assess whether a mean change in YMRS or HRSD scores was significantly different from zero and a sign test was used to assess whether an equal proportion of individuals saw an increase in YMRS scores as a decrease from baseline to study exit and similarly for HRSD scores.

For effectiveness, we used a weighted and centered least squares method [27,28] to estimate the average effect of delivering an intervention on primary outcomes (toward and away energy) and secondary outcomes (*d* and *m* scores) as a function of time in the study conditional on the participant being available for randomization. A linear working model was used to estimate these effects. A similar approach was used to estimate moderation of intervention effects except for making requisite changes to the linear working model. The study was not powered for moderation analyses; any subsequent findings may be spurious and are, therefore, reported in Multimedia Appendix 1A. Robust SEs were calculated using a sandwich estimator [29].

The only source of missingness was if a participant did not complete an in-app assessment at their next assessment window after being randomized. As specified in our protocol [14], additional variables were controlled for that predicted missingness if more than 10% of the data was missing. Potential variables included age, sex, diagnosis, time of day, day in the study, count of prior interventions delivered, count of prior completed in-app assessments, count of prior missing data points, toward energy reported immediately before randomization, and away energy reported immediately before randomization. Linear models were built for the logit function of expected missingness as a function of these potential variables. The best model was selected according to quasi information criterion [30]. Variables in the best model were then controlled for when estimating intervention effects.

For analyses, hypothesis tests were 2-tailed, and the significance threshold was $P<.05$.

Results

Participant Flow

Thirty individuals with BP were enrolled between September 2019 and September 2020 (see Multimedia Appendix 1B for the Consolidated Standards Of Reporting Trials diagram). The study ended in October 2020 because enrollment goals were met and data collection was completed. All participants completed the first interview. One participant missed the exit interview and was not included when analyzing safety outcomes. One participant never set up the study app, and 3 participants never linked their Fitbit to the study. Four participants never logged their symptoms and were never randomized to receive an intervention. These 4 participants were not included when analyzing effectiveness outcomes. Six participants never used Fitbit. All participants were included when analyzing feasibility outcomes.

Sample Characteristics

Table 1 summarizes sample characteristics. They had an average (SD) age of 42.70 (11.11) years and were 60% female. The majority were White (83%), non-Hispanic (93%), and diagnosed with bipolar I disorder (80%).

Table 1. Characteristics of the sample population (N=30).

Variables	Value
Age (years), mean (SD)	42.70 (11.11)
Female, n (%)	18 (60)
Race, n (%)	
White	25 (83)
Black or African American	2 (7)
Asian	0 (0)
American Indian or Alaskan Native	1 (3)
More than 1	2 (7)
Hispanic	2 (7)
Diagnosis, n (%)	
Bipolar I disorder	24 (80)
Bipolar II disorder	6 (20)
Bipolar disorder not otherwise specified	0 (0)
Baseline HRSD[a] score, mean (SD)	6.20 (5.78)
Baseline YMRS[b] score, mean (SD)	1.83 (3.29)

[a]HRSD: Hamilton Rating Scale for Depression.

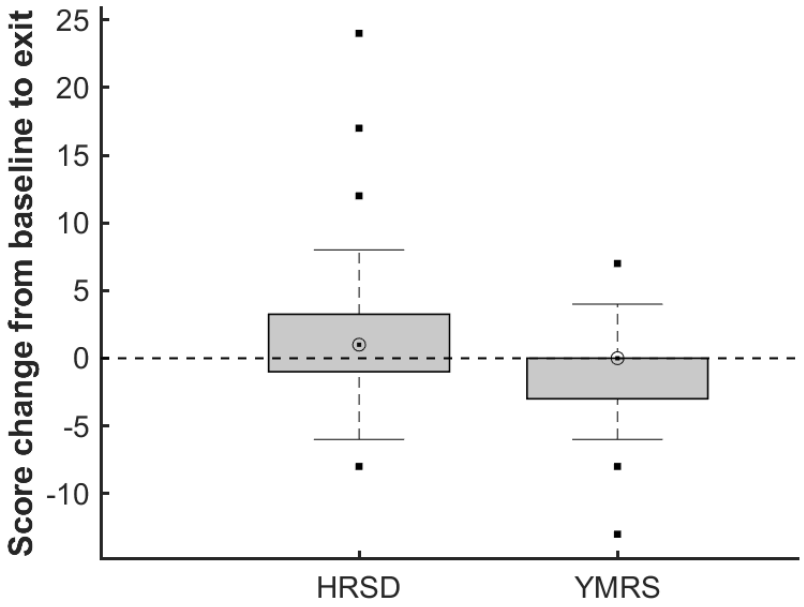
[b]YMRS: Young Mania Rating Scale.

Safety Outcomes

Figure 1 illustrates the change from baseline to study exit in HRSD and YMRS scores for the 29 participants who completed baseline and exit interviews. Depressive severity increased slightly with an average increase in HRSD score of 2.1 points (t_{28}=1.75, P=.09) and with 15 participants seeing an increase in HRSD scores compared to 9 participants seeing a decrease (mean decrease 63%, z=1.22, P=.22). Five participants saw no change in HRSD scores. Manic severity decreased slightly with an average decrease in YMRS score of 1.2 points (t_{28}=1.74, P=.09) and with 4 participants seeing an increase in YMRS scores compared to 9 participants seeing a decrease (mean decrease 31%, z=−1.39, P=.17). Sixteen participants saw no change in YMRS scores.

Given that participation was associated with a slight increase in depressive symptoms, we investigated whether ACT interventions contributed to these changes. Among participants who were randomized at least once, the average HRSD score still increased over the study, but only by 1.2 points (t_{24}=1.15, P=.26) rather than 2.1 points. Average YMRS score decreased by the same 1.2 points (t_{24}=1.52, P=.14) as in the full sample.

Figure 1. Box plot of change in manic and depressive severity over course of 6-week study, as measured by the HRSD and the YMRS. Extreme values, marked with a dot, are any data points less than the 25th percentile minus 1.5 times the interquartile range or greater than the 75th percentile plus 1.5 times the interquartile range. HRSD: Hamilton Rating Scale for Depression; YMRS: Young Mania Rating Scale.



Feasibility

Our second analysis investigated whether participants were available for randomization, that is, whether a participant logged their symptoms in the app at one of the 84 times points. Participants were available for randomization for an average of 66% of the time points. Excluding the 4 participants who never logged symptoms, availability increased to an average of 76% of the time points. In addition, participants were available at least once a day for an average of 74% of the days, which was significantly larger than our prespecified target of 60% (t_{29}=2.15, *P*=.04). Again excluding participants who never logged symptoms, participants were available at least once a day for an average of 85% of the days.

Effectiveness

Our final analysis explored whether ACT interventions had a short-term impact on mood and behavior. Primary outcomes were missing for 12% of the times that a participant was randomized. Because this amount was larger than a prespecified cutoff of 10%, all models in this section controlled for the following covariates predicting missingness: time of day, count of prior missing values, count of prior logs, and self-reported toward behavior immediately prior to randomization (see Multimedia Appendix 1C for details on model selection and final models).

Adjusting for these covariates, we found that ACT interventions did not have a significant impact on toward behavior (β=−.006; *z*=−0.11; *P*=.91) or away behavior (β=.093; *z*=1.65; *P*=.10). ACT interventions did, however, significantly increase average depressive score *d* (β=.57, *z*=2.39, *P*=.02) and manic score *m* (β=.19; *z*=2.67; *P*=.008). The day in the study, ranging from day 1 through day 42, did not significantly moderate the effect of the intervention on any of the 4 outcomes.

Given that the ACT interventions worsened mood, we explored the symptoms that might be more greatly impaired and interventions that might be more impairing. Table 2 reports the average effects for all 6 symptoms from the digiBP survey. Symptoms that were significantly impacted by the interventions were fidgeting (β=.130; *z*=2.61; *P*=.009) and irritability (β=.129; *z*=3.34; *P*=.001).

We looked at the type of intervention in 2 ways: the 3 ACT processes (ie, openness, engagement, and awareness) or the 4 quadrants in the ACT matrix (ie, toward behavior, away behavior, internal experiences, and who/what matters). Table 3 reports average effects by intervention type. In the former method, awareness interventions led to a significant increase in average depressive score *d* (β=.65; *z*=2.04; *P*=.04) and manic score (β=.31; *z*=3.04; *P*=.002). Openness or engagement intervention did not lead to a significant increase in any score. In the latter method, interventions focused on who or what matters significantly increased average depressive score *d* (β=0.86; *z*=2.54; *P*=.01), whereas interventions focused on away behaviors significantly increased average manic score *m* (β=.92; *z*=2.09; *P*=.04).

Table 2. Average effects of acceptance and commitment therapy interventions on individual symptoms.

Outcome	β	95% CI	*z* score	*P* value
Depressed mood	.056	(−.012 to .123)	1.62	.11
Fatigue	.026	(−.081 to .132)	0.47	.13
Fidgeting	.130	(.033 to .227)	2.61	.009
Increased energy	.004	(−.050 to .058)	0.13	.89
Rapid speech	.030	(−.010 to .070)	1.47	.14
Irritability	.129	(.053 to .204)	3.34	.001

Table 3. Average effects of interventions on secondary outcomes by intervention type.

Type	Depressive score *d*				Manic score *m*			
	β	95% CI	*z* score	*P* value	β	95% CI	*z* score	*P* value
Engagement	.58	(−.05 to 1.28)	1.82	.07	.19	(−.06 to .43)	1.49	.14
Awareness	.65	(.025 to 1.28)	2.04	.04	.31	(.11 to .50)	3.04	.002
Openness	.45	(−.26 to 1.16)	1.25	.21	.09	(−.24 to .42)	0.54	.59
Toward behaviors	.32	(−.31 to .96)	0.99	.32	.14	(−.09 to .37)	1.22	.22
Away behaviors	.83	(−.63 to 2.29)	1.12	.26	.92	(.06 to 1.79)	2.09	.04
Internal experiences	.48	(−.19 to 1.15)	1.40	.16	.23	(−.06 to .52)	1.53	.13
Who/what matters	.86	(.19 to 1.52)	2.54	.01	.08	(−.17 to .34)	0.65	.52

Discussion

We presented results from a pilot MRT on delivering mobile ACT interventions to individuals with BP (n=30). This MRT was one of the two parallel studies, with the other focusing on distressed first-generation college students. The goal was to collect evidence to make a go or no-go decision about pursuing a larger study on effectiveness and personalization of mobile ACT in BP. For reasons described below, we concluded such a study is not warranted. Despite this negative conclusion, we discuss several important findings that inform future mobile studies in BP.

A main concern with mobile interventions is whether users sufficiently use the app to have a measurable effect. Thus, we wanted to know if participants met our benchmark for logging symptoms regularly and subsequently being available to receive interventions. The key number here is that participants were available for randomization at an average of 66% of twice-daily time points, because this number determines suitable sample sizes for MRTs [12,31]. Note that participants were not renumerated for interacting with the app. In fact, 66% number includes 4 participants who never logged symptoms. Participants were reminded by push notifications to log symptoms, which is likely to be crucial for achieving high utilization of the app.

Although limited use of the app does not appear to be a limiting factor, safety and effectiveness does appear to be a factor. The first hint that the mobile ACT may have a detrimental effect arose when analyzing safety outcomes. There was a slight 2-point increase in depressive symptoms as measured by the HRSD. Although this increase was nonsignificant, the small sample limits our ability to detect significant changes. On a positive note, there was only a 1-point increase in the HRSD when excluding patients who never were randomized and even a slight 1-point decrease in manic symptoms as measured by the YMRS. Of course, we did not randomize individuals to participate in this study vs not, so we cannot conclude that these changes were due to study participation.

When we look at effectiveness, the picture becomes clearer. Delivering an ACT intervention led to a significant increase in depressive and manic symptoms at the next time period. That is, if a person was delivered an intervention in the morning, they were expected to report higher mood symptoms in the evening than a person who was not delivered an intervention. Similarly, if a person was delivered an intervention in the evening, they were expected to report higher mood symptoms in the morning on the next day than a person who was not delivered an intervention. We looked more closely to determine which symptoms might be more greatly affected. Irritability and fidgeting were greatly affected compared to fatigue, depressed mood, increased energy, or rapid speech. Thus, symptoms induced by ACT interventions were more akin to dysphoric or anxious depression as opposed to anhedonic depression or euphoric mania.

So why might ACT interventions make someone more irritable or agitated? We looked at different factors. Interventions focused on raising a person's awareness of internal experiences (eg, emotions) had the most significant effect. Furthermore, if a person was currently depressed or manic, then interventions had even larger effect on depressive symptoms. This may be a manifestation of a feature of BP known as emotional reactivity, whereby a person's emotions react more intensely when provoked [32]. These findings suggest that awareness of negative emotions and thoughts, especially when a person is

already experiencing intense mood symptoms, can increase irritability and agitation in individuals with BP.

The downside of raising awareness helps shed light on other digital interventions in BP. A qualitative study from a prior RCT identified common explanations for why individuals with BP did not complete web-based psychoeducation that included difficulties with acute phases of BP and not wanting to think about one's illness. Moreover, the MONARCA I and II trials found that their mobile intervention group involving computer-based monitoring experienced a nonsignificant increase in average depressive symptoms and nonsignificant decrease in manic symptoms as measured by the HRSD and YMRS [33,34]. The RADMIS trial found their mobile intervention group experienced a nonsignificant decrease in average HRSD scores and a significant increase in average YMRS scores [35].

These findings provide 3 lessons for future mobile studies in BP. The first is that given raising awareness of mood can worsen mood, monitoring mood may carry risks. Put differently, focusing on what is not going well, such as by asking about current mood or behavior, may be distressing for people with BP. Partly in conflict, individuals with BP had previously endorsed raising self-awareness as the best reason for digital self-monitoring of symptoms [36]. Moreover, the World Health Organization recommends that individuals with BP should monitor their mood [37]. One possible compromise would be passive solutions for monitoring mood [38], such as voice patterns [33,34,39], phone keystroke data [40], global positioning system [41], or phone metadata [42]. It is also possible that monitoring of mood *alone* in combination with therapist support may serve a helpful function.

The second lesson is that mood might not be a suitable outcome. As noted earlier, ACT is designed to increase psychological flexibility, that is, not get rid of unwanted thoughts or emotions, but rather pursue what matters despite them. To this point, our primary outcomes were energy devoted to behaviors that move a person toward who or what matters to them, and away from unwanted emotions and thoughts. The only estimated effect of ACT interventions that was near zero was toward energy. Future studies may want to align outcomes with the cognitive behavioral process that is targeted. Given that instability of mood can be as debilitating in BP as mood itself [43,44], a better outcome may be mood stability. Consistency in values-based behavior and workable responses to mood changes may also be an outcome congruent with the goal of psychological flexibility.

The suitability of mood as an outcome may be especially salient when the intervention (like ACT) incorporates mindfulness. Part of mindfulness is increased awareness of symptoms. Through this lens, it is possible that ACT interventions had the desired effect: participants become more aware of unwanted emotions, and consequently, were reporting higher symptoms. This awareness may have temporarily intensified the experience of the emotion but could promote long-term self-efficacy, mindfulness, or psychological flexibility. Furthermore, becoming aware of unpleasant emotions can be both irritating and agitating. Awareness alone may be insufficient to help individuals. Although interventions in openness and engagement were included, interventions were randomized, and thus were not delivered in a particular order that may have benefited those developing awareness of intense and unpleasant emotions. Future work may examine the impact of the order of interventions, such as building skills in awareness, followed by openness, and in parallel, altering behavioral engagement. Fittingly, although the interventions led to higher reported symptoms on average, it did not deter participants from engaging with the app. In fact, we found, it increased the likelihood of logging symptoms at the next time point.

The third lesson is that there might be a better mobile intervention for BP. ACT was chosen for its potential to reach a transdiagnostic audience, with our 2 initial samples chosen at the extreme ends of impairment, but it may be better to focus on interventions specific to BP, such as interpersonal and social rhythms therapy. Alternatively, a mobile intervention may need to be augmented by clinical support. Regardless, it is still notable that a mobile intervention can have a significant effect, though expectedly small (~0.1 standardized effect size), for a pilot study not powered for effectiveness. This information could inform future MRTs on BP as effect sizes are needed, for sample, size calculations [12,31].

There are several limitations to keep in mind. First, we did not collect data on psychological or pharmacological treatment. Thus, we did not control for treatment. Second, our sample was small and relatively homogeneous being 60% female, 83% White, and 93% non-Hispanic. Third, limited data were collected on engagement with the app or with the intervention. The study did not, for example, determine who viewed the introductory ACT video, if individuals would have engaged with the app were to remove the in-app assessments, or if participants engaged with those ACT processes targeted by the intervention. Regarding this last point, however, we did recently publish a preplanned, interim, and qualitative analysis of participant open-ended responses to behavior prompts and intervention prompts [45]. By reading and coding these responses, these qualitative analyses were investigated if participants responded in a way that is congruent with the targeted ACT process and was intended to show intervention fidelity, supporting the study's reporting of preplanned primary and secondary analyses of the MRT. Fourth, participants wore Fitbit activity trackers, which induce behavioral changes. Fifth, the ACT survey used, including the primary outcomes of toward and away energy, was developed for this study and has yet to be validated.

A final limitation is that the time frame between intervention and assessment, which is roughly half a day, may have been too short to observe behavioral changes induced from ACT. The motivating ACT literature has shown that low-dose versions of ACT have been effective but have not examined how quickly people respond. Therefore, while the present study adds to the ACT literature by demonstrating that ACT interventions can have significant effect on mood in a short time frame, it remains unclear if ACT interventions could similarly affect behavior in this time frame. With limited investigation into a short time frame in the ACT literature, the short time frame for the present study was primarily motivated by the microrandomized trial literature, which has shown that behavioral interventions in

other contexts could have effects in time frames as short as 30 minutes [27].

In summary, we presented results from a pilot MRT delivering a mobile ACT intervention in BP. Because the ACT intervention worsened mood, we concluded a larger MRT was not warranted for BP. However, our results are informative for future studies on BP: it is feasible to get measurable effects in MRTs with a small sample size; estimated availability and effects can inform sample sizes for MRTs; self-reported mood may not be the best target; and interventions may need to manage the consequences of raising symptom awareness.

Acknowledgments

This research was supported by the Heinz C Prechter Bipolar Research Program at the University of Michigan (MGM); the Richard Tam Foundation (MGM); Health and Human Services, Department of National Institutes of Health (ALC, grant number K01 MH112876); and by the University of Wisconsin-Madison's Ira and Ineva Reilly Baldwin Idea Endowment Seed Grant (EBKT). The authors would especially like to thank Tijana Sagorac Gruichich for her help in coordinating the study and for her continued involvement in this research.

Conflicts of Interest

MGM has consulted for Janssen and Otsuka Pharmaceuticals and received research support from Janssen in the past 5 years, all unrelated to the current work. EBKT has consulted with Macmillan Learning on work unrelated to this research.

Multimedia Appendix 1
Accompanying appendix reporting (A) Moderation analyses, (B) CONSORT Diagram, (C) Missingness procedure, and (D) Lagged effects.
[DOCX File , 43 KB - mental_v10i1e43164_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 87 KB - mental_v10i1e43164_app2.pdf]

References

1. McCance-Katz EF. The national survey on drug use and health: 2017. Substance Abuse and Mental Health Services Administration. 2017. URL: https://www.samhsa.gov/data/sites/default/files/nsduh-ppt-09-2018.pdf [accessed 2019-05-07]
2. Cook BL, Trinh NH, Li Z, Hou SS, Progovac AM. Trends in racial-ethnic disparities in access to mental health care, 2004-2012. Psychiatr Serv 2017;68(1):9-16 [FREE Full text] [doi: 10.1176/appi.ps.201500453] [Medline: 27476805]
3. Rowan K, McAlpine DD, Blewett LA. Access and cost barriers to mental health care, by insurance status, 1999-2010. Health Aff 2013;32(10):1723-1730 [FREE Full text] [doi: 10.1377/hlthaff.2013.0133] [Medline: 24101061]
4. Schmitt SK, Phibbs CS, Piette JD. The influence of distance on utilization of outpatient mental health aftercare following inpatient substance abuse treatment. Addict Behav 2003;28(6):1183-1192. [doi: 10.1016/s0306-4603(02)00218-6] [Medline: 12834661]
5. Rhodes KV, Vieth TL, Kushner H, Levy H, Asplin BR. Referral without access: for psychiatric services, wait for the beep. Ann Emerg Med 2009;54(2):272-278. [doi: 10.1016/j.annemergmed.2008.08.023] [Medline: 18950898]
6. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: a randomized controlled clinical trial. Behav Res Ther 2016;78:30-42. [doi: 10.1016/j.brat.2016.01.001] [Medline: 26848517]
7. Mattila E, Lappalainen R, Välkkynen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016;4(3):e90 [FREE Full text] [doi: 10.2196/mhealth.5241] [Medline: 27468653]
8. Ly KH, Asplund K, Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. Internet Interv 2014;1(3):95-101. [doi: 10.1016/j.invent.2014.06.003]
9. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
10. Ahtinen A, Mattila E, Välkkynen P, Kaipainen K, Vanhala T, Ermes M, et al. Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. JMIR Mhealth Uhealth 2013;1(2):e11 [FREE Full text] [doi: 10.2196/mhealth.2596] [Medline: 25100683]
11. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. Behav Modif 2017;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]

12. Liao P, Klasnja P, Tewari A, Murphy SA. Sample size calculations for micro-randomized trials in mHealth. Stat Med 2016;35(12):1944-1971 [FREE Full text] [doi: 10.1002/sim.6847] [Medline: 26707831]
13. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
14. Kroska EB, Hoel S, Victory A, Murphy SA, McInnis MG, Stowe ZN, et al. Optimizing an acceptance and commitment therapy microintervention via a mobile app with two cohorts: protocol for micro-randomized trials. JMIR Res Protoc 2020;9(9):e17086 [FREE Full text] [doi: 10.2196/17086] [Medline: 32965227]
15. McInnis MG, Assari S, Kamali M, Ryan K, Langenecker SA, Saunders EFH, Prechter Bipolar Clinical Research Collaborative. Cohort profile: the Heinz C. Prechter longitudinal study of bipolar disorder. Int J Epidemiol 2018;47(1):28-28n [FREE Full text] [doi: 10.1093/ije/dyx229] [Medline: 29211851]
16. Nurnberger JI, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH genetics initiative. Arch Gen Psychiatry 1994;51(11):849-859. [doi: 10.1001/archpsyc.1994.03950110009002] [Medline: 7944874]
17. Polk KL, Schoendorff B, Webster M, Olaz FO. The Essential Guide to the ACT Matrix : A Step-By-Step Approach to Using the ACT Matrix Model in Clinical Practice. Oakland, CA: New Harbinger Publications, Inc; 2016.
18. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978;133:429-435. [doi: 10.1192/bjp.133.5.429] [Medline: 728692]
19. Williams JPW. Structured interview guides for the Hamilton rating scales. In: The Hamilton Scales. Berlin, Heidelberg: Springer; 1990:48-63.
20. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30(6):473-483. [Medline: 1593914]
21. Açma A, Carrat F, Hejblum G. Comparing SF-36 scores collected through web-based questionnaire self-completions and telephone interviews: an ancillary study of the SENTIPAT multicenter randomized controlled trial. J Med Internet Res 2022;24(3):e29009 [FREE Full text] [doi: 10.2196/29009] [Medline: 35266869]
22. García M, Rohlfs I, Vila J, Sala J, Pena A, Masiá R, REGICOR Investigators. Comparison between telephone and self-administration of short form health survey questionnaire (SF-36). Gac Sanit 2005;19(6):433-439 [FREE Full text] [doi: 10.1016/s0213-9111(05)71393-5] [Medline: 16483520]
23. Ware JE. SF-36 Health Survey: Manual and Interpretation Guide. Boston: The Health Institute, New England Medical Center; 1993.
24. Watson EK, Firman DW, Baade PD, Ring I. Telephone administration of the SF-36 health survey: validation studies and population norms for adults in Queensland. Aust N Z J Public Health 1996;20(4):359-363 [FREE Full text] [doi: 10.1111/j.1467-842x.1996.tb01046.x] [Medline: 8908757]
25. Sagorac Gruichich T, David Gomez JC, Zayas-Cabán G, McInnis MG, Cochran AL. A digital self-report survey of mood for bipolar disorder. Bipolar Disord 2021;23(8):810-820 [FREE Full text] [doi: 10.1111/bdi.13058] [Medline: 33587813]
26. Cochran A, Belman-Wells L, McInnis M. Engagement strategies for self-monitoring symptoms of bipolar disorder with mobile and wearable technology: protocol for a randomized controlled trial. JMIR Res Protoc 2018;7(5):e130 [FREE Full text] [doi: 10.2196/resprot.9899] [Medline: 29748160]
27. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of heartSteps. Ann Behav Med 2019;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
28. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. J Am Stat Assoc 2018;113(523):1112-1121 [FREE Full text] [doi: 10.1080/01621459.2017.1305274] [Medline: 30467446]
29. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Soft 2006;15(2). [doi: 10.18637/jss.v015.i02]
30. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics 2001;57(1):120-125. [doi: 10.1111/j.0006-341x.2001.00120.x] [Medline: 11252586]
31. Seewald N, Sun J, Liao P. MRT-SS calculator: an R shiny application for sample size calculation in micro-randomized trials. ArXiv Preprint posted online August 5, 2016 [FREE Full text]
32. Henry C, Phillips M, Leibenluft E, M'Bailara K, Houenou J, Leboyer M. Emotional dysfunction as a marker of bipolar disorders. Front Biosci (Elite Ed) 2012;4(7):2622-2630 [FREE Full text] [doi: 10.2741/e578] [Medline: 22652673]
33. Faurholt-Jepsen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, et al. Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. Psychol Med 2015;45(13):2691-2704. [doi: 10.1017/S0033291715000410] [Medline: 26220802]
34. Faurholt-Jepsen M, Frost M, Christensen EM, Bardram JE, Vinberg M, Kessing LV. The effect of smartphone-based monitoring on illness activity in bipolar disorder: the MONARCA II randomized controlled single-blinded trial. Psychol Med 2020;50(5):838-848. [doi: 10.1017/S0033291719000710] [Medline: 30944054]

35. Faurholt-Jepsen M, Lindbjerg Tønning M, Fros M, Martiny K, Tuxen N, Rosenberg N, et al. Reducing the rate of psychiatric re-admissions in bipolar disorder using smartphones-the RADMIS trial. Acta Psychiatr Scand 2021;143(5):453-465. [doi: 10.1111/acps.13274] [Medline: 33354769]
36. Van Til K, McInnis MG, Cochran A. A comparative study of engagement in mobile and wearable health monitoring for bipolar disorder. Bipolar Disord 2020;22(2):182-190 [FREE Full text] [doi: 10.1111/bdi.12849] [Medline: 31610074]
37. mhGAP intervention guide for mental, neurological and substance use disorders in non-specialized health settings: mental health gap action programme (mhGAP), version 2.0. World Health Organization. 2016. URL: https://apps.who.int/iris/handle/10665/250239 [accessed 2023-03-22]
38. Victory A, Letkiewicz A, Cochran AL. Digital solutions for shaping mood and behavior among individuals with mood disorders. Curr Opin Syst Biol 2020;21:25-31 [FREE Full text] [doi: 10.1016/j.coisb.2020.07.008] [Medline: 32905495]
39. Gideon J, McInnis MG, Provost EM. Improving cross-corpus speech emotion recognition with adversarial discriminative domain generalization (ADDoG). IEEE Trans Affect Comput 2021;12(4):1055-1068. [doi: 10.1109/taffc.2019.2916092]
40. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a biAffect digital phenotyping study. J Med Internet Res 2018;20(7):e241 [FREE Full text] [doi: 10.2196/jmir.9775] [Medline: 30030209]
41. Depp CA, Bashem J, Moore RC, Holden JL, Mikhael T, Swendsen J, et al. GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. npj Digit Med 2019;2(1):108. [doi: 10.1038/s41746-019-0182-1]
42. Faurholt-Jepsen M, Busk J, Þórarinsdóttir H, Frost M, Bardram JE, Vinberg M, et al. Objective smartphone data as a potential diagnostic marker of bipolar disorder. Aust N Z J Psychiatry 2019;53(2):119-128. [doi: 10.1177/0004867418808900] [Medline: 30387368]
43. Cochran AL, McInnis MG, Forger DB. Data-driven classification of bipolar I disorder from longitudinal course of mood. Transl Psychiatry 2016;6(10):e912 [FREE Full text] [doi: 10.1038/tp.2016.166] [Medline: 27727242]
44. Cochran AL, Schultz A, McInnis MG, Forger DB. Testing frameworks for personalizing bipolar disorder. Transl Psychiatry 2018;8(1):36 [FREE Full text] [doi: 10.1038/s41398-017-0084-4] [Medline: 29391394]
45. Hoel S, Victory A, Sagorac Gruichich T, Stowe ZN, McInnis MG, Cochran A, et al. A mixed-methods analysis of mobile ACT responses from two cohorts. Front Digit Health 2022;4:869143 [FREE Full text] [doi: 10.3389/fdgth.2022.869143] [Medline: 35633737]

Abbreviations

ACT: acceptance and commitment therapy
BP: bipolar disorder
HRSD: Hamilton Depression Rating Scale
MRT: microrandomized trial
RCT: randomized control trial
YMRS: Young Mania Rating Scale

Edited by J Torous; submitted 01.10.22; peer-reviewed by FJ Keefe, E Morton; comments to author 10.11.22; revised version received 29.01.23; accepted 13.02.23; published 20.04.23.

Please cite as:
Cochran A, Maronge JM, Victory A, Hoel S, McInnis MG, Thomas EBK
Mobile Acceptance and Commitment Therapy in Bipolar Disorder: Microrandomized Trial
JMIR Ment Health 2023;10:e43164
URL: https://mental.jmir.org/2023/1/e43164
doi:10.2196/43164
PMID:37079363

Original Paper

A Motivational Interviewing Chatbot With Generative Reflections for Increasing Readiness to Quit Smoking: Iterative Development Study

Andrew Brown[1], BASc; Ash Tanuj Kumar[1], BSc; Osnat Melamed[2,3], MSc, MD; Imtihan Ahmed[1], BSc, MASc; Yu Hao Wang[1], BASc; Arnaud Deza[1], BASc; Marc Morcos[1]; Leon Zhu[1]; Marta Maslej[4], PhD; Nadia Minian[2,3,5,6,7], PhD; Vidya Sujaya[1], BA; Jodi Wolff[2], BA, MSc; Olivia Doggett[8], BA, MI; Mathew Iantorno[8], BA, MA; Matt Ratto[8], PhD; Peter Selby[2,3,9], MBBS, MHSc; Jonathan Rose[1,2], PhD

[1]The Edward S Rogers Sr Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada

[2]INTREPID Lab, Centre for Addiction and Mental Health, Toronto, ON, Canada

[3]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[4]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada

[5]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada

[6]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON, Canada

[7]Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada

[8]Faculty of Information, University of Toronto, Toronto, ON, Canada

[9]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Jonathan Rose, PhD
The Edward S Rogers Sr Department of Electrical & Computer Engineering
University of Toronto
10 King's College Rd
Toronto, ON, M5S 3G4
Canada
Phone: 1 416 978 6992
Email: jonathan.rose@ece.utoronto.ca

Abstract

Background: The motivational interviewing (MI) approach has been shown to help move ambivalent smokers toward the decision to quit smoking. There have been several attempts to broaden access to MI through text-based chatbots. These typically use scripted responses to client statements, but such nonspecific responses have been shown to reduce effectiveness. Recent advances in natural language processing provide a new way to create responses that are specific to a client's statements, using a *generative* language model.

Objective: This study aimed to design, evolve, and measure the effectiveness of a chatbot system that can guide ambivalent people who smoke toward the decision to quit smoking with MI-style generative reflections.

Methods: Over time, 4 different MI chatbot versions were evolved, and each version was tested with a separate group of ambivalent smokers. A total of 349 smokers were recruited through a web-based recruitment platform. The first chatbot version only asked questions without reflections on the answers. The second version asked the questions and provided reflections with an initial version of the reflection generator. The third version used an improved reflection generator, and the fourth version added extended interaction on some of the questions. Participants' readiness to quit was measured before the conversation and 1 week later using an 11-point scale that measured 3 attributes related to smoking cessation: readiness, confidence, and importance. The number of quit attempts made in the week before the conversation and the week after was surveyed; in addition, participants rated the perceived empathy of the chatbot. The main body of the conversation consists of 5 scripted questions, responses from participants, and (for 3 of the 4 versions) generated reflections. A pretrained transformer-based neural network was fine-tuned on examples of high-quality reflections to generate MI reflections.

Results: The increase in average confidence using the nongenerative version was 1.0 (SD 2.0; P=.001), whereas for the 3 generative versions, the increases ranged from 1.2 to 1.3 (SD 2.0-2.3; P<.001). The extended conversation with improved

generative reflections was the only version associated with a significant increase in average importance (0.7, SD 2.0; *P*<.001) and readiness (0.4, SD 1.7; *P*=.01). The enhanced reflection and extended conversations exhibited significantly better perceived empathy than the nongenerative conversation (*P*=.02 and *P*=.004, respectively). The number of quit attempts did not significantly change between the week before the conversation and the week after across all 4 conversations.

Conclusions: The results suggest that generative reflections increase the impact of a conversation on readiness to quit smoking 1 week later, although a significant portion of the impact seen so far can be achieved by only asking questions without the reflections. These results support further evolution of the chatbot conversation and can serve as a basis for comparison against more advanced versions.

(JMIR Ment Health 2023;10:e49132) doi:10.2196/49132

KEYWORDS

conversational agents; chatbots; behavior change; smoking cessation; motivational interviewing; deep learning; natural language processing; transformers; generative artificial intelligence; artificial intelligence; AI

Introduction

Background

Tobacco use is the leading preventable cause of premature death in Canada, killing 45,000 Canadians every year [1], with 4.6 million Canadians ensnared by the addiction [2]. A key step in smoking cessation is the decision by the smoker to quit; yet, 80% of all smokers are *ambivalent* [3] about quitting and make no current effort to stop smoking [4]. In this paper, we report on an automated method to engage ambivalent smokers in a web-based conversation with the aim of moving them toward the decision to quit. This goal is distinct from smoking cessation efforts that assume the smoker is ready and willing to quit; however, the decision to quit is a necessary precursor of any quit attempt.

Smokers can be guided toward the decision to quit by a widely used talk therapy approach known as motivational interviewing (MI) [5]. MI guides individuals toward healthy behavior change by helping them explore their ambivalence. As MI relies on highly trained clinicians working at hospitals and specialized clinics, it is both expensive and difficult to access. Clinicians are usually engaged only *after* a health issue occurs, whereas earlier engagement with a more accessible chatbot could improve health outcomes and even prevent illness or death. For every 2 smokers helped to quit, 1 life is saved from a tobacco related death [6]. If it were possible to *automate* an MI-style conversation and deploy it directly to smokers on the web, smokers could be helped much sooner.

However, it is challenging for a machine to achieve the level of understanding and facility needed to practice MI. Prior efforts at automated therapy, beginning with ELIZA [7] and proceeding through many generations of dialogue systems [8-10], suffer from 2 key issues. First, because most of the outgoing text is scripted, these systems have difficulty responding to the specific things that an individual says. These responses are often seen by users as either repetitive or too generic [11]. Second, many chatbots do not permit free-form text input, which prevents users from expressing themselves fully. Recent dramatic advances in *natural language generation* [6,12-14] have produced language models that can generate very humanlike responses that are more relevant to the free-form dialogue of a human.

In this paper, we present the design of several versions of a chatbot, called *MIBot*, that makes use of these new kinds of language models to generate context-specific responses to user statements in combination with scripted interactions. We also present a scientific infrastructure for measuring the impact of MIBot on recruited smokers.

This project is a collaboration among a group of MI researchers and clinicians, computer engineers, and social scientists. The clinicians, who work with patients to resolve their ambivalence toward quitting smoking using the MI approach, bring their expertise to the initial design and iterative steps of the chatbot. The computer engineers bring experience in methods of software and natural language generation. The social scientists draw on perspectives in human-computer interaction and notions of trust. The ongoing decision-making and iterative improvement of the chatbot described in this paper follows the principles and rationale of a co-design process [15]. As noted by Donia and Shaw [15], such processes are key to appropriate outcomes when dealing with health care and artificial intelligence technologies. The group has met biweekly for >2 years, and each subgroup has sought to learn from the other so that a true cross-disciplinary interaction and purposeful evolution of the chatbot take place. The interactions explore the tension between what is needed in a successful MI conversation and what is possible to automate in limited time. We have all enjoyed learning across traditional disciplinary boundaries.

This paper is organized as follows: the *Prior Related Work* subsection reviews MI and related work on conversational agents and chatbots, as well as the relevant new capabilities in natural language generation. The *Methods* section describes the 4 versions of the chatbot and the training of the neural networks needed to provide the generative responses, as well as the recruitment and measurement methods. The *Results* section presents the results from the 4 versions of the chatbot, and the *Discussion* section contrasts these results and the impact of the different versions.

Prior Related Work

The MI Approach

MI is a counseling approach that helps patients increase motivation toward changing unhealthy behaviors, including addictions [5]. The goal of MI is to move a person away from

ambivalence, a conflicted state where opposing attitudes or feelings coexist in an individual toward changing a behavior. MI counselors use a structured conversation that includes open-ended questions and reflective listening, which encourage a patient to contemplate the roots of a behavior and guide them toward overcoming their ambivalence.

A core skill used by MI practitioners is reflective listening [5,16,17]. A reflective listener responds to patient statements with words that both reflect what is said and guide the patient toward continued exploration of their thoughts and feelings about change. These responses, called *reflections* in MI, can be simple or complex. A simple reflection repeats or rephrases what the patient has said to convey understanding and invite continuation of the conversation. A complex reflection attempts to infer something relevant to the prior conversation or guess something that might be relevant to a recent utterance, which also invites continued contemplation on the part of the patient.

MI has been shown to be a successful therapeutic tool for motivating many behavior changes [18], including smoking cessation [19,20].

Natural Language Processing and Generative Networks

Natural language processing (NLP) is a subfield of linguistics, computer science, and machine learning concerned with the interactions between computers and human (natural) languages [21]. Over the last 11 years, there have been significant advances in the field of NLP [8]. A key step was the invention of limited-sized word vectors or embeddings, in which it has been shown that a small-sized vector of real numbers (from 50 to 300) could represent the meaning of individual words or parts of words (called *tokens*) [22,23]. Hereinafter, for simplicity, we will refer to these elements as words. These word vectors make it possible to know whether 2 words have similar meaning through a numerical comparison and have led to significant advances in the subfields of speech recognition, natural language understanding (NLU), question answering (QA), and natural language generation [13,14,21,24,25].

In particular, the advent of the *transformer*-based neural network has dramatically improved the state of the art in most NLP subfields [12,13]. This success comes when these networks are scaled up to relatively large sizes and *trained* on large amounts of human-written text; for example, OpenAI's generative pretrained transformer 2 (GPT-2) model consists of 1.5 billion parameters that are trained on 40 GB of text gathered from various domains [13].

Transformer networks can be used to both analyze and *generate* language given an input text. It is this generative capability that we make extensive use of in this work. This was a relatively unknown capability that recently came to broad impact with the release of ChatGPT [26] in November 2022.

There now exist many such large language models that have been fully *pretrained* on massive corpora of text gathered from several sources on the internet and elsewhere [13,14,27]. A common use case in the field of deep learning and NLP is to take such pretrained models and *fine-tune* [8] them for a specific prediction task that takes language as input. To *fine-tune* a model means to train it on a (typically much smaller) data set to become more proficient at a specific task.

MI Chatbots

There have been several prior attempts to automate MI using a chatbot across different domains, including stress management, sex health education, and smoking cessation [19,20,28-32]; for example, Park et al [28] designed a conversational sequence using MI to aid in stress management. This conversation was deployed to 30 graduate students to compare its efficacy with that of human-to-human MI. It posed thought-provoking questions combined with scripted reflections. Participants reported that they were satisfied with the evocative questions but were dissatisfied with the prewritten reflective statements.

Almusharraf et al [29] and Almusharraf [30], the predecessors of this work, designed and tested an MI chatbot used for motivating smoking cessation. The chatbot used NLP classifiers to select scripted responses that guided a client through the conversation. Both the questions and the reflections in this chatbot were scripted. In a study involving 97 participants, Almusharraf et al [29] found that the average confidence to quit on an 11-point scale increased 1 week after the conversation by 0.8 (P<.001).

He et al [19] created both an MI chatbot and a non-MI chatbot to investigate whether chatbots can motivate smoking cessation. In an experiment with 153 participants, differences in motivation to quit smoking and perceived empathy were compared between a chatbot that used MI and one that did not. Both chatbots used evoking questions and prewritten statements for reflections. There were no significant differences between the chatbots on engagement, therapeutic alliance, or perceived empathy. Notably, for both chatbots, participants reported significantly increased motivation to quit smoking.

The chatbots developed to date used scripted statements based on keyword detection or neural net–based classification of the users' utterances. Scripted responses are often interpreted by humans as generic and repetitive [11], making it possible that the scripted reflections used in prior work contributed to poor user satisfaction and perceived empathy. We hypothesize that a chatbot capable of delivering context-specific MI reflections will better motivate smokers to move toward the decision to quit.

Some works exploring the use of generated reflections within MI-style conversational agents demonstrate this capacity; for example, Shen et al [33] showed how transformer-based models can produce context-specific generative reflections. Although these reflections were used to train practitioners and were not patient facing, they highlight the capacity of transformer-based models to produce good-quality patient-specific reflections. Similarly, Saiyed et al [20] created a technology-assisted MI chatbot for smoking cessation. The chatbot was designed to onboard participants, use MI, and refer participants to human-to-human treatment. It used intent classifiers and transformers to understand and generate utterances, including MI reflections. In a pilot trial of 34 smokers, participants reported that the chatbot had a strong competency in MI but

only scored 3 out of 5 on user satisfaction, leaving room for improvement.

The goal of this study was to determine the impact of several versions of an MI-oriented chatbot, which uses generative reflections, on moving smokers toward the decision to quit smoking. Most of the prior work used scripted statements based on keyword detection or neural net–based classification of the users' utterances. It is possible that scripted reflections caused prior work to score poorly on user satisfaction and perceived empathy. We hypothesize that context-specific MI reflections, such as those developed by Shen et al [33], deployed in a chatbot intervention will better motivate smokers to move toward the decision to quit.

Methods

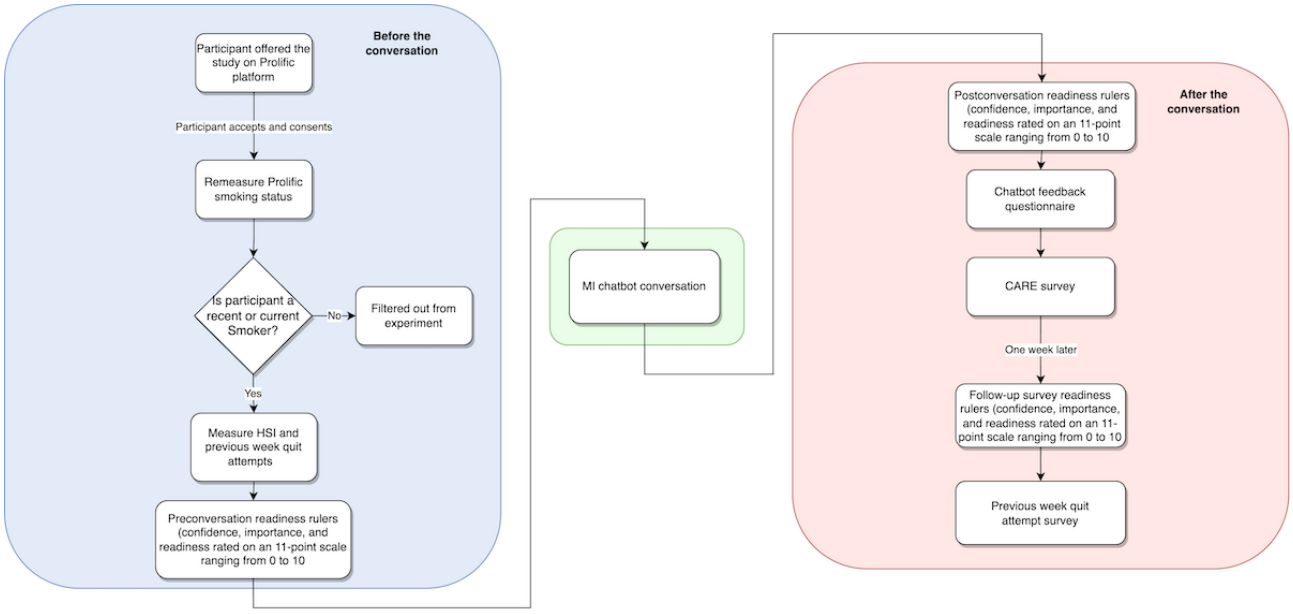
In this section, we describe the overall structure of the MIBot conversation and its several versions and provide the specifics of how the generative model is trained to provide MI-style reflections. We also describe how the effectiveness of the conversation is measured.

Overall Chatbot System Study Design

Figure 1 illustrates the procedure used to evaluate each version of the chatbot. Details of the steps are described in Textbox 1.

The following subsections provide the details of the interaction and evaluation process.

Figure 1. Overall design of each study for each chatbot version. CARE: consultation and relational empathy; HSI: Heaviness of Smoking Index; MI: motivational interviewing.



Textbox 1. Details of the steps involved in the evaluation of the chatbot.

- Participants are recruited through the web-based Prolific [34] paid-recruitment system, the details of which are provided in *Participant Recruitment and Screening* subsection. Each participant who is offered the opportunity to participate (by Prolific) is asked to review an informed consent document. The interaction proceeds if the participant provides consent.
- Participants are taken to a custom website and asked several questions, after which they fill out 3 surveys on readiness to quit, heaviness of smoking, and number of quit attempts, as described in the *Preconversation Surveys and Screening* subsection.
- Next, the participant is presented with a text chat window in which the chatbot, called *MIBot*, begins to interact with them. The conversation begins with an introductory section, after which the participant is asked whether they wish to chat about smoking, as part of a motivational interviewing–style permission-asking approach [5]. If the participant agrees, the conversation continues; otherwise, it terminates.
- The core conversation, described in the *Core Conversation* subsection, ensues.
- After the conversation, the participant is asked to respond to another readiness-to-quit survey, the consultation and relational empathy measure, and other qualitative questions, as described in the *Postconversation Survey* subsection.
- Finally, the participant is directed back to the Prolific system so that it can record the successful conclusion of the tasks.
- One week *after* the end of the conversation, the participant is invited to a second task (also within the Prolific system), which is to answer the survey and questions described in the *One-Week Follow-Up Survey* subsection. Participants are not paid unless they fill out this follow-up survey, and their submission is reviewed for data quality.

Core Conversation

The core conversation consists of 5 open-ended questions that make use of the "running head start" [5] method of MI. The underlying theory of the running head start method is that ambivalent smokers spend very little time thinking about their smoking addiction and do it by habit [35]. Therefore, a key first step is to bring the habit to their attention and ask them to contemplate it. Here, a clinician inquires what participants like about smoking and what they do not like and then uses these reasons as a basis for further discussion and contemplation. This approach is realized through asking these five questions:

1. To start, what is the thing you like most about smoking?
2. Now, what is the thing you like least about smoking?
3. Now, what is one thing about your smoking habit that you would like to change?
4. What will it look like when you have made this change in your smoking habit?
5. Finally, what are the steps you need to take to make this change?

The first 2 questions are based on the running head start approach, and the subsequent 3 questions attempt to stimulate contemplation around the addiction.

One important aspect of this chatbot is that participants respond with *free-form text*. This means that they can provide any textual response in English as opposed to making a selection from scripted responses [10]. In allowing free-form responses, the conversation may more closely align with a human–clinical provider conversation.

In this paper, we will describe and evaluate several versions of the conversation. For most of the versions, the chatbot generates an MI-style *reflection* of the free-form response to each question that is specific to the words of the response, as described in the *Prior Related Work* subsection. Here, we use a fine-tuned transformer-based neural network [13] to generate that reflection. The data and training of this neural network are described in the *Reflection Generation Design and Training* subsection.

After each reflection is provided, the chatbot (in most versions) asks, "Did that make sense?" If the participant responds in a way that indicates *yes*, the chatbot offers a thank you. If the participant's answer is equivalent to a *no*, the chatbot thanks the participant for helping it improve. As the participant can write a free-form text answer, they may respond to the chatbot's question in many ways (eg, offering a longer, possibly corrective answer).

Chatbot Versions

The ongoing goal of the long-term project is to continuously improve the chatbot system through iteration, and, in this paper, we report on the impact that 4 different versions have on readiness to quit smoking in recruited participants. It should be noted that this means that participants are not randomized to different conditions, and the results will be subject to cohort effects. In future work, once we believe that the chatbot is having a sufficient impact, we will measure it against a control. Each new version increases the complexity of the interaction. Table 1 provides a short description of each version. The differences among the versions are more readily seen in Multimedia Appendix 1, which provides an example of the full conversation for each version, taken from actual conversations with participants.

Table 1. Chatbot versions.

Version	Description	Period of experiment
MIBot v4.7	Asks *just* the 5 questions shown in the *Core Conversation* subsection but *does not* provide reflections (instead, responds "Thank you")	July 26, 2022, to August 2, 2022
MIBot v5.0	Asks the 5 questions and provides MI[a]-style reflective answers, as described in the *Core Conversation* subsection; uses the early version of the generator, described in the *Reflection Generation and Training* subsection	August 12 to 19, 2022
MIBot v5.1	Same as v5.0 but uses the significantly improved reflection generator, described in the *Reflection Generation and Training* subsection	August 16 to 23, 2022
MIBot v5.2	On the basis of v5.1, with extensions to the sequence of questions 1 and 2; in addition, if the answer to question 3 relates to *reduction* of smoking, it changes question 4 to be specifically about reduction; furthermore, it extends the interaction around question 3, as described in the *Enhanced Conversation* subsection	November 22 to 29, 2022

[a]MI: motivational interviewing.

Enhanced Conversation Structure

We have described the 4 versions of the chatbot in Table 1, but MIBot v5.2 needs to be presented in more detail because of its increased complexity. The structure of the enhanced conversation is presented in Textbox 2.

Textbox 2. Structure of the enhanced conversation (MIBot v5.2).

- After the reflection is generated for question 1 and question 2, the chatbot asks *what else* the participant likes (or dislikes in the case of question 2) about smoking. If the answer provides more reasons, these are also reflected. If there are no further reasons, then the chatbot moves on to the next question without generating a reflection or asking for validation of the previous reflections.
- If the participant's response to question 3 is related to the *reduction* of smoking, a common answer, then the chatbot switches to a different dialogue stream. The following question arises next: "It's great to hear you want to reduce your smoking. What would it look like when you have reduced your smoking addiction?" After the reflection of the response to this question, the chatbot waits silently for 30 s to encourage the participant to respond on their own. If participants do not respond within 30 s, the chatbot prompts a response by asking the following question: "Could you elaborate on what I said?"

Preconversation Surveys and Screening

Before interacting with the chatbot, to confirm the participant's smoking status and ensure that this status has not changed since the administration of Prolific's own screening survey, each participant is first asked to respond (again) to the same screening question administered by Prolific. Participants who provide a response that is inconsistent with that in their prior Prolific-administered survey, indicating that they do not identify as smokers, are not allowed to proceed with the study.

Next, the participant is asked to respond to a Heaviness of Smoking Index (HSI) survey [36], a validated survey that combines *cigarettes per day* (CPD) and *time to the first cigarette of the day* (TTF). After this, the participant indicates how many quit attempts they had made the previous week, as shown in Figure S1 in Multimedia Appendix 2.

Finally, participants are asked to fill out the readiness ruler [37] survey, a validated tool for tracking progress in MI sessions. Our version is shown in Figure S2 in Multimedia Appendix 2, which rates on an 11-point scale participants' confidence that they could quit smoking now (ie, confidence), their readiness to quit now (ie, readiness), and how important they feel it is for them to quit (ie, importance).

Our protocol imposes a second screen that is based on the readiness ruler responses, as follows: since an MI conversation is targeted toward ambivalent smokers, it is not useful if the participant is already very confident that they will be able to quit; therefore, participants are only included if they have a confidence level of ≤5, with 1 exception: if those with a confidence level of >5 also rate the importance >5 points below this confidence level, there is a contradiction that implies the presence of ambivalence. These participants have confidence that they could quit, but they do not think it is important to do so and hence the contradiction.

Postconversation Survey

After the conversation, the participant is asked to fill out a readiness ruler survey [37] similar to the one they completed before the conversation and then respond to the consultation and relational empathy (CARE) survey [38]. The latter is a validated tool developed to assess empathy in a primary care patient–provider relationship. Empathy in the therapeutic encounter is linked with patient satisfaction and positive health outcomes [39]. The CARE survey examines empathy in the encounter by asking patients to rate the ability of the provider (in this case, the chatbot) to (1) appreciate their perspective; (2) communicate back this understanding; and (3) given this understanding, be helpful to them. The CARE measure has 10 statements that are rated using a 6-point Likert scale with total scores ranging from 0 to 50.

Finally, the participants are asked to respond to the following qualitative questions:

1. What are 3 words that you would use to describe the chatbot?
2. What would you change about the conversation?
3. Did the conversation help you realize anything about your smoking behavior? Why or why not?

One-Week Follow-Up Survey

One week after engaging in the conversation, participants are contacted through the Prolific platform to complete 2 more surveys: the first is a reprise of the readiness ruler survey [37] to determine to determine the effect on the principle outcome after a week has passed, which would be a more permanent effect than immediately after the conversation, and the second concerns 3 questions relating to quit attempts made during the preceding week. The first 2 questions are the same as the last 2 questions (Q3 and Q4) shown in Figure S1 in Multimedia Appendix 2, and the third is given in Figure S3 in Multimedia Appendix 2.

Determining Whether Ambivalence Moved Toward Quitting or Continuing to Smoke

An underlying goal of this work is to help smokers begin to resolve their ambivalence toward smoking. It is possible to classify the data from each participant as belonging to 1 of 3 outcome categories: the participant moved toward changing their smoking addiction (toward quit class), they moved toward maintaining their smoking addiction (toward smoke class), or there was no change (same class).

To place participants into 1 of these 3 classes, we made use of 2 outcome data points. First, we computed the preconversation data to 1-week–later change in confidence from the readiness ruler, which can range in value from −6 to +10. A more positive value suggests a stronger move toward the quit class, whereas a negative value suggests a move toward the smoke class. Second, to gain a more accurate signal, we combined this number with a subjective evaluation of the participant's answer to the third 1-week–later question: "Did the conversation help you realize anything about your smoking behavior? Why or why not?" If the participant stated that they realized something that helped them change their smoking addiction toward quitting and had a positive change in confidence, they were placed in the *toward quit* class. If the participant stated that they realized

that they wished to sustain their smoking addiction, and the confidence change was negative, they were placed in the *toward smoke* class. If neither category fit, we placed them in the *same* class.

Reflection Generation Training

One of the key contributions of this work is the novel way that MI-consistent reflections are generated in response to participant responses to the 5 questions shown in the *Core Conversation* subsection. Here, we made use of recent advances in NLP, and specifically in text generation, as described in the *Prior Related Work* subsection. The reflection generation neural network evolved from the one described in the studies by Ahmed [40] and Ahmed et al [41]. It makes use of the pretrained GPT-2 XL transformer-based neural network model [13], which is *fine-tuned*, as described in the *Prior Related Work* subsection. In this subsection, we provide more detail on how the generators used were trained.

There are 2 versions of the reflection generator that are evaluated in this study. The generators are fine-tuned using example sequences of text, which we call a triplet, consisting of a question, a response, and a reflection.

In version 1 of the generator, the fine-tuning question-and-response data set came from 2 sources: the first was our prior work [40,41], and the second data source was from earlier deployments of MIBot, before the creation of MIBot v4.7. The reflections used came from a variety of sources: from previous versions of this chatbot that were deemed to be acceptable MI reflections by MI-literate researchers or actual reflections produced by MI-literate researchers or MI-expert clinicians.

We used the *hit rate* metric to evaluate the quality of a generator, which is the number of MI-consistent reflections generated divided by the total number of reflections generated. The hit rate was measured on a validation set of question prompts and human responses that did not overlap with the training set. The MI-consistency of the reflection was judged by a single human rater trained in MI literacy. The hit rate of version 1 of the generator was approximately 76% (25/33) on a validation set of 33 prompts and responses. It is important to note that a hit rate of <100% means that some fraction of the generated responses will not be consistent with MI and may indeed make counterintuitive or simply wrong statements. In our experience, the most common type of error is a misstatement of the clear intent of the human; for example, when a user suggested that they would like to quit smoking, the chatbot would sometimes generate a reflection that implied that the user would like to continue smoking.

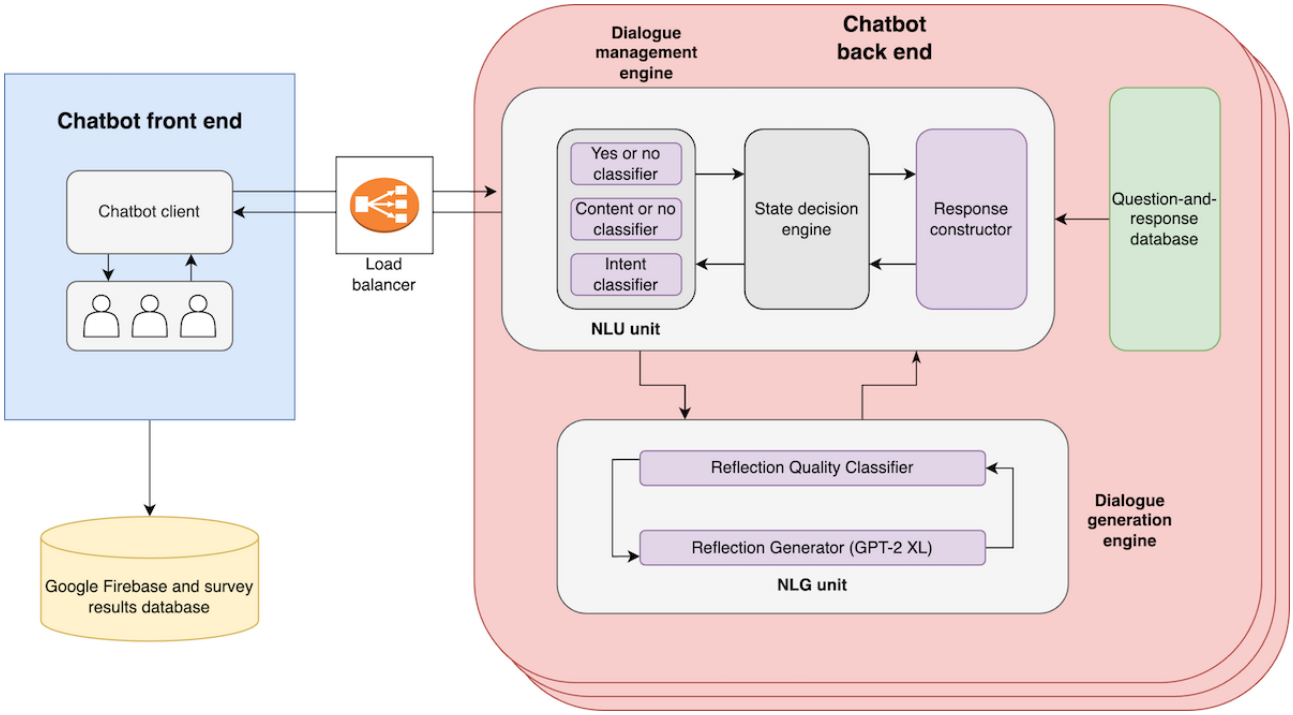
To address the rate of poor reflections, we developed version 2 of the generator with 2 significant enhancements. First, a larger set of 301 fine-tuning triplets were collected over approximately 10 months of deploying the chatbot, making use of the various responses from smokers who had been recruited in a similar manner, as described in the *Participant Recruitment and Screening* subsection. This second data set did not include any of the data from the earlier chatbot version [40,41]. Only MI-consistent reflections were used, which were sourced from MI clinicians, MI-literate researchers, or version 1 of the generator. The labeling and selection of the MI-consistent reflections were improved by using multiple human raters and a carefully controlled decision tree to determine the validity of the reflections. The new rating scheme itself was stricter than the one used in version 1, which caused the hit rate to go down—not because the generation was worse but because of the stricter rating. The hit rate of the new generator was measured to be 55.1% (166/301) on a set of reflections.

The second enhancement was the implementation of a separate *classifier* neural network, trained to determine whether a reflection is MI-consistent, given the prompt, response, and reflection in a triplet. The classifier was used to filter out poor reflections and therefore increase the overall hit rate of generation. This makes use of the fact that all modern neural network–style generators can easily generate many reflections because the generation process is carried out through *sampling* from a probability distribution [8]. The classifier is called the reflection quality classifier, and an earlier version of it is described in the studies by Ahmed [40] and Ahmed et al [41]. This reflection quality classifier, based on the Bidirectional Encoder Representations from Transformers (BERT) [27] pretrained neural network, was fine-tuned using a data set of 740 examples, both positive and negative. Using our validation data, we achieved an accuracy of 70.3% (71/101) on the question, prompt, and response triplets.

Software System

Figure 2 illustrates the structure of the software system used in the studies. Once the Prolific system transfers a participant to our system, they are brought to our chatbot front end, which exists on a web page. This web page connects to a backend database (based on Google Firebase [42]) that records the entire conversation and all data associated with the surveys and information that the participant provides. It also connects to the chatbot backend that contains compute accelerators capable of running the GPT-2 XL neural network in an amount of time that is acceptable to participants (provided by Amazon Elastic Compute Cloud [EC2] [43]). The chatbot backend is split into a dialogue management engine and a dialogue generation engine. The dialogue management engine uses an intent classifier, a yes or no classifier, and a content or no content classifier to classify incoming user utterances using NLU techniques and controls the current state and direction of the conversation. Responses are constructed using a combination of the dialogue generation engine and question-and-response database. The dialogue generation engine uses the GPT-2 XL neural network to generate MI reflections as described in the *Reflection Generation Training* subsection.

Figure 2. Chatbot architecture. GPT-2 XL: generative pretrained transformer 2 extra large; NLG: natural language generation; NLU: natural language understanding.



Participant Recruitment and Screening

Participants were recruited through the Prolific [34] web-based recruitment system and were paid a total of £5 (US $6.25) for completing 2 tasks 1 week apart. The inclusion criteria, based on Prolific's screening filters, were as follows: can be located in any country, minimum age 18 years, fluent in English, smoking status as per 1 of these 2 choices: "I am a current smoker (smoke at least 5 cigarettes a day and have smoked this amount for at least 1 year)" or "I am a recent smoker (smoke at least 5 cigarettes a day and have smoked this amount for <1 year)," and minimum approval rate of 90% on participant's prior Prolific studies.

In addition, the Prolific system was set to recruit an equal number of men and women. However, because there was a subsequent screening of participants as described in the *Preconversation Surveys and Screening* subsection, the final number of participants was not balanced between men and women.

Data from the participants who completed part 1 and part 2 of the study were manually reviewed for data inclusion on the following criteria:

1. Participant properly filled out each survey metric with realistic values.
2. Participant responded to the chatbot with apparent honesty (eg, no toxic language or apparent ulterior motives).
3. Participant met the additional screening criteria as described in the *Preconversation Surveys and Screening* subsection.

Ethics Approval

This research was approved by the University of Toronto Research Ethics Board under protocol number 35567, as amended, on June 29, 2022. All participants provided consent before participating in the study.

Statistical Analysis

Significance testing was completed within and across chatbot versions. Within each chatbot version, we compared readiness ruler responses, quit attempts, and ambivalence resolution counts before and after the conversation. For readiness rulers, a 2-tailed *t* test was applied to examine changes in readiness ruler attributes (ie, readiness, confidence, and importance) from before the conversation to 1 week later. A Fisher exact test was used to evaluate significant changes in quit attempts from before the conversation to 1 week later. We also compared changes in readiness ruler attributes, CARE survey data, and reduction of smoking across chatbot versions. To compare readiness rulers and CARE survey data, a Welch *t* test was used, and to compare changes in reduction of smoking, we used a 2-sample proportion test (*z* test).

For all tests, a significance level of $P<.05$ was considered statistically significant. Statistical analysis was completed using the SciPy library for the Python programming language [44].

Results

Overview

This section reports on the impact of the 4 versions of the chatbot on recruited participants. We begin with the recruitment yield and data inclusion and then provide the demographics of participants. Next, we present the readiness rulers, quit attempts, CARE survey and HSI results, and ambivalence resolution counts.

Recruitment Results

Figure 3 depicts the study procedures (also described in the *Overall Chatbot System Study Design* subsection), showing the points at which participants enter and (may) exit the study. Table 2 presents the specific exit and entry numbers for each version of the chatbot that was deployed. Of the 517 participants who completed part 1 and part 2 of the study, 168 (32.5%) were filtered out across all 4 chatbot versions using the preconversation survey criteria described in the *Preconversation Surveys and Screening* subsection, leaving 349 (67.5%) for inclusion in the analysis.

Figure 3. Flowchart of study procedure.

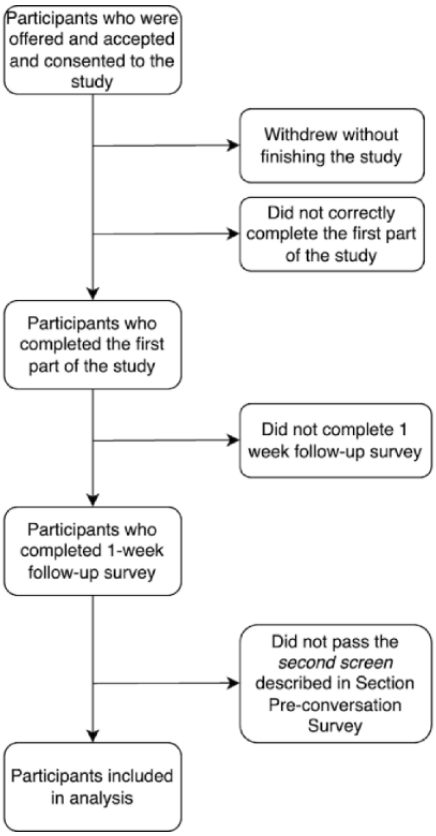


Table 2. Participant count for each chatbot version corresponding to the study flowchart (Figure 3).

Version	Withdrew without finishing the study, n (%)	Did not correctly complete the first part of the study, n (%)	Participants who completed the first part of the study, n (%)	Did not complete 1 week follow-up survey, n (%)	Participants who completed 1 week follow-up survey, n (%)	Did not pass the second screen described in the *Preconversation Surveys and Screening* subsection, n (%)	Values, n (%)
MIBot v4.7 (n=119)	17 (14.3)	9 (8.8)[a]	93 (91.2)[a]	0 (0)[b]	93 (100)[b]	41 (44.1)[b]	52 (55.9)[b]
MIBot v5.0 (n=171)	23 (13.5)	3 (2)[c]	145 (98)[c]	4 (2.8)[d]	141 (97.2)[d]	43 (30.5)[e]	98 (69.5)[e]
MIBot v5.1 (n=169)	24 (14.2)	4 (2.8)[f]	141 (97.2)[f]	1 (0.7)[g]	140 (99.3)[g]	41 (29.3)[h]	99 (70.7)[h]
MIBot v5.2 (n=195)	36 (18.5)	6 (3.8)[i]	153 (96.2)[i]	10 (6.5)[j]	143 (93.5)[j]	43 (30.1)[k]	100 (69.9)[k]

[a]n=102.
[b]n=93.
[c]n=148.
[d]n=145.
[e]n=141.
[f]n=145.
[g]n=141.
[h]n=140.
[i]n=159.
[j]n=153.
[k]n=143.

HSI Survey

Table 3 provides the data from the HSI survey that was applied to the most recent versions of the chatbot (not including MIBot v4.7, for which these data were not collected).

Table 3. Heaviness of Smoking Index (HSI) measures.

Version	Daily number of cigarettes, mean (SD)	Time to first cigarette (answering the question "How soon after you wake up do you smoke your first cigarette?") and count, n (%)	HSI, mean (SD)
MIBot v5.0 (n=98)	11.9 (9.3)	• <5 min: 26 (26.5) • 6-30 min: 15 (15.3) • 31-60 min: 31 (31.6) • >60 min: 26 (26.5)	1.8 (1.5)
MIBot v5.1 (n=99)	11.1 (7.9)	• <5 min: 26 (26.3) • 6-30 min: 27 (27.3) • 31-60 min: 29 (29.3) • >60 min: 17 (17.2)	1.8 (1.5)
MIBot v5.2 (n=100)	9.9 (6.1)	• <5 min: 15 (15) • 6-30 min: 32 (32) • 31-60 min: 18 (18) • >60 min: 35 (35)	1.6 (1.4)

Participant Demographics

Multimedia Appendix 3 provides participants' demographic data, collected by Prolific when participants enrolled on the platform. Prolific allows for these data to be revoked or changed (eg, refer to the attribute *Student Status* in Multimedia Appendix 3).

Readiness Rulers

Participant responses to readiness rulers were collected before the chatbot conversation, immediately after, and 1 week later, as described in the *Preconversation Surveys and Screening*, *Postconversation Survey*, and *One-Week Follow-Up Survey* subsections.

Table 4 provides the means and SDs of *confidence* to quit smoking for each chatbot version before and after the

conversation and 1 week later, as well as the average change after 1 week and its statistical significance.

Table 5 shows the participants' average value for their rated *importance* to quit smoking across each experiment at each collection time, with the same format as Table 4.

Table 6 shows the participants' average value for their rated *readiness* to quit smoking across each experiment at each collection time, with the same format as Table 4.

Figure 4 provides the distribution of confidence scores in each of the 4 versions of the chatbot for each of the 3 collection points. The distributions of the importance and readiness values are provided in Multimedia Appendix 4.

Table 7 presents the number of participants whose confidence increased from before the conversation to 1 week later, the number of participants whose confidence decreased during this period, and the number of participants whose confidence stayed the same.

Table 4. Average confidence before, after, and 1 week after the conversation.

Version	Before the conversation, mean (SD)	After the conversation, mean (SD)	One week later, mean (SD)	Change from before the conversation to 1 week later, mean (SD)	*P* value (from paired *t* test; before the conversation to 1 week later)
MIBot v4.7	3.6 (2.2)	4.5 (2.4)	4.7 (2.6)	1.0 (2.0)	<.001
MIBot v5.0	3.5 (2.7)	4.1 (2.9)	4.7 (2.9)	1.2 (2.0)	<.001
MIBot v5.1	3.2 (2.2)	3.9 (2.1)	4.4 (2.4)	1.3 (2.3)	<.001
MIBot v5.2	3.3 (2.3)	4.1 (2.5)	4.7 (2.7)	1.3 (2.0)	<.001

Table 5. Average importance before, after, and 1 week after the conversation.

Version	Before the conversation, mean (SD)	After the conversation, mean (SD)	One week later, mean (SD)	Change from before the conversation to 1 week later, mean (SD)	*P* value (from paired *t* test; before the conversation to 1 week later)
MIBot v4.7	5.1 (3.1)	5.5 (3.1)	5.3 (3.1)	0.3 (1.6)	.41
MIBot v5.0	5.2 (3.0)	5.7 (3.0)	5.6 (2.8)	0.4 (1.5)	.03
MIBot v5.1	5.2 (2.8)	5.7 (2.8)	5.5 (2.9)	0.3 (1.3)	.17
MIBot v5.2	5.5 (2.9)	6.0 (2.8)	6.2 (2.8)	0.7 (2.0)	<.001

Table 6. Average readiness before, after, and 1 week after the conversation.

Version	Before the conversation, mean (SD)	After the conversation, mean (SD)	One week later, mean (SD)	Change from before the conversation to 1 week later, mean (SD)	*P* value (from paired *t* test; before the conversation to 1 week later)
MIBot v4.7	4.3 (2.7)	4.6 (2.6)	4.8 (2.8)	0.4 (1.5)	.09
MIBot v5.0	4.3 (2.7)	4.4 (2.8)	4.4 (2.7)	0.1 (1.8)	.75
MIBot v5.1	4.4 (2.4)	4.6 (2.4)	4.6 (2.6)	0.2 (1.5)	.14
MIBot v5.2	4.9 (2.8)	5.3 (2.7)	5.4 (2.9)	0.4 (1.7)	.01

Figure 4. Distribution of all 4 versions' confidence values before, after, and 1 week after the conversation.

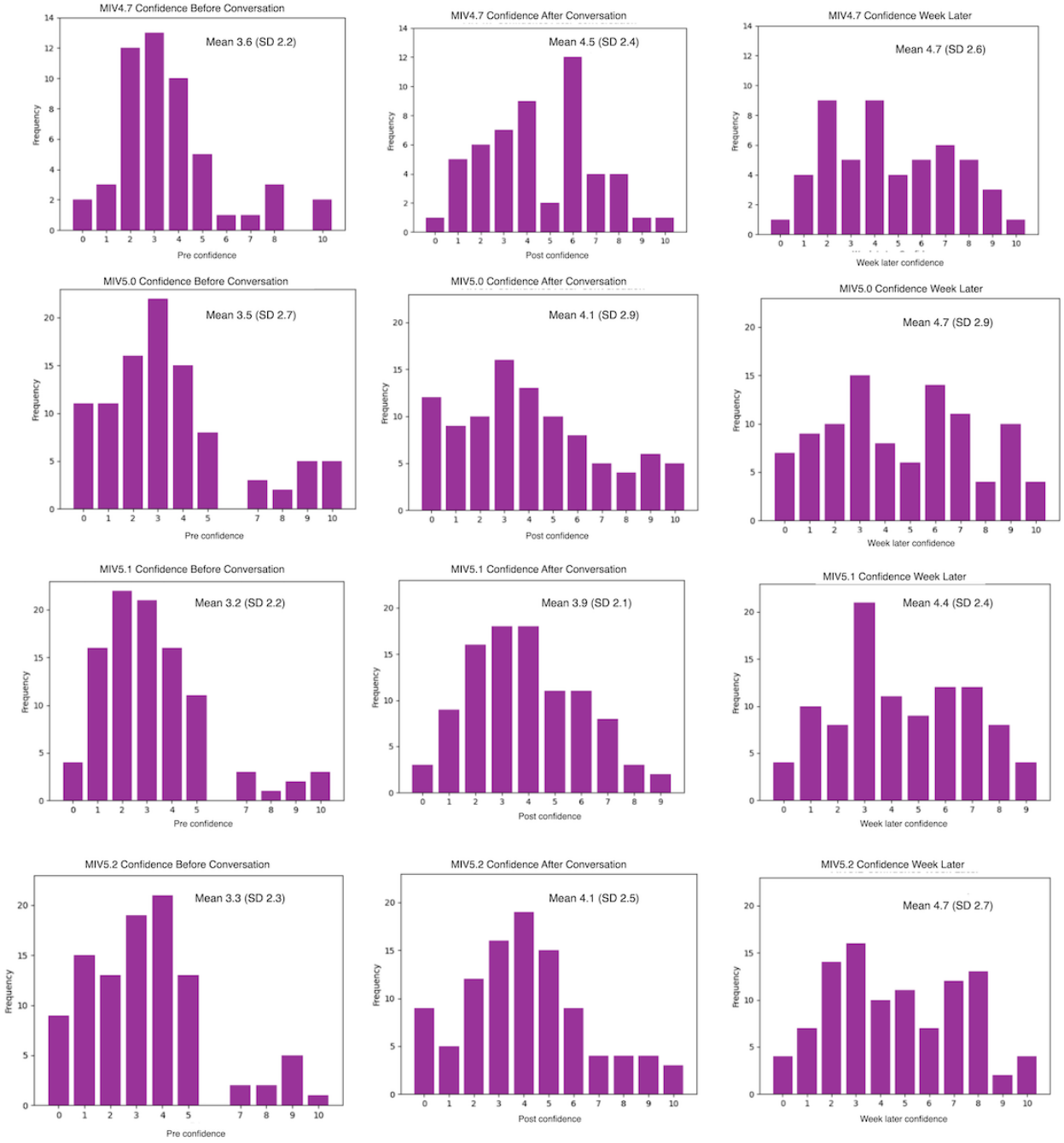


Table 7. Participants whose confidence increased, decreased, or stayed the same from before the conversation to 1 week later.

Version	Confidence increased, n (%)	Confidence decreased, n (%)	Confidence stayed the same, n (%)
MIBot v4.7 (n=52)	31 (59.6)	8 (15.4)	13 (25)
MIBot v5.0 (n=98)	52 (53)	12 (12.3)	34 (34.7)
MIBot v5.1 (n=99)	50 (50.5)	21 (21.2)	28 (28.3)
MIBot v5.2 (n=100)	61 (61)	16 (16)	23 (23)

Quit Attempts and Reduction of Smoking

Table 8 provides the number and percentage of participants, for each chatbot version, who made at least 1 quit attempt (defined as going 24 hours without smoking a cigarette) in the week before engaging in the conversation and in the week after the conversation. It should be noted that our evaluation of MIBot v4.7 did not include a survey for the number of quit attempts before the conversation.

Table 9 shows the count of participants who, after talking to the chatbot, reduced the number of cigarettes they smoke as

well as those who did not reduce the number of cigarettes they smoke. The binary result of *reduce/did not reduce* was determined by setting the result to *reduce* if any one of the conditions listed in Figure S2 in Multimedia Appendix 2 was selected.

Table 8. Participants who made quit attempts before and after the conversation.

Version	Participants with preconversation quit attempt, n (%)	Participants with postconversation quit attempt, n (%)	*P* value (from Fisher exact test)
MIBot v4.7 (n=52)	N/A[a]	18 (35)	N/A
MIBot v5.0 (n=98)	38 (39)	33 (34)	.55
MIBot v5.1 (n=99)	26 (26)	25 (25)	.99
MIBot v5.2 (n=100)	40 (40)	38 (38)	.88

[a]N/A: not applicable.

Table 9. Count of participants who reduced the number of cigarettes they smoke as well as those who did not reduce the number of cigarettes they smoke.

Version	Reduced smoking after talking to the chatbot, n (%)	Did not reduce smoking after talking to the chatbot, n (%)	*P* value (from 2-sample proportion test [*z* test] against MIBot v4.7)
MIBot v4.7 (n=52)	37 (71.2)	15 (28.8)	N/A[a]
MIBot v5.0 (n=98)	68 (69.4)	30 (30.6)	.82
MIBot v5.1 (n=99)	67 (67.7)	32 (32.3)	.66
MIBot v5.2 (n=100)	74 (74)	26 (26)	.70

[a]N/A: not applicable.

CARE Measure

Table 10 provides means and SDs for the CARE survey for each version of the chatbot as well as the results of comparisons between each version and MIBot v4.7.

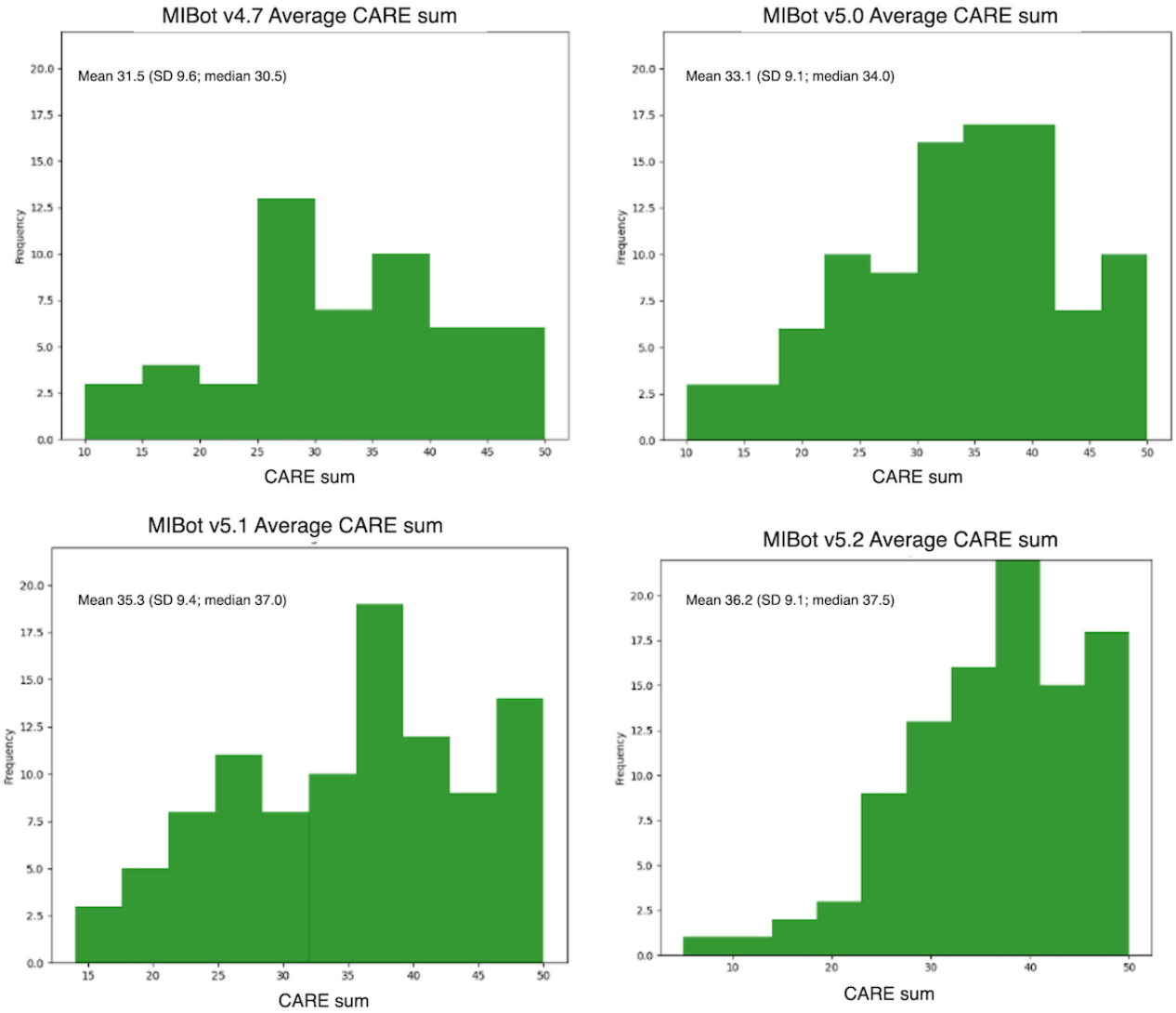
Figure 5 shows a histogram for each average CARE score distribution. Within each plot, the mean, median, and SD values are given.

Table 10. Consultation and relational empathy measure.

Version	Mean (SD)	*P* value (from the Welch *t* test against MIBot v4.7)
MIBot v4.7	31.5 (9.6)	N/A[a]
MIBot v5.0	33.1 (9.1)	.24
MIBot v5.1	35.3 (9.4)	.02
MIBot v5.2	36.2 (9.1)	.004

[a]N/A: not applicable.

Figure 5. Consultation and relational empathy (CARE) survey distribution.



Did Ambivalence Change and in What Direction?

Table 11 presents the counts of participants who were classified as (1) moving in the direction toward quitting, (2) moving toward smoking, or (3) staying the same, as described in the *Determining Whether Ambivalence Moved Toward Quitting or Smoking* subsection. Across all 4 chatbot versions, none of the values in each class were statistically significant. Multimedia Appendix 5 provides the classification for each participant and the raw data upon which the classification was based.

Table 11. Counts of quit, smoke, and same ambivalence classes.

Version	Toward Quit class, n (%)	Toward Smoke class, n (%)	Stayed Same class, n (%)
MIBot v4.7 (n=52)	17 (32.7)	2 (3.8)	33 (63.5)
MIBot v5.0 (n=98)	26 (26.5)	1 (1)	71 (72.4)
MIBot v5.1 (n=99)	20 (20.2)	4 (4)	75 (75.8)
MIBot v5.2 (n=100)	30 (30)	6 (6)	64 (64)

Discussion

Principal Findings

The long-term goal of this work is to evolve a chatbot to have an impact on a smoker's readiness to quit smoking, with a focus in this study on whether generative reflections can improve chatbot efficacy. Although the readiness ruler measures 3 attributes—confidence, importance, and readiness—the confidence measure (which relates to self-efficacy) most successfully predicts quitting success. The more confident someone is, the more likely they are to make a quit attempt and succeed [45-47]. Table 4 shows that all 4 versions of the conversation achieved a statistically significant improvement in confidence 1 week after the conversation took place. The average increase in confidence ranged from 1 to 1.3 on an

11-point scale. This finding is consistent with that of He et al [19] who found that a short chatbot intervention about smoking cessation can have a significant impact on quitting intentions and behaviors. In addition, although the average increase is greater for the later versions (MIBot v5.0 through MIBot v5.2), these are not statistically significant changes between versions of the chatbot (eg, *P*=.43 for MIBot v5.2 vs MIBot v4.7).

Although we hypothesized that generative responses that are specific to what a smoker says would lead to better outcomes, this result suggests that simply asking questions is sufficient to evoke most of the impact on confidence that we observed. However, there is some evidence to suggest that the improvements to the conversation beyond MIBot v4.7 (ie, generative reflections and extended dialogue) had a positive impact on participants' readiness to quit with respect to increases in importance and readiness (refer to Tables 5 and 6, which show that MIBot v5.2 is the only version associated with significant increases in these attributes). In addition, the perceived empathy of MIBot v5.2 is significantly greater than that of MIBot v4.7 (Table 10). This makes intuitive sense because a response that addresses what a person says should be perceived as more empathetic than a response such as "Thank you for answering." Our result contrasts the finding of He et al [19] of no significant difference in perceived empathy between a chatbot that performs MI and one that does not. It is possible that the use of generative reflections (vs the scripted reflections and responses of the study by He et al [19]) is the cause of the difference.

A study by Bikker et al [48] showed that a smoking cessation conversation by human practitioners received a high score on the CARE survey, that is, 46 (with 48% of the nurses achieving a perfect score), which is much higher than the score achieved by MIBot v5.2, that is, an average of 36 (with only 3/100, 3% of the interactions receiving a perfect score). Thus, although much of the benefits for confidence in quit readiness can be attributed to simply asking MI questions, there may be other benefits related to producing generative responses and reflections for importance and readiness, as well as the perceived empathy of the chatbot. These findings are encouraging and support further evolution of this capability.

Recruitment and Demographics

The demographic characteristics of participants in our study (Multimedia Appendix 3) notably differ from those of participants in prior MI intervention studies in 2 ways. First, their mean age (ie, 30 years) is somewhat lower than that of participants in prior studies of human-to-human MI interventions (ie, 35 years) [49]. Second, we balanced our sample of men and women, whereas in many MI studies, approximately 68% of the participants tend to be women [49]. Third, based on the HSI survey (Table 3), participants in our study tended to smoke fewer cigarettes (ie, a mean of 10.8 cigarettes daily) than participants in studies of human-to-human MI interventions (ie, 16 cigarettes on average) [49]. These findings suggest that the participants in our study are younger and overall lighter smokers than those in typical MI studies.

Figure 3 and Table 2 show the number of participants entering each study and the number of exits from the study. The last but one column ("Did not pass the second screen described in the *Preconversation Surveys and Screening* subsection") of Table 2 shows that many of the participants (168/349, 48.1%) did not meet the secondary screening criteria: that they were confident that they could quit smoking and thought that doing so was important. Carpenter et al [50] show that, globally, 20% of smokers are in a similar state, already motivated to quit. It may be that the younger demographic of this study accounts for this difference.

Quit Attempts and Reduction of Smoking

The number of quit attempts related to interacting with each version of the chatbot did not significantly change from the week before the conversation to the week after (Table 8). However, the percentages of participants who attempted to quit across all versions were in the 25% (25/99) to 40% (40/100) range, much higher than the 11% that has been reported occurring 4 to 8 weeks after human MI interventions [49]. This difference in quit attempts may be related to the demographic differences we observed in our sample compared with those in other MI studies. We speculate that the groups in our studies were more likely to make quit attempts because they are a younger and less addicted population, as discussed in the *Recruitment and Demographics* subsection.

Across all conversations, Table 9 shows that a large fraction of the participants (246/349, 70.5%) did make some kind of smoking reduction attempt—meaning that they checked one of the boxes in Figure S3 in Multimedia Appendix 2. However, the differences in percentages were not significant among the different groups and chatbot versions.

Resolution of Ambivalence

We used an alternative measure of the chatbot's impact by classifying participants, based on their ambivalence status, as moving toward quitting, moving toward smoking, or staying the same. There was no significant difference among the chatbot versions in the percentage of participants belonging to each category (Table 11).

It is possible that the participants who resolved their ambivalence toward quitting were just ready to do this and were going to do it anyway or that the conversation was just the push they needed to go there.

It is important to consider the possibility that the 3.7% (13/349) of the participants who resolved their ambivalence to continue smoking were hurt by the interaction with the chatbot. We manually reviewed each of these conversations, and for 85.1% (297/349) of the conversations, we did not see evidence of harmful statements made by the chatbot that could have contributed to this resolution. For the other 14.9% (52/349) of the conversations, the chatbot produced poor reflections, which may have caused participants to be less likely to quit or to believe that they had less of a chance to do so; for example, in response to a participant expressing the idea that quitting *cold turkey* was their best approach to quitting, the chatbot responded, "A smoker can't really do that," which is quite inappropriate.

Limitations

Our findings should be considered in the context of several limitations. First, the self-reported measures used in our study to evaluate the various chatbot versions (ie, readiness ruler, HSI survey, CARE measure, self-reported cigarette consumption, and change in smoking behavior) are potentially less accurate than a clinician-administered survey. Research suggests that participants in health studies tend to underreport unhealthy behaviors and overreport intentions to improve [51]. This tendency may account for some differences in smoking behaviors observed between our sample and other MI studies (eg, quit attempts). In addition, the data suggested that the metric *number of quit attempts* was interpreted differently by different participants; therefore this metric is not particularly reliable.

Furthermore, participants were informed that the aim of the conversation was to help improve the chatbot, which may have led them to respond in what they believed to be a desirable way after the conversation, rather than expressing their true feelings. Participants recruited in this study were also financially compensated, contingent on a review of their responses, which may have led them to agree with statements on the surveys even when they disagreed (ie, acquiescence bias [52]). Although such tendencies would apply to all chatbot versions (and not apply to comparisons among them), they limit the conclusions drawn about pre- to postconversation comparisons. Nevertheless, 1 purpose of the survey administered 1 week later was to give participants time to forget their answers to the initial surveys and to see whether the impact persists over time.

Of the 654 participants who accepted and consented to the study, 105 (16.1%) did not finish the entire study. We speculate that this dropout was caused by several factors: some of the participants may have realized that they were unwilling to discuss their smoking addiction, whereas others may have encountered technical difficulties or became distracted by other tasks because they worked from their own homes.

There is also variance in characteristics among the populations in our chatbot versions. This effect is known as the cohort effect [53] and can be seen in Tables 5 and 6 where we see variation in participant starting values on the readiness rulers. Each population sample has different characteristics and thus has different starting values. This makes comparison among studies difficult because we lose relative significance. In this study, the iterative nature of the chatbot motivated evaluations at different temporal periods. However, to draw appropriate conclusions about the impacts of different versions, future research should randomize smokers to interact with one of the various versions, or with a control in a randomized controlled trial, to eliminate such cohort effects.

Conclusions

In this paper, we have presented a scientific and engineering framework for measuring the effect of an automated conversation on a smoker's readiness to quit smoking. Using this framework, we evaluated how 4 versions of the conversation affected this readiness. We found that simply asking relevant questions about smoking was sufficient to confer benefits on the confidence attribute of the readiness ruler, whereas generated reflections may additionally increase other readiness attributes, while making the chatbot appear more empathetic.

Moving forward, we propose 2 avenues for future progress. First, we intend to design a more complex, longer conversation that uses more aspects of a clinician-delivered MI conversation. We hypothesize that a longer, more nuanced interaction will help to invoke more contemplation in participants. Second, we intend to improve the quality of the reflections generated. More recent models such as GPT-3 [14], ChatGPT [26], and GPT-4 [54] have been shown to be very powerful natural language generating machines, and these capabilities may be leveraged in service of generating more effective responses and reflections, with more clinical impact. If future versions of the chatbot are capable of achieving outcomes comparable with those achieved by human-delivered MI, their widespread delivery has strong potential to help smokers quit smoking sooner, perhaps preventing illness and even loss of life.

Acknowledgments

This research was funded by a Natural Sciences and Engineering Research Council of Canada Discovery grant (RGPIN-2019-04395), a New Frontiers in Research Fund grant (NFRFE-2021-00449), and an AMS Healthcare 2021 Compassion and Artificial Intelligence grant. OM acknowledges receiving salary support through the New Investigator Award at the department of family and community medicine at the University of Toronto.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of the 4 versions of chatbot conversation.
[DOCX File , 17 KB - mental_v10i1e49132_app1.docx]

Multimedia Appendix 2
Chatbot survey screenshots.
[DOCX File , 907 KB - mental_v10i1e49132_app2.docx]

Multimedia Appendix 3
Demographics of participants.
[DOCX File , 15 KB - mental_v10i1e49132_app3.docx]

Multimedia Appendix 4
Importance and readiness values.
[DOCX File , 530 KB - mental_v10i1e49132_app4.docx]

Multimedia Appendix 5
Chatbot feedback and confidence change with ambivalence resolution label.
[DOCX File , 66 KB - mental_v10i1e49132_app5.docx]

References

1. Overview of Canada's tobacco strategy. Health Canada. 2018. URL: https://www.canada.ca/en/health-canada/services/publications/healthy-living/canada-tobacco-strategy/overview-canada-tobacco-strategy.html [accessed 2023-01-26]
2. Canadian Tobacco, Alcohol and Drugs Survey (CTADS): summary of results for 2017. Health Canada. 2018. URL: https://www.canada.ca/en/health-canada/services/canadian-alcohol-drugs-survey/2017-summary.html [accessed 2023-01-26]
3. Wilson SJ, Creswell KG, Sayette MA, Fiez JA. Ambivalence about smoking and cue-elicited neural activity in quitting-motivated smokers faced with an opportunity to smoke. Addict Behav 2013 Feb;38(2):1541-1549 [FREE Full text] [doi: 10.1016/j.addbeh.2012.03.020] [Medline: 22483100]
4. Velicer WF, Fava JL, Prochaska JO, Abrams DB, Emmons KM, Pierce JP. Distribution of smokers by stage in three representative samples. Prev Med 1995 Jul;24(4):401-411. [doi: 10.1006/pmed.1995.1065] [Medline: 7479632]
5. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. 3rd edition. New York, NY: The Guilford Press; 2012.
6. West R. The clinical significance of "small" effects of smoking cessation treatments. Addiction 2007 Apr;102(4):506-509. [doi: 10.1111/j.1360-0443.2007.01750.x] [Medline: 17362283]
7. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966 Jan;9(1):36-45 [FREE Full text] [doi: 10.1145/365153.365168]
8. Jurafsky D, Martin J. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. London, UK: Pearson Education; 2008.
9. Martinengo L, Jabir AI, Goh WW, Lo NY, Ho MR, Kowatsch T, et al. Conversational agents in health care: scoping review of their behavior change techniques and underpinning theory. J Med Internet Res 2022 Oct 03;24(10):e39243 [FREE Full text] [doi: 10.2196/39243] [Medline: 36190749]
10. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
11. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
12. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv. Preprint posted online June 12, 2017 [FREE Full text] [doi: 10.5860/choice.189890]
13. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language Models are Unsupervised Multitask Learners. 2019. URL: https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf [accessed 2023-01-26]
14. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. arXiv. Preprint posted online May 28, 2020 [FREE Full text] [doi: 10.48550/arXiv.2005.14165]
15. Donia J, Shaw J. Co-design and ethical artificial intelligence for health: myths and misconceptions. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. 2021 Presented at: AIES '21; May 19-21, 2021; Virtual Event p. 77 URL: https://dl.acm.org/doi/10.1145/3461702.3462537 [doi: 10.1145/3461702.3462537]
16. Braillon A, Taiebi F. Practicing "reflective listening" is a mandatory prerequisite for empathy. Patient Educ Couns 2020 Sep;103(9):1866-1867. [doi: 10.1016/j.pec.2020.03.024] [Medline: 32487470]
17. Diener I, Kargela M, Louw A. Listening is therapy: patient interviewing from a pain science perspective. Physiother Theory Pract 2016 Jul;32(5):356-367. [doi: 10.1080/09593985.2016.1194648] [Medline: 27351690]
18. Burke BL, Arkowitz H, Menchola M. The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. J Consult Clin Psychol 2003 Oct;71(5):843-861. [doi: 10.1037/0022-006X.71.5.843] [Medline: 14516234]
19. He L, Basar E, Wiers RW, Antheunis ML, Krahmer E. Can chatbots help to motivate smoking cessation? A study on the effectiveness of motivational interviewing on engagement and therapeutic alliance. BMC Public Health 2022 Apr 12;22(1):726 [FREE Full text] [doi: 10.1186/s12889-022-13115-x] [Medline: 35413887]

20. Saiyed A, Layton J, Borsari B, Cheng J, Kanzaveli T, Tsvetovat M, et al. Technology-assisted motivational interviewing: developing a scalable framework for promoting engagement with tobacco cessation using NLP and machine learning. Procedia Comput Sci 2022;206:121-131 [FREE Full text] [doi: 10.1016/j.procs.2022.09.091]
21. Chowdhury GG. Natural language processing. Annu Rev Inf Sci Technol 2005 Jan 31;37(1):51-89 [FREE Full text] [doi: 10.1002/aris.1440370103]
22. Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP '14; October 25-29, 2014; Doha, Qatar p. 1532-1543 URL: https://aclanthology.org/D14-1162.pdf [doi: 10.3115/v1/d14-1162]
23. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv. Preprint posted online January 16, 2013 [FREE Full text] [doi: 10.3126/jiee.v3i1.34327]
24. Malik M, Malik MK, Mehmood K, Makhdoom I. Automatic speech recognition: a survey. Multimed Tools Appl 2020 Nov 10;80(6):9411-9457 [FREE Full text] [doi: 10.1007/s11042-020-10073-7]
25. Storks S, Gao Q, Chai JY. Recent advances in natural language inference: a survey of benchmarks, resources, and approaches. arXiv. Preprint posted online April 2, 2019 [FREE Full text]
26. ChatGPT: optimizing language models for dialogue. OpenAI. 2022. URL: https://openai.com/blog/chatgpt/ [accessed 2023-02-24]
27. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online October 11, 2018 [FREE Full text] [doi: 10.48550/arXiv.1810.04805]
28. Park S, Choi J, Lee S, Oh C, Kim C, La S, et al. Designing a chatbot for a brief motivational interview on stress management: qualitative case study. J Med Internet Res 2019 Apr 16;21(4):e12231 [FREE Full text] [doi: 10.2196/12231] [Medline: 30990463]
29. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing-based chatbot through iterative interactions. J Med Internet Res 2020 Nov 03;22(11):e20251 [FREE Full text] [doi: 10.2196/20251] [Medline: 33141095]
30. Almusharraf F. Motivating smokers to quit through a computer-based conversational system. University of Toronto. 2018. URL: https://www.eecg.toronto.edu/~jayar/pubs/theses/Almusharraf/FahadAlmusharraf.pdf [accessed 1999-11-30]
31. Shingleton RM, Palfai TP. Technology-delivered adaptations of motivational interviewing for health-related behaviors: a systematic review of the current research. Patient Educ Couns 2016 Jan;99(1):17-35 [FREE Full text] [doi: 10.1016/j.pec.2015.08.005] [Medline: 26298219]
32. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J Med Internet Res 2021 Mar 23;23(3):e24850 [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]
33. Shen S, Welch C, Mihalcea R, Pérez-Rosas V. Counseling-style reflection generation using generative pretrained transformers with augmented context. In: Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue. 2022 Presented at: SIGDIAL '20; July 1-3, 2020; Virtual Event p. 10-20 URL: https://aclanthology.org/2020.sigdial-1.2.pdf [doi: 10.18653/v1/2022.acl-long.221]
34. Palan S, Schitter C. Prolific.ac—a subject pool for online experiments. J Behav Exp Finance 2018 Mar;17:22-27 [FREE Full text] [doi: 10.1016/j.jbef.2017.12.004]
35. Bartlett YK, Gartland N, Wearden A, Armitage CJ, Borrelli B. "It's my business, it's my body, it's my money": experiences of smokers who are not planning to quit in the next 30 days and their views about treatment options. BMC Public Health 2016 Aug 04;15(1):716 [FREE Full text] [doi: 10.1186/s12889-016-3395-0] [Medline: 27488047]
36. Heatherton TF, Kozlowski LT, Frecker RC, Rickert W, Robinson J. Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. Br J Addict 1989 Jul;84(7):791-799. [doi: 10.1111/j.1360-0443.1989.tb03059.x] [Medline: 2758152]
37. Readiness ruler. Center for Evidence-Based Practices | Case Western Reserve University. 2021. URL: https://case.edu/socialwork/centerforebp/resources/readiness-ruler [accessed 2023-01-26]
38. Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. Fam Pract 2004 Dec;21(6):699-705. [doi: 10.1093/fampra/cmh621] [Medline: 15528286]
39. Mercer SW, Neumann M, Wirtz M, Fitzpatrick B, Vojt G. General practitioner empathy, patient enablement, and patient-reported outcomes in primary care in an area of high socio-economic deprivation in Scotland--a pilot prospective study using structural equation modeling. Patient Educ Couns 2008 Nov;73(2):240-245. [doi: 10.1016/j.pec.2008.07.022] [Medline: 18752916]
40. Ahmed I. Automatic generation and detection of motivational interviewing-style reflections for smoking cessation therapeutic conversations using transformer-based language models. TSpace. 2022. URL: https://tspace.library.utoronto.ca/handle/1807/123170 [accessed 2023-01-26]
41. Ahmed I, Rose J, Keilty E, Cooper C, Selby P. Generation and classification of motivational-interviewing-style reflections for smoking behaviour change using few-shot learning with transformers. TechRxiv. Preprint posted online June 13, 2022 . [doi: 10.36227/techrxiv.20029880]

42. Home page. Firebase. URL: https://firebase.google.com/ [accessed 2023-01-29]
43. AWS documentation. Amazon. URL: https://docs.aws.amazon.com/ [accessed 2023-01-29]
44. Home page. SciPy. URL: https://scipy.org/ [accessed 2023-01-29]
45. Smit ES, Hoving C, Schelleman-Offermans K, West R, de Vries H. Predictors of successful and unsuccessful quit attempts among smokers motivated to quit. Addict Behav 2014 Sep;39(9):1318-1324 [FREE Full text] [doi: 10.1016/j.addbeh.2014.04.017] [Medline: 24837754]
46. Gwaltney CJ, Metrik J, Kahler CW, Shiffman S. Self-efficacy and smoking cessation: a meta-analysis. Psychol Addict Behav 2009 Mar;23(1):56-66 [FREE Full text] [doi: 10.1037/a0013529] [Medline: 19290690]
47. Von Ah D, Ebert S, Ngamvitroj A, Park N, Kang DH. Factors related to cigarette smoking initiation and use among college students. Tob Induc Dis 2005 Dec 15;3(1):27-40 [FREE Full text] [doi: 10.1186/1617-9625-3-1-27] [Medline: 19570288]
48. Bikker AP, Fitzpatrick B, Murphy D, Mercer SW. Measuring empathic, person-centred communication in primary care nurses: validity and reliability of the Consultation and Relational Empathy (CARE) Measure. BMC Fam Pract 2015 Oct 23;16(1):149 [FREE Full text] [doi: 10.1186/s12875-015-0374-y] [Medline: 26493072]
49. Heckman CJ, Egleston BL, Hofmann MT. Efficacy of motivational interviewing for smoking cessation: a systematic review and meta-analysis. Tob Control 2010 Oct;19(5):410-416 [FREE Full text] [doi: 10.1136/tc.2009.033175] [Medline: 20675688]
50. Carpenter MJ, Hughes JR, Solomon LJ, Callas PW. Both smoking reduction with nicotine replacement therapy and motivational advice increase future cessation among smokers unmotivated to quit. J Consult Clin Psychol 2004 Jun;72(3):371-381. [doi: 10.1037/0022-006X.72.3.371] [Medline: 15279521]
51. Johnson TP, O'Rourke DP, Burris JE, Warnecke RB. An investigation of the effects of social desirability on the validity of self-reports of cancer screening behaviors. Med Care 2005 Jun;43(6):565-573. [doi: 10.1097/01.mlr.0000163648.26493.70] [Medline: 15908851]
52. Kuru O, Pasek J. Improving social media measurement in surveys: avoiding acquiescence bias in Facebook research. Comput Human Behav 2016 Apr;57:82-92 [FREE Full text] [doi: 10.1016/j.chb.2015.12.008]
53. Willets RC. The cohort effect: insights and explanations. Br Actuar J 2011 Jun 10;10(4):833-877 [FREE Full text] [doi: 10.1017/S1357321700002762]
54. OpenAI. GPT-4 technical report. arXiv. Preprint posted online March 15, 2023 2023 [FREE Full text] [doi: 10.48550/arXiv.2303.08774]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
CARE: consultation and relational empathy
CPD: cigarettes per day
EC2: Elastic Compute Cloud
GPT-2: generative pretrained transformer 2
HSI: Heaviness of Smoking Index
MI: motivational interviewing
NLP: natural language processing
NLU: natural language understanding
QA: question answering
TTF: time to the first cigarette of the day

Edited by J Torous; submitted 25.05.23; peer-reviewed by R Wiers, C Heaukulani; comments to author 21.07.23; revised version received 28.07.23; accepted 29.07.23; published 17.10.23.

Please cite as:
Brown A, Kumar AT, Melamed O, Ahmed I, Wang YH, Deza A, Morcos M, Zhu L, Maslej M, Minian N, Sujaya V, Wolff J, Doggett O, Iantorno M, Ratto M, Selby P, Rose J
A Motivational Interviewing Chatbot With Generative Reflections for Increasing Readiness to Quit Smoking: Iterative Development Study
JMIR Ment Health 2023;10:e49132
URL: https://mental.jmir.org/2023/1/e49132
doi:10.2196/49132
PMID:37847539

Review

Digital Health Technology to Support Health Care Professionals and Family Caregivers Caring for Patients With Cognitive Impairment: Scoping Review

Mohamed-Amine Choukou[1,2], PhD; Funminiyi Olatoye[1], MSc; Reg Urbanowski[1], EdD; Maurizio Caon[3], PhD; Caroline Monnin[4], MLIS

[1]Department of Occupational Therapy, College of Rehabilitation Sciences, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada

[2]Centre on Aging, University of Manitoba, Winnipeg, MB, Canada

[3]School of Management of Fribourg, University of Applied Sciences and Arts Western Switzerland (HES-SO), Fribourg, Switzerland

[4]Neil John Maclean Health Sciences Library, University of Manitoba, Winnipeg, MB, Canada

Corresponding Author:
Mohamed-Amine Choukou, PhD
Department of Occupational Therapy
College of Rehabilitation Sciences
Rady Faculty of Health Sciences, University of Manitoba
771 McDermot Ave
Winnipeg, MB, R3E 0T6
Canada
Phone: 1 2043334778
Email: amine.choukou@umanitoba.ca

Abstract

Background: Digital health technology is a promising way of supporting health care providers and family caregivers as they care for patients with cognitive impairment.

Objective: This scoping review aimed to portray the use of digital health technology to assist health care providers and family caregivers in caring for patients with cognitive impairment who live in the community or in a facility.

Methods: We conducted a scoping review of peer-reviewed scientific articles available in MEDLINE, PsycINFO, Scopus, and CINAHL with Full Text, as well as gray literature available in preprint servers, theses depositories, and various national and international dementia organizations' websites. The search yielded 975 articles, of which we included 7 (0.7%) in the review.

Results: Of the 7 interventions included in the retrieved manuscripts, 2 (29%) were digital calendar reminder systems to support activities of daily living and medication management; 2 (29%) were apps on tablet devices to simulate the presence of family before therapy interventions; 1 (14%) was a social robot used in therapeutic sessions to include elements of musicotherapy, reminiscence, cognitive games, and relaxation; 1 (14%) was a commercially available computer system that provides access to various recreational leisure activities; and 1 (14%) was a web-based self-management support system that helps family caregivers to deal with behavior changes in a relative with dementia. Of the 7 articles, only 1 (14%) reported on the use of a behavior change theory, namely a comprehensive process model of engagement coupled with cognitive stimulation therapy.

Conclusions: Literature on the topic is scarce, recent, and heterogeneous. There is a clear need for a theoretical framework to conceptualize and govern the use of behavior change models that incorporate technology for patients with cognitive impairment.

(JMIR Ment Health 2023;10:e40330) doi:10.2196/40330

KEYWORDS

digital health; behavior change; mental health; cognitive impairment

Introduction

Background

The aging of the population will intensify in the coming decades, increasing the prevalence of disease and disability, in particular, the impairment of cognitive functions [1]. In Canada, for example, approximately 1 in 4 seniors aged ≥85 years have been diagnosed with dementia [2]. The growing number of people with cognitive impairment will result in high demands on health care systems. Cognitive impairment reduces the quality of life of older adults and increases the risk of dementia and mortality [3,4]. In this paper, the term *cognitive impairment* refers to the various etiologies known to characterize cognitive impairment, such as vascular conditions, neurological conditions, and stroke. The ability of people with cognitive impairment to retain information, learn new things, or concentrate is impaired, leading to a decrease in autonomy and independence and poor social participation. Cognitive impairment affects the patients and those around them, particularly family caregivers and the health care professionals providing care to alleviate the condition [5,6]. Family caregivers' activities require a constant connection to the person with cognitive impairment and the activities they require to perform to function and live safely (eg, maintain a career, home, and family life). The burden experienced by people taking care of a person with cognitive impairment is associated with an adverse emotional state and can be more substantial in terms of physical sequelae and financial and social consequences [7].

Responsible health and social systems should support caregivers to help maintain persons with cognitive impairment in their homes for as long as possible. From the health care provider's perspective, the delivery of quality patient-oriented care seems to be determined by their ability to understand the particular care and heterogeneous communication needs of people with cognitive impairment [8,9]. Digital health technology is a promising way of supporting patients with cognitive impairment in that technology can be an enabler of behavior change in patients, family caregivers, and health care professionals. One could imagine how technology may foster communication among users (patients, caregivers, and care providers) and affect daily routines among all stakeholders. This paper examines the use of digital health technology as a plausible way of supporting health care providers and family caregivers as they care for people with cognitive impairment. According to the US Food and Drug Administration, digital health technologies "use computing platforms, connectivity, software, and sensors for health care and related uses" [10]. Digital health technology has been deemed valuable and efficient in supporting health care providers and family caregivers in different populations, such as children with special health care needs [11], patients undergoing hematopoietic cell transplantation [12], and patients with Parkinson disease [13], as well as patients requiring cardiovascular care [14], pulmonary rehabilitation [15,16], mental health rehabilitation [17], and cognitive rehabilitation [18]. Digital health technology faces challenges at individual and system levels, such as lack of presence and in-person contact [19] as well as ethics and data governance [20]. The design of a digital health technology for patients with cognitive impairment is itself challenging because of their complex needs and because they may be in no condition to engage in the design processes. Creative technologists should be aware of the importance of actively including the end users in the design process [21] and that the involvement of end users with cognitive impairment can be a complex task [22], which is even more critical when older adults are the end users [23]. When designing a digital health technology that persons with cognitive impairment will use, the task is cumbersome for many reasons related to awareness about their health needs, knowledge of potential ideas based on technology, and level of technology literacy.

In practice, digital health technology is understudied in people with cognitive impairment, family caregivers, and health care providers. This paper focuses on information and communication technologies deployed with or without additional digital technologies to look more closely at digital health technology that claims to support health care professionals and family caregivers caring for patients with cognitive impairment by providing them with digital support and a sense of presence using technologies such as, but not limited to, anthropomorphic agents, digital tools, robots, apps, and multimedia systems. The typical intervention sought in this review would be based on technology that allows digital dyadic communication between patients with cognitive impairment and their health care providers and family caregivers.

Objectives

In this paper, we aimed to depict the use of digital health technology to assist health care professionals and family caregivers in caring for patients with cognitive impairment who live in the community or in a facility. Recent literature portraying available technology applications to support the informal caregiver of a person with cognitive impairment showed the impact of technology in reducing the caregiver's burden and the patient's social isolation [24]. According to the literature, social support is still needed to improve the adherence and effectiveness of digital health technology–enabled telerehabilitation [25]. Although studies show an impact of technology on occupational performance and quality of life in community and health care settings [26,27], more research on the impact of digital health technology settings [26,27] and users' existing knowledge and informational needs [25,27] as well as feasibility and acceptability [25] is needed. Furthermore, Bell et al [28] advocate for including motivational frameworks and behavior change interventions in digital health intervention development, increasing service engagement in young people and lived experience involvement in digital intervention development. The objectives of this paper were to (1) identify which digital health interventions were tested in patients with cognitive impairment, health care professionals, and family caregivers, as well as the associated behavior change strategies; (2) identify the behavior change theories involved in the included literature that relates to patients, health care providers, and family members; (3) provide an overview of the effects of these interventions on patients with cognitive impairment, health care professionals, and family caregivers; and (4) identify and comprehend the benefits, challenges, and influencers of using the digital health technology to assist health care professionals

and family members caring for patients with cognitive impairment.

Methods

Design

We performed a scoping review of peer-reviewed scientific articles following the recommendations of Arksey and O'Malley [24] and Levac et al [25]. We followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [26]. Information was extracted from the included manuscripts using customized calibrated forms that were tested by the team before their use. Data charting was performed by 1 coauthor per paper and was validated by a second team member. There was no consultation exercise at the end of this scoping review.

Search Strategy

The dementia and cognitive impairment search concept was adapted from the Cochrane Dementia and Cognitive Improvement group [27]. To develop the search strategy, a health sciences librarian (CM) consulted with the team (MAC, FO, RU, and MC). Using a combination of controlled vocabulary and keywords, the team developed search concepts for telepresence and behavior changes (Multimedia Appendix 1). The telepresence search concept incorporated Medical Subject Headings (MeSH) terms and keywords representative of robotics and telecommunications, including the MeSH terms *Artificial Intelligence*, *Human-Robot Interactions*, *Self-help Devices*, *Wearable Electronic Devices* and *Robotics*. The behavior changes concept included the MeSH terms *Health Behavior*, *Self-control*, *Self-examination*, and *Health Risk Behaviors*. In addition, we excluded articles indexed as child or infant only and limited them to English.

To locate gray literature, the librarian searched in preprint servers, theses depositories, and various national and international dementia organizations' websites with the keywords *dementia* or *cognitive impairment and technology*. The search extraction took place in April 2021 in the following databases: MEDLINE (Ovid; 1946-2021), PsycINFO (Ovid; 1806-2021), Scopus (1970-2021), and CINAHL with Full Text (EBSCOhost; 1981-2021). All records were exported to EndNote (version X9; Clarivate), and duplicates were removed [29]. The screening process was completed in Rayyan software [30].

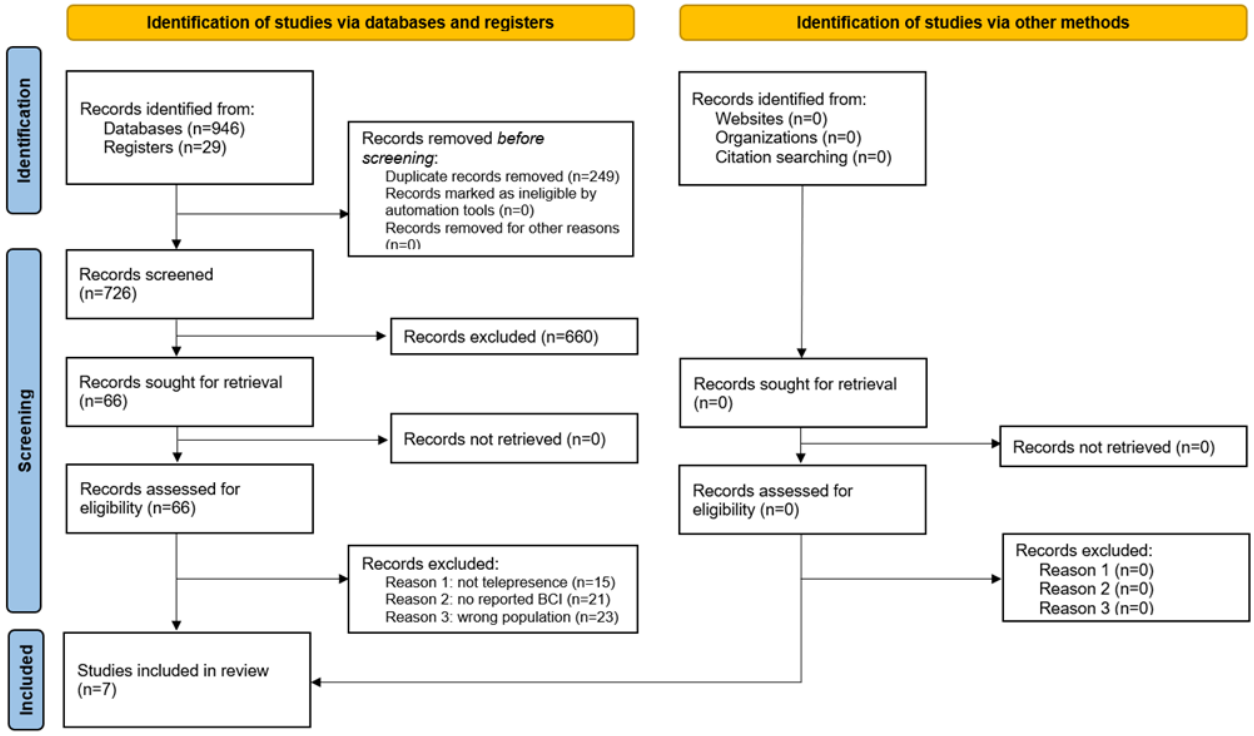
Inclusion and Exclusion Criteria

To be included, the articles had to be written in English and present original qualitative or quantitative empirical data related to the use of a digital health technology (eg, a mobile app or a telepresence robot) to support health care delivery, family caregiving, and behavior change in patients with cognitive impairments. The articles had to refer to a part of, or an entire, behavior change approach and the related qualitative and quantitative instruments used to measure a patient's behavior change, with reference to the behavior change technique taxonomy formulated by Michie et al [31]. The included articles had to be focused on adults with cognitive impairment (eg, dementia, traumatic brain injury, or stroke) and their health care providers and family caregivers. Search results reporting on surveys, technical development, perspectives, and literature reviews were excluded. Two independent reviewers performed the study selection. An article had to be judged relevant by both reviewers to be included. A third reviewer was consulted when the 2 first reviewers could not reach a consensus.

Study Selection Process

The search yielded 975 articles, with 726 (74.5%) remaining after duplicates were removed (Figure 1). After screening titles and abstracts, of the 726 articles, 66 (9.1%) were retained and read in their entirety to determine eligibility. As a result, 89% (59/66) of the articles were rejected for various reasons (Figure 1). Thus, of the initial 975 articles, 7 (0.7%) were finally included in this scoping review.

Figure 1. Study selection process.



Data Extraction and Data Analysis

We gathered and charted the following information: the technology-enabled presence and telepresence interventions; technology used; profiles of the patients with cognitive impairment, family caregivers and health care professionals; outcome measures; intervention outcomes; and barriers to intervention implementation. Information about the factors affecting the success of digital health technology implementation was extracted from the discussion and limitations sections of the retrieved manuscripts and underwent thematic analysis. Two reviewers independently extracted the data, and all team members revised the extracted information and adjusted or completed the extraction if necessary. After a series of iterations and content validation, we charted the data and presented them in a tabular format in Microsoft Excel 2016.

Results

Characteristics of the Included Studies

The included studies were published between 2012 and 2020 in the United States (2/7, 29%), Canada (1/7, 14%), Mexico (1/7, 14%), Sweden (1/7, 14%), the Netherlands (1/7, 14%), and Japan (1/7, 14%). Of the 7 studies, 6 (86%) were peer-reviewed research articles, and 1 (14%) was a thesis presented to obtain the degree of doctor of occupational therapy [32]. Of the 7 manuscripts, 6 (86%) presented qualitative and quantitative information from small-scale empirical studies involving 2 to 18 participants in each profile studied (patients with cognitive impairment, health care providers, and family caregivers) [32-37]. The manuscripts centered around adults and older adults with Alzheimer disease, dementia, and Parkinson disease dementia living in various settings, such as geriatric residence, memory care unit, and home. Only 14% (1/7) of the studies presented data from a randomized controlled trial involving 81 family caregivers [38].

Digital Health Interventions: Characteristics and Outcomes

Table 1 describes the included digital health interventions and their effects on patients with cognitive impairment, health care professionals, and family caregivers. As per the objective of the scoping review, the included manuscripts presented data from interventions targeting patients and health care providers (3/7, 43%) [33-35]; patients and family caregivers (1/7, 14%) [32]; patients with cognitive impairment, health care providers, and family caregivers (2/7, 29%) [36,37]; and family caregivers only (1/7, 14%) [38].

The included studies involved the following presence and telepresence interventions, all detailed in Table 1:

- Two digital calendar reminder systems to support activities of daily living [33] and medication management [36]
- Two apps on tablet devices to simulate the presence of family before therapy interventions [35] and to reduce agitations and support caregivers' education [32]
- A social robot used in therapeutic sessions to include elements of musicotherapy, reminiscence, cognitive games, and relaxation [34]
- A commercially available computer system that provides access to various recreational leisure activities to encourage the engagement of people with dementia in activities and social interactions [37]
- A web-based self-management support system that helps family caregivers deal with behavior changes in a relative with dementia [38]

Table 1. Intervention description and effects on patients with cognitive impairment, health care professionals, and family caregivers.

Technology used	Intervention description	Patients	Health care professionals	Family caregivers	Outcome measure	Outcome	Reference number
Digital calendar with active reminders to support activities of daily living—description: RemindMe, a digital calendar developed to support people with cognitive impairment that includes three core functions: (1) scheduling of activities and customizing reminders in a user-friendly digital calendar, (2) active confirmation of reminders sent by SMS text message requiring users to acknowledge the prompt actively by responding to the SMS text message, and (3) a feedback system that registers information based on the user's interaction; use: stand-alone	Patients received an individualized introduction, a written manual, and individual weekly conversations for 2 months with follow-up assessments after 2 and 4 months. All occupational therapists received training. The patients received rehabilitation as usual at the rehabilitation clinic together with structured support in the use of RemindMe as a support for memory and to plan and structure activities of daily living; setting: rehabilitation clinics	Eight patients (6 women) with a median age of 58 (range 26-68) years from 3 rehabilitation clinics in Sweden, who had cognitive impairment and needed support with planning, organizing, and remembering to perform activities of daily living	Seven female occupational therapists, specialists in rehabilitation for people with neurological impairments in primary care, with 2 to 25 years of experience; 5 had a bachelor's degree in occupational therapy, and 2 had a bachelor's degree with 1 year's postgraduate education; all were experienced in using digital health technology for their patients, but no one was specialized in developing information and communication technologies	N/A[a]	• Patient satisfaction (assessed with QUEST 2.0[b]) • Field notes (patients' opinions or interests or intentions) • Logs in the calendar • Interviews with occupational therapists (appropriateness, patient ability, implementability, and fit of RemindMe within the infrastructure) • Data regarding the cost for patients to use RemindMe	• Acceptability: +[c] • Demand: + • Implementation: challenges • Practicality: challenges • Integration: +	[33]

Technology used	Intervention description	Patients	Health care professionals	Family caregivers	Outcome measure	Outcome	Reference number
Social robot Eva—description: an assistive anthropomorphic robot used in therapeutic sessions to include elements of musicotherapy, reminiscence, cognitive games (complete with wisdom sayings), and relaxation; use: stand-alone	Cognitive stimulation therapy that aims to actively mentally stimulate people with dementia through cognitive activities and reminiscence, multisensory stimulation, and group social contact. The cognitive stimulation therapy included a set of therapeutic sessions conducted by the robot Eva. The study was conducted in a lounge of the geriatric residence where all participants live; setting: nursing home	Nine people (6 women) with dementia living in a geriatric residence, aged between 74 and 95 (mean 83.77, SD 8.13) years	Six caregivers (4 women) participated in the study; they had an average of 3.2 years of experience as caregivers in a geriatric residence	N/A	• Frequency and intensity of problematic behaviors (assessed with the NPI-NH[d]) • Qualitative analysis	• Decrease in frequency and intensity of problematic behaviors, agitation, depression, delusion, apathy, and irritability (–[e]) • Behavior and mood change: cooperation, positive mood, less isolation, and positive facial and corporal expressions (+) • Activities of daily living: feeding improvement, activity-level change, fewer naps, and participation in residence activities (+) • Socialization: talkative, singing and laughing, and influence on other residents (+) • Prevalence of the impact: fewer problematic behaviors, impact on caregivers' burden, and sustained impact on behavior (+)	[34]
Tablet device (large screen size to accommodate potential visual impairment)—description: tablet-simulated presence therapy intervention; use: human enacted	Family caregivers asked to self-record a personalized 1-minute video based on how they would typically communicate with the study participant. Family videos are recorded either directly into the tablet device on the unit or at home using a personal device and later uploaded to the tablet device. Content tailored to help with particular challenges (eg, medication refusal and responsive [aggressive] behaviors); setting: hospital	Four hospitalized older adults (3 women) with dementia, aged 69 to 80 years. They had different types of diagnoses, including Alzheimer disease, vascular dementia, and Parkinson disease dementia	Two experienced full-time nurses who had basic training in dementia care. One had practiced 14 years in dementia care, and the other had 18 years of experience	N/A	• Filmed footage and interviews		[35]

Technology used	Intervention description	Patients	Health care professionals	Family caregivers	Outcome measure	Outcome	Reference number
						• Patients: positive changes in the mood of all 4 older adults (+) • How health care providers and health care professionals used the tablet device to interact with the patients significantly affected the outcomes. The care interaction among the patients required a person-centered approach. When a video delivered an effective message, it bridged the connection and helped older adults with dementia feel safe (+)	
Tablet device—description: apps for reducing agitations and supporting caregivers' education; use: human enacted	Tablet device–based program for family caregivers' education and reducing problem behaviors in older adults with cognitive impairments. Four apps were used for the agitation-reduction component of the program. For the educational component, meetings or email correspondence with family caregivers occurred to discuss important information about dementia and how to interact with people with dementia; setting: care unit	Five older adults with dementia	N/A	Eight family caregivers received educational information; data were only collected from 6 of the participants	• Participants' scores on the Agitated Behavior Scale • Interviews	• Patients: a notable reduction in agitation during the use of the apps (–) • Family caregivers: knowledge about specific aspects of dementia increased significantly after participating in the program (+)	[32]
		Five older adults residing in the MCU, aged ≥50 years	Seven health care professionals in the MCU, aged ≥18 years, interact directly with the individuals in the MCU		• Interviews with health care professionals and family caregivers	• Health care professionals and family members reported benefits for residents, such as enjoyment, interactions and connections with others, and mental stimulation (+)	[37]

Technology used	Intervention description	Patients	Health care professionals	Family caregivers	Outcome measure	Outcome	Reference number
A computer system designed for older adults in community settings—description: a commercially available system that encourages the engagement of people with dementia in activities and social interactions by providing access to various recreational leisure activities; use: human enacted	The system was made available for use, as part of a 6-month study, in a 26-apartment MCU[f] for people with mild-to-severe dementia. The system was placed in the dining room of the MCU and was used there by health care professionals. It was wheeled into a side room or resident's room by the researchers for individual sessions; setting: care unit			Four family caregivers of older adults living in the MCU, aged ≥18 years, who have visited their relative residing in the MCU at least monthly in the year preceding the study			
Medication reminder device—description: an automatic pill dispenser with audible and visual stimuli that remind users when to take their medication. When the alarm rings, the correct dose of medication is released into the lid opening. Users must then invert the device to obtain medication and stop the alarm; use: human enacted	If participants did not find any difficulty using the device, the device and its use were customized. Customized conditions included medication loaded into the device, loading schedule, location in the home, time of the alarm, and other individual considerations. The caregivers monitored device use during the first week of its use. They were asked to provide minimal assistance to users in using the devices (only when required); setting: home care ad community care	A total of 18 older adults (15 women) with dementia (mean age 81.2, SD 6.2, years), living at home, with a history of missed medication doses, overdoses, or need for verbal reminders to take medication once or more during a week	Eight caregivers and visiting nurses	Ten family members living separately	• Self-administration medication rate is defined as the ratio of the number of doses taken independently to the number of all prescribed doses during 1 week • Open questioning of users and their caregivers	• Self-administration medication rate during 1 week showed improvement at 3 months (13/18 users; +) • Reminder devices can improve medication adherence (+) • Caregivers reported maintenance of normal blood pressure, reduction of caregivers' burdens, and decreased care costs (+) • Users reported gaining self-confidence and success at learning the skills necessary to use the device (+)	[36]
Web-based self-management support—description: web-based self-management support intervention to help family caregivers deal with behavior changes in a relative with dementia; email contact of a specialist dementia nurse and web-based videos and e-bulletins; use: human enacted		N/A	N/A	A total of 81 family caregivers (partner or relative) of people with dementia who live at home. They had contact with the person with dementia at least once a week			[38]

Technology used	Intervention description	Patients	Health care professionals	Family caregivers	Outcome measure	Outcome	Reference number
	Family caregivers received 3 personal email contacts with a specialist dementia nurse (during 12 weeks). The nurse supported the family caregivers in managing behavior changes by giving feedback on assignments and tailoring support to the personal needs of the family caregivers. Other family caregivers received links to 6 web-based videos with assignments about different types of behavior changes, and they could choose how many videos they watched and assignments they completed. A third group of family caregivers received 6 e-bulletins containing practical information about different types of changes in behavior and how to manage them; setting: web based (caregiver support)				• Primary outcome variable (self-efficacy) was measured using the Trust in Our Own Abilities instrument, a questionnaire in Dutch • Presence and reaction scores for mood and behavior problems, measured by the Revised Memory And Behavioral Problem Checklist • Occurrence of disruptive behavior and family caregivers' reaction • Dyadic Relationship Scale used to assess relationship between person with dementia and family caregiver	• The web-based self-management support intervention involving email contacts did not lead to positive effects compared with web-based interventions without personal email contacts (–) • The medium intervention involving web-based videos and e-bulletins showed no statistical improvements compared with the minor intervention involving e-bulletins only (no significant change)	

[a]N/A: not applicable.

[b]QUEST 2.0: Quebec User Evaluation of Satisfaction With Assistive Technology, version 2.0.

[c]+: positive effect.

[d]NPI-NH: Neuropsychiatric Inventory, Nursing Home version.

[e]Negative effect.

[f]MCU: memory care unit.

Overall, using multiple methods and designs, the studies showed positive effects on patients with cognitive impairment, health care providers, and family caregivers (Table 1), without adverse effects on any digital health technology users. The studies reported some barriers faced by health care providers and family caregivers to implementing and using digital health technology.

Behavior Change Theories Involved in the Included Literature

The study by Cruz-Sandoval et al [34] was the only one reporting on using a behavior change theory, which consisted of applying a comprehensive process model of engagement and cognitive stimulation therapy. None of the other studies explicitly applied a behavior change theory.

What Factors Influence Digital Health Technology's Success in Assisting Health Care Professionals and Family Caregivers Caring for Patients With Cognitive Impairment?

Thematic analysis of the information extracted from the discussion and limitations sections of the included papers about the factors influencing digital health technology success revealed 18 factors. Table 2 summarizes these factors into 3 categories: benefits, challenges, and influencers of digital health technology success in each of the 3 groups: patients with cognitive impairment, health care professionals, and family caregivers.

Table 2. Benefits, challenges, and influencers of using digital health technology to support patients with cognitive impairment, health care professionals, and family caregivers.

	Benefits	Challenges	Influencers

Patient with cognitive impairment	• Engagement in the activity
	  • Habit forming [33]
	  • Gaining self-esteem, motivation, and "having something to which to look forward" [37]
	  • Technology replaced or augmented activities [37]
		• Relevance and quality of the content
		  • Video content does not have to be older for residents to relate to it [37]
		  • Older content may be frustrating ("Residents were frustrated knowing they should recognize materials from a certain era but not being able to do so") [37]
		• Technical issues
		  • Technical issues and lack of user-friendliness of the system are frustrating for people with dementia [37]
		  • Improper operation of the device may be a reason for cessation [36]
		• Patient-centered content
		  • Applications designed for the general population and ill-suited for patients with cognitive impairment [37]
		  • Gaming stations not suited to disease-related disabilities (eg, wheelchair vs access to station and station height) [37]
			• Planning and delivery of activities
			  • Whether the system was used one-on-one or in a group (intervention would probably be more beneficial in a one-on-one setting or small group) [37]
			  • Frequency of use [37]

Health care professional	• Support and resources
	  • Equipped, trained, and having access to educational resources [32]
	  • The use of technology, specifically the iPad, throughout the program provided an easy and convenient way to deliver both interventions and education [32]
		• Workload
		  • Weekly telephone calls during a more extended period can become time consuming and problematic to implement in practice [33]
		  • Telephone calls might reveal other, new, problems, and therefore frequent telephone calls can be problematic to implement in daily practice [33]
		  • Lack of resources, especially time and personnel [37]
		• Technology use
		  • The difficulty of keeping up to date with rapid technological development [33]
		  • Knowing how to use a tablet device and video applications [35]
		  • Not knowing how to use features or where applications were located [37]
		  • Limited previous experience in providing self-management support through email contacts [38]
		• Attractiveness of content
		  • Boredom caused by limited content (too few episodes of television shows or movies) [37]
			• Involvement and professional development
			  • Involvement as a facilitator and motivator [37]
			  • Instruction and increased awareness of the importance of the integrated use of the various elements in an intervention [38]
			  • Adding information about the apps and creating case studies [32]
			• Quality of content and convenience of technology
			  • Service accessed from various computers [37]
			  • Convenience and availability of information and media [37]
			  • Collaboration with families to improve care [35] (ie, building video content)
			• Sustainability
			  • Long-term use: the health care professionals would have everything they need in 1 place for both education and intervention purposes [32]
			• Cost
			  • Apps can be downloaded for little to no cost [32]

	Benefits	Challenges	Influencers

Family caregiver	• Involvement
	  • Involvement leads to a better relationship [35]
		• Collaboration needed with a health care provider
		  • The family caregiver asked the occupational therapist for support in deciding about content building (what to include in videos) [35]
			• Involvement
			  • Involvement leads to a better relationship as well as safer and higher quality of care [35]
			  • Involvement as a facilitator and motivator [37]
			  • Caregiver's help is a prerequisite for the use of the medication reminder device [36]
			  • Technology should be developed as a technical aid for use with family caregivers' help [36]
			• Resources and skills
			  • Family caregivers all had internet access and were often relatively young and well educated [38]
			  • Increased comfort with technology and finding information makes family caregivers more self-assured with accessing and using knowledge about dementia [32]
			• Ease of use of the technology
			  • Adding information about the apps and creating case studies [32]

How Does Digital Health Technology Support Patients With Cognitive Impairment?

Although digital health technology encourages patients to participate in activities (videoconferencing and visualization of memories through photos and videos), they may encounter technical difficulties, find content less relevant than expected, and be less adapted to their conditions. The quality of activity planning seems to be a determinant of the success of digital health technology implementation.

How Does Digital Health Technology Support Health Care Professionals' Work With Patients With Cognitive Impairment?

Although health care providers feel supported and resourceful when incorporating digital health technology into their interventions, they may face technical difficulties, find content less appealing than expected, and feel burdened by the increased workload. The success of digital health technology implementation seems to be influenced by health care providers' active participation in the development and integrated use of the technology, collaboration with families, the technical quality of the digital health technology, and its convenience and cost.

How Does Digital Health Technology Support Family Caregivers of Patients With Cognitive Impairment?

Although family caregivers believe that their involvement in the therapy provided to their relative leads to a better relationship, they may require more support from, and collaboration with, health care providers. The success of digital health technology implementation seems to be influenced by family caregivers' participation as active facilitators in the delivery of care, the ease of use of digital health technology, access to the internet, and the caregivers' literacy and technology literacy.

Discussion

Principal Findings

This scoping review aimed to portray the use of digital health technology to assist health care professionals and family caregivers caring for patients with cognitive impairment who live in the community or in a facility. Literature on the topic is scarce, recent, and heterogeneous. Previous literature [39] explored the use of eHealth technologies as a plausible approach to supporting aging with cognitive impairment and identified cognitive training solutions and supportive web platforms as the most effective interventions on a limited number of outcomes. More research is needed to solve methodological difficulties observed in the current literature, according to Dequanter et al [39], and development should focus on solutions for leisure and reminiscence as well as outcomes directly relevant to independent living. Di Lorito et al [40] gathered evidence on the effectiveness of digital health interventions on physical, cognitive, behavioral, and psychological outcomes as well as activities of daily living in people with dementia and mild cognitive impairment. The authors recommend considering different modalities of supervision while administering digital health interventions. "A mix of remote and face-to-face delivery could maximize benefits and optimize costs," according to Di Lorito et al [40]. This paper is the first literature review that gathers the available digital health technology to support health care professionals and family caregivers caring for patients with cognitive impairment and synthesizes the implementation outcomes of using these technologies in the community or in a facility. We were able to retrieve only 7 manuscripts, including

6 (86%) peer-reviewed journal articles and 1 (14%) doctoral thesis, reporting on studies conducted over the last decade involving various technologies. Of these 7 studies, 4 (57%) reported on systems embedded into computers and tablet devices as gaming systems, self-management systems, and apps, whereas 3 (43%) reported on digital calendar reminder systems to support activities of daily living and medication management and a social robot used in therapeutic sessions. The interactions among patients with cognitive impairment, health care providers, and family caregivers identified in the included manuscripts were diverse and heterogeneous. Overall, the literature reported on the use of digital health technology to support bilateral interactions between patients and health care providers or between patients and family caregivers or a trilateral integration with support provided to all 3 groups, whereas 14% (1/7) of the studies reported on technology to support family caregivers only (people with dementia indirectly). The trilateral integration of technology complements the portrait presented by Huisman et al [41], who reported on technology targeting a person with dementia, the informal caregiver, or both.

Our scoping review indicates that the retrieved publications lack clarity regarding adopting a behavior change theory and associated model to carry out digital health technology intervention. Most (6/7, 86%) of the studies examined in this review did not include any explicit behavior change theory or model. The study by Cruz-Sandoval et al [34] was the only one reporting on using a behavior change theory, namely a comprehensive process model of engagement followed by cognitive stimulation therapy [34]. Two authors involved other theories such as activity theory [37] and simulated presence therapy intervention [35], but these are not behavior change models. The rest (4/7, 57%) of the articles did not mention any behavior change theory or model. This phenomenon can be explained by a lack of application of theoretical frameworks to conceptualize and govern the behavior change models that incorporate technology for patients with cognitive impairment. The shortcomings of using general theories and models for designing interventions aimed at this profile of patients with cognitive impairment were discussed in a recent scoping review of behavior change theories in adults without dementia [42]. The study aimed to adapt and develop a model promoting physical activity in people with dementia (the “PHYT in dementia”) [42] and presented an ad hoc model for the promotion of physical activity. Our scoping review indicates the direction that future studies may follow to eventually observe outcomes with a larger impact, that is, crafting a model that can be applied and replicated to inform the design of systems that support people with dementia to change their behavior.

Digital health technology can take multiple forms depending on the intervention it should deliver. It is important to note that only 14% (1/7) of the studies reported on implementing a robot as a conversational agent [34], and no other study presented another type of conversational agent, which can provide exciting opportunities for developing new ways of communicating with people with cognitive impairment. As other researchers have stated [43], implementing conversational agents to deliver interventions to people with dementia is still a poorly explored domain that deserves more attention and interdisciplinary work. This knowledge gap creates a new quandary in using established theoretical frameworks (eg, behavior change theories in our case) when building and deploying digital health technology. Our findings show that the use of digital health technology may bring real-world evidence [34] that is not supported by existing theoretical frameworks. This gap in real-world evidence documentation is a whole understudied research question. One can wonder whether technology must be used to support the adoption of behavior change approaches or whether it is introducing a new strategy that may prove effective but has not yet been defined and theorized. This dilemma may also raise the fundamental question of the necessity for behavior change theories to assist health care professionals and family caregivers caring for patients with cognitive impairment. Digital health technology may be perceived as a novel technique that has shown effectiveness (real-world proof) but has yet to be conceptualized as a posteriori. In this context, Bell et al [28] emphasize the importance of including perspectives of young people and lived experience, motivational frameworks, and behavior change interventions in developing digital health solutions. Our scoping review revealed 2 critical factors to consider when comprehending digital health technology for patients with cognitive impairment. The first factor is the operation mode. Digital health technology can be designed to operate with, or independently of, other hardware or software. A technology that operates independently of other technologies is called a stand-alone technology, which is usually easier to learn and use than a non–stand-alone technology. The second factor is the number of users. Although some of the digital health technologies presented in the retrieved studies are considered stand-alone technologies, the interventions needed more support from others. Understandably, end users may need the support of other people around them, especially when the end users are patients with cognitive impairment. Often, digital health technology is designed to enable family caregivers to access it, too, even if they are not required to act to deliver the intervention.

The design of technology and the inherent intervention seems critical in implementing and using digital health technology with end users. Di Lorito et al [40] recommend considering accessibility, acceptability, and sustainability for end users as prerequisites for the development of digital health interventions. Beyond exploring the effects of digital health technology on patients with cognitive impairment and their interaction with health care providers and family caregivers, this scoping review aimed to identify what determines the success of a digital health intervention. Design wise, only 29% (2/7) of the studies included in this review followed a user-centered design approach, despite the importance of such methodology when addressing a specific population with particular needs, namely patients with cognitive impairment [21-23]. A secondary observation from our scoping review is that designing a digital health technology for a patient with cognitive impairment necessitates collaboration with secondary stakeholders such as family caregivers and tertiary stakeholders from specialized institutions and clinics, in addition to actively and repeatedly involving the end user. Hung et al [35] followed a thorough user-centered approach involving patients and caregivers using video-ethnographic methods, video-recorded observations, and health care professionals’

interviews. Andreassen et al [33] validated the app with end users [44] and caregivers [45], but there is no explicit design process described. Including caregivers in the design is uncommon (in this scoping review, only a single study [35] did so) and represents a future research possibility. It is also important to note the complete absence of co-design and cocreation of solutions when this has been demonstrated to be a valuable methodology [46]. Co-design can be challenging for people with severe dementia or cognitive impairments, but it is still viable for people with less complex conditions as well as caregivers [47,48]. Because of their firsthand experience, caregivers can contribute to cocreating solutions for contexts they know well. This type of information is difficult to capture, formalize, and convey in interviews but is more naturally communicated and incorporated in cocreation endeavors.

Strengths and Limitations

The findings of this study contribute to filling a significant knowledge gap regarding the value of using emerging digital health technologies to assist health care professionals and family caregivers in caring for patients with cognitive impairment. Although the portrait is limited to 7 manuscripts, this scoping review presents many compiled results and recommendations based on the primary authors' observations. This knowledge is one of the priorities identified in the literature over the last decade, and its importance has grown since the COVID-19 pandemic significantly affected the delivery of rehabilitation services. This paper has presented the evidence as of April 2021, and there may be ongoing yet unpublished research work applied to the COVID-19 context. Although the findings are helpful for pandemic-related decision-making, the presented findings were interpreted without regard for the pandemic. This review did not focus on addressing cognitive impairment from a therapeutic perspective, and we did not discuss the use of digital health technology as a clinical approach. Although it was built as exhaustive as possible, our search strategy may have lacked some terms; for example, we did not include critical MeSH terms such as *Mobile Applications* and *Telemedicine* because these were expected to yield the same manuscripts as search strings 1 to 6 (Multimedia Appendix 1). In addition, we planned to include only peer-reviewed articles, but we added a doctoral thesis because of the scarcity of articles. Therefore, we discourage hasty generalization of the results owing to the heterogeneity of the technologies and contexts of the interventions and the small number of manuscripts included.

Conclusions

This paper showed that despite a plethora of theoretical frameworks available on the use of behavior change models, there is an apparent lack of application of theories incorporating technology for patients with cognitive impairment. This scoping review depicted existing digital health technology to support health care providers and family caregivers while caring for patients with cognitive impairment and highlighted the lack of reliance on behavior change theories and models despite studies showing their positive effects on patients with cognitive impairment, health care providers, and family caregivers with no adverse effects on any digital health technology users. This scoping review depicted the benefits, challenges, and influencers of using digital health technology to support patients with cognitive impairment, health care providers, and family caregivers to bolster future research and implementation.

Acknowledgments

This research was supported by a joint grant from Victoria Hospital Foundation and Research Manitoba.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 13 KB - mental_v10i1e40330_app1.docx]

References

1. Crimmins EM, Kim JK, Langa KM, Weir DR. Assessment of cognition using surveys and neuropsychological assessment: the Health and Retirement Study and the Aging, Demographics, and Memory Study. J Gerontol B Psychol Sci Soc Sci 2011 Jul;66 Suppl 1(Suppl 1):i162-i171 [FREE Full text] [doi: 10.1093/geronb/gbr048] [Medline: 21743047]
2. Canadian Institutes for Health Information. Dementia in Canada: Summary. Canadian Institutes for Health Information. 2022. URL: https://www.cihi.ca/en/dementia-in-canada/dementia-in-canada-summary [accessed 2022-11-24]
3. Roberts R, Knopman DS. Classification and epidemiology of MCI. Clin Geriatr Med 2013 Nov;29(4):753-772 [FREE Full text] [doi: 10.1016/j.cger.2013.07.003] [Medline: 24094295]
4. Park M, Kwon DY, Jung JM, Han C, Jo I, Jo SA. Mini-Mental Status Examination as predictors of mortality in the elderly. Acta Psychiatr Scand 2013 Apr;127(4):298-304. [doi: 10.1111/j.1600-0447.2012.01918.x] [Medline: 22901036]
5. Millán-Calenti JC, Tubío J, Pita-Fernández S, González-Abraldes I, Lorenzo T, Maseda A. Prevalence of cognitive impairment: effects of level of education, age, sex and associated factors. Dement Geriatr Cogn Disord 2009 Nov 10;28(5):455-460. [doi: 10.1159/000257086] [Medline: 19907183]

6. Geda YE. Mild cognitive impairment in older adults. Curr Psychiatry Rep 2012 Aug 7;14(4):320-327 [FREE Full text] [doi: 10.1007/s11920-012-0291-x] [Medline: 22773365]
7. Paradise M, McCade D, Hickie IB, Diamond K, Lewis SJ, Naismith SL. Caregiver burden in mild cognitive impairment. Aging Ment Health 2015 Jan 28;19(1):72-78. [doi: 10.1080/13607863.2014.915922] [Medline: 24866046]
8. Dewing J, Dijk S. What is the current state of care for older people with dementia in general hospitals? A literature review. Dementia (London) 2016 Jan 23;15(1):106-124. [doi: 10.1177/1471301213520172] [Medline: 24459188]
9. Kracht F, Boekholt M, Schumacher-Schönert F, Nikelski A, Chikhradze N, Lücker P, et al. Describing people with cognitive impairment and their complex treatment needs during routine care in the hospital - cross-sectional results of the intersec-CM study. BMC Geriatr 2021 Jul 12;21(1):425 [FREE Full text] [doi: 10.1186/s12877-021-02298-4] [Medline: 34253180]
10. What is digital health? U.S. Food & Drug Administration. 2020 Sep 22. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2022-11-28]
11. Bird M, Li L, Ouellette C, Hopkins K, McGillion MH, Carter N. Use of synchronous digital health technologies for the care of children with special health care needs and their families: scoping review. JMIR Pediatr Parent 2019 Nov 21;2(2):e15106 [FREE Full text] [doi: 10.2196/15106] [Medline: 31750840]
12. Raj M, Gupta V, Hoodin F, Yahng L, Braun T, Choi SW. Evaluating health technology engagement among family caregivers of patients undergoing hematopoietic cell transplantation. Res Sq 2021 May 14:7058 [FREE Full text] [doi: 10.21203/rs.3.rs-427058/v1] [Medline: 34013246]
13. Luis-Martínez R, Monje MH, Antonini A, Sánchez-Ferro A, Mestre TA. Technology-enabled care: integrating multidisciplinary care in Parkinson's disease through digital technology. Front Neurol 2020 Oct 30;11:575975 [FREE Full text] [doi: 10.3389/fneur.2020.575975] [Medline: 33250846]
14. Whitelaw S, Pellegrini DM, Mamas MA, Cowie M, Van Spall HG. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021 Mar;2(1):62-74 [FREE Full text] [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
15. Vaish H. Is digital technology the solution for pulmonary rehabilitation in developing countries during the COVID-19 pandemic? Turk Thorac J 2021 Apr 07;22(2):186-187. [doi: 10.5152/turkthoracj.2021.20224]
16. Fekete M, Fazekas-Pongor V, Balazs P, Tarantini S, Nemeth AN, Varga JT. Role of new digital technologies and telemedicine in pulmonary rehabilitation : smart devices in the treatment of chronic respiratory diseases. Wien Klin Wochenschr 2021 Nov 30;133(21-22):1201-1207 [FREE Full text] [doi: 10.1007/s00508-021-01930-y] [Medline: 34460006]
17. Lopes RP, Barroso B, Deusdado L, Novo A, Guimarães M, Teixeira JP, et al. Digital technologies for innovative mental health rehabilitation. Electronics 2021 Sep 14;10(18):2260. [doi: 10.3390/electronics10182260]
18. van der Ham IJ, van der Vaart R, Miedema A, Visser-Meily JM, van der Kuil MN. Healthcare professionals' acceptance of digital cognitive rehabilitation. Front Psychol 2020;11:617886 [FREE Full text] [doi: 10.3389/fpsyg.2020.617886] [Medline: 33324309]
19. Henoch I, Strang S, Larnebratt C, Hermansson J. Digital health support – a good thing, but not for me. Experiences of family caregivers and healthcare staff. British J Guidance Counselling 2020 Oct 02:1-12. [doi: 10.1080/03069885.2020.1826030]
20. Manteghinejad A, Javanmard SH. Challenges and opportunities of digital health in a post-COVID19 world. J Res Med Sci 2021;26:11 [FREE Full text] [doi: 10.4103/jrms.JRMS_1255_20] [Medline: 34084190]
21. Newell A, Arnott J, Carmichael A, Morgan M. Methodologies for involving older adults in the design process. In: Universal Acess in Human Computer Interaction. Coping with Diversity. Berlin, Heidelberg: Springer; 2007.
22. Oskar W. User Centered Design for Persons with Disabilities – How persons with cerebral palsy can be included in the design process. Digitala Vetenskapliga Arkivet. 2011 Nov 26. URL: http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A654239&dswid=-5246 [accessed 2022-11-05]
23. Eisma R, Dickinson A, Goodman J, Mival O, Syme A, Tiwari L. Mutual inspiration in the development of new technology for older people. INCLUDE. 2003. URL: https://www.researchgate.net/publication/229029733_Mutual_inspiration_in_the_development_of_new_technology_for_older_people [accessed 2022-11-29]
24. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Social Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
25. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5(1):69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
26. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Tunçalp, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/m18-0850]
27. Ball E, Owen-Booth B, Gray A, Shenkin SD, Hewitt J, McCleery J. Aromatherapy for dementia. Cochrane Database Syst Rev 2020 Aug 19;8(8):CD003150 [FREE Full text] [doi: 10.1002/14651858.CD003150.pub3] [Medline: 32813272]
28. Bell I, Pot-Kolder RM, Wood SJ, Nelson B, Acevedo N, Stainton A, et al. Digital technology for addressing cognitive impairment in recent-onset psychosis: a perspective. Schizophr Res Cogn 2022 Jun;28:100247 [FREE Full text] [doi: 10.1016/j.scog.2022.100247] [Medline: 35281550]

29. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016 Jul;104(3):240-243 [FREE Full text] [doi: 10.3163/1536-5050.104.3.014] [Medline: 27366130]
30. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
31. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug 20;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
32. Olexsovich A. Interpersonal APProach to dementia: an iPad-based program for caregiver education and decreasing problem behaviors in older adults with cognitive impairments. Duquesne University. 2016. URL: https://dsc.duq.edu/cgi/viewcontent.cgi?article=1049&context=etd [accessed 2022-11-28]
33. Andreassen M, Hemmingsson H, Boman I, Danielsson H, Jaarsma T. Feasibility of an intervention for patients with cognitive impairment using an interactive digital calendar with mobile phone reminders (RemindMe) to improve the performance of activities in everyday life. Int J Environ Res Public Health 2020 Mar 26;17(7):2222 [FREE Full text] [doi: 10.3390/ijerph17072222] [Medline: 32224975]
34. Cruz-Sandoval D, Morales-Tellez A, Sandoval EB, Favela J. A social robot as therapy facilitator in interventions to deal with dementia-related behavioral symptoms. In: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. 2020 Presented at: HRI '20: ACM/IEEE International Conference on Human-Robot Interaction; Mar 23 - 26, 2020; Cambridge United Kingdom. [doi: 10.1145/3319502.3374840]
35. Hung L, Au-Yeung A, Helmer C, Ip A, Elijah L, Wilkins-Ho M, et al. Feasibility and acceptability of an iPad intervention to support dementia care in the hospital setting. Contemp Nurse 2018 Aug 10;54(4-5):350-361. [doi: 10.1080/10376178.2018.1505436] [Medline: 30068237]
36. Kamimura T, Ishiwata R, Inoue T. Medication reminder device for the elderly patients with mild cognitive impairment. Am J Alzheimers Dis Other Demen 2012 Jun 27;27(4):238-242 [FREE Full text] [doi: 10.1177/1533317512450066] [Medline: 22739031]
37. Lazar A, Demiris G, Thompson HJ. Evaluation of a multifunctional technology system in a memory care unit: opportunities for innovation in dementia care. Inform Health Soc Care 2016 Dec 28;41(4):373-386 [FREE Full text] [doi: 10.3109/17538157.2015.1064428] [Medline: 26819070]
38. Huis In Het Veld JG, Willemse BM, van Asch IF, Groot Zwaaftink RB, Verkade P, Twisk JW, et al. Online self-management support for family caregivers dealing with behavior changes in relatives with dementia (part 2): randomized controlled trial. J Med Internet Res 2020 Mar 25;22(2):e13001 [FREE Full text] [doi: 10.2196/13001] [Medline: 32130142]
39. Dequanter S, Gagnon MP, Ndiaye MA, Gorus E, Fobelets M, Giguère A, et al. The effectiveness of e-health solutions for aging with cognitive impairment: a systematic review. Gerontologist 2021 Sep 13;61(7):e373-e394 [FREE Full text] [doi: 10.1093/geront/gnaa065] [Medline: 32525977]
40. Di Lorito C, Bosco A, Rai H, Craven M, McNally D, Todd C, et al. A systematic literature review and meta-analysis on digital health interventions for people living with dementia and Mild Cognitive Impairment. Int J Geriatr Psychiatry 2022 May 04;37(6):10.1002/gps.5730 [FREE Full text] [doi: 10.1002/gps.5730] [Medline: 35588315]
41. Huisman C, Huisman E, Kort H. Technological applications contributing to relieve care burden or to sleep of caregivers and people with dementia: a scoping review from the perspective of social isolation. Front Public Health 2022 Mar 29;10:797176 [FREE Full text] [doi: 10.3389/fpubh.2022.797176] [Medline: 35425752]
42. Di Lorito C, Pollock K, Harwood R, das Nair R, Logan P, Goldberg S, et al. A scoping review of behaviour change theories in adults without dementia to adapt and develop the 'PHYT in dementia', a model promoting physical activity in people with dementia. Maturitas 2019 Mar;121:101-113 [FREE Full text] [doi: 10.1016/j.maturitas.2019.01.008] [Medline: 30686623]
43. Rampioni M, Stara V, Felici E, Rossi L, Paolini S. Embodied conversational agents for patients with dementia: thematic literature analysis. JMIR Mhealth Uhealth 2021 Jul 16;9(7):e25381 [FREE Full text] [doi: 10.2196/25381] [Medline: 34269686]
44. Baric V, Andreassen M, Öhman A, Hemmingsson H. Using an interactive digital calendar with mobile phone reminders by senior people - a focus group study. BMC Geriatr 2019 Apr 23;19(1):116 [FREE Full text] [doi: 10.1186/s12877-019-1128-9] [Medline: 31014276]
45. Bolic Baric V, Tegelström V, Ekblad E, Hemmingsson H. Usability of RemindMe - an interactive web-based mobile reminder calendar: a professional's perspective. Stud Health Technol Inform 2015;217:247-254. [Medline: 26294480]
46. Rodgers PA. Co-designing with people living with dementia. CoDesign 2017 Feb 15;14(3):188-202. [doi: 10.1080/15710882.2017.1282527]
47. Vollmer Dahlke D, Ory MG. Emerging issues of intelligent assistive technology use among people with dementia and their caregivers: a U.S. Perspective. Front Public Health 2020 May 21;8:191 [FREE Full text] [doi: 10.3389/fpubh.2020.00191] [Medline: 32528920]

48. Wang G, Marradi C, Albayrak A, van der Cammen TJ. Co-designing with people with dementia: a scoping review of involving people with dementia in design research. Maturitas 2019 Sep;127:55-63 [FREE Full text] [doi: 10.1016/j.maturitas.2019.06.003] [Medline: 31351521]

Abbreviations

MeSH: Medical Subject Headings
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by J Torous; submitted 15.06.22; peer-reviewed by M Bardus, I Wilson, T Holthe; comments to author 03.07.22; revised version received 25.09.22; accepted 09.11.22; published 11.01.23.

Please cite as:
Choukou MA, Olatoye F, Urbanowski R, Caon M, Monnin C
Digital Health Technology to Support Health Care Professionals and Family Caregivers Caring for Patients With Cognitive Impairment: Scoping Review
JMIR Ment Health 2023;10:e40330
URL: https://mental.jmir.org/2023/1/e40330
doi:10.2196/40330
PMID:36630174

Original Paper

Young People's Trust in Cocreated Web-Based Resources to Promote Mental Health Literacy: Focus Group Study

Sachiyo Ito-Jaeger[1,2], PhD; Elvira Perez Vallejos[1,2], PhD; Saruka Logathasan[1], BMedSci; Thomas Curran[3], PhD; Paul Crawford[1], PhD

[1]Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, United Kingdom

[2]National Institute for Health and Care Research (NIHR), Nottingham Biomedical Research Centre, Nottingham, United Kingdom

[3]Department of Psychological and Behavioural Science, London School of Economics, London, United Kingdom

Corresponding Author:
Sachiyo Ito-Jaeger, PhD
Faculty of Medicine and Health Sciences
University of Nottingham
Institute of Mental Health
Innovation Park, Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 1158231294
Email: sachiyo.ito-jaeger@nottingham.ac.uk

Abstract

Background: There is a pressing need to create resources to promote mental health literacy among young people. Digital media is one of the methods that can be used to successfully promote mental health literacy. Although digital mental health resources are generally favorably perceived by young people, one of the essential factors in whether they choose to use these interventions is trust.

Objective: The objective of this study was to explore young people's trust-related concerns about and recommendations for the cocreated mental health website "What's Up With Everyone" by using TrustScapes. Our aim was to use the findings to improve the trustworthiness of the website and to inform future creators of web-based mental health resources.

Methods: In total, 30 young people (mean age 19, SD 1.509; range 17-21 years) participated in TrustScapes focus groups. Thematic analysis was carried out to analyze both the TrustScapes worksheets and audio transcripts.

Results: Qualitative analysis revealed that the mental health website contains elements perceived to be both trustworthy and untrustworthy by young people. The relatable and high-quality design, which was achieved by cocreating the website with a team of design professionals and young people, was considered to increase trust. Creators' credibility also positively affected trust, but the logos and other information about the creators were recommended to be more salient for users. Suggestions were made to update the privacy policy and cookie settings and include communication functions on the platform to improve the trustworthiness of the website.

Conclusions: Factors perceived to be trustworthy included the website's relatable, high-quality design and creators' credibility, whereas those perceived to be untrustworthy included the privacy policy and cookie settings. The findings highlighted the significance of collaborating with end users and industrial partners and the importance of making the trust-enabling factors salient for users. We hope that these findings will inform future creators of web-based mental health resources to make these resources as trustworthy and effective as possible.

(JMIR Ment Health 2023;10:e38346) doi:10.2196/38346

KEYWORDS

trust; mental health; web; young people; cocreation; mental health literacy; qualitative study; thematic analysis; trustworthy; digital mental health; internet; digital health; mobile phone

Introduction

Background

Overview

Mental health problems are increasing globally [1]. Mental ill-health is of particular concern for young people, as 20% of them face mental health issues in any given year [1], and 75% of mental health issues develop by age 24 [2]. The World Health Organization urges increased investment in mental health in all facets, including in mental health awareness, to promote mental health literacy [3]. Mental health literacy is defined as "understanding how to obtain and maintain positive mental health; understanding mental disorders and their treatments; decreasing stigma related to mental disorders; and, enhancing help-seeking efficacy (knowing when and where to seek help and developing competencies designed to improve one's mental health care and self-management capabilities)" [4]. There is a pressing need to create resources to promote mental health literacy among young people.

Digital Mental Health

Digital media is one of the methods that can be used to successfully promote mental health literacy [5-7]. Advancement in digital technology allows mental health information to reach a wide and large audience. Web-based mental health resources have great advantages, such as accessibility, cost-effectiveness, instantaneity, and anonymity [8,9]. They can attract especially young people because such resources are favorably perceived by them and are effective for them [10,11]. According to a recent systematic review of studies from various countries, many young people have had experience in using web-based mental health resources and are prepared to use them [12]. For example, in the United Kingdom, approximately 70% of university students reported that they were confident in using a computer or telephone to seek information about mental health issues [13]. In the United States, 87% of young people aged 14 to 22 years have accessed health information via the web, and 90% of those who had experienced depressive symptoms reported accessing mental health information via the web [14].

Trust in Digital Health

Although digital mental health resources are generally favorably perceived by young people, one of the essential factors in whether they choose to use these interventions is trust. Trust is defined as "a positive belief about the perceived reliability of, dependability of, and confidence in a person, object, or process" and is closely related to credibility [15]. It has been described as a relationship between a trustor and a trustee "with optimistic anticipation that the trustee will fulfil the trustor's expectations" [16]. Trust is imperative in health care, as it is a crucial factor to promote health care access and improve treatment outcomes [17]. Trust is also of great importance when one decides whether to use a particular kind of technology [18] as it is "a precursor to successful and effective adoption, interaction and ongoing commitment in the digital space" [19]. Young people encounter an excessive number of web-based mental health resources, and thus, they face significant challenges in evaluating and selecting which resources to use [20]. As these resources vary largely in their accuracy [20], it is critical that the most informationally accurate web-based resources are perceived as trustworthy by the target users and thus selected for use.

Some factors are found to encourage or discourage users' trust in digital health. Factors that positively affect trust in digital health resources include stakeholder engagement, creators' credibility and reputation, validity of the information, fair data access, clear design, and ease of use [9,16,21]. More specifically, engaging stakeholders and taking their suggestions into account appears to be an important condition to increase trust in digital health [16]. In addition, stakeholders are more likely to trust digital health interventions created by public institutions (eg, universities and health services) than those created by private companies (eg, pharmaceutical companies) [16].

In contrast, factors that negatively influence trust in digital health include fear of data exploitation, concerns about privacy, poor information quality, inadequate publicity, excessive cost, and defective technology [9,16,21]. For example, end users find it difficult to trust some digital health systems owing to the fear of data exploitation from third parties [16]. Privacy is one of the major concerns for young people when they access digital mental health interventions [9,22,23].

Objective

The objective of this study was to explore young people's trust-related concerns about the cocreated mental health website, What's Up With Everyone (WUWE; refer to the *Methods* section for details). In addition, we asked young people to generate both realistic and idealistic recommendations to improve the trustworthiness of the website. We aimed to gain deep insight into young people's perceptions about trust in the web-based mental health resource by using TrustScapes (refer to the *Methods* section for details). Our aim was not only to use the findings to improve the trustworthiness of the website but also to inform the future development of web-based mental health resources.

Methods

Ethics Approval

This study was approved by the research ethics committee of the University of Nottingham (application ID CS-2019-R30).

Participants

After obtaining approval, we recruited an independent sample of young people (ie, different from those who participated in the cocreation process) solely for the purpose of evaluating the website, by distributing flyers and posting the information on a web-based recruitment site. The flyers, containing the description of the study and the eligibility criteria for participation, were distributed via email to mental health organizations, high schools, colleges, and universities in the United Kingdom. As the target audience of the website is a diverse group of young people, we tried to recruit young people from diverse backgrounds to be inclusive and hear diverse opinions. The eligibility criteria for participation included the following: (1) English speaker living in the United Kingdom, (2) aged between 17 and 21 years at the time of recruitment,

and (3) has access to the internet and a computer or tablet. Potential participants were asked to submit the Expression of Interest form.

The WUWE Website

In this study, we explored what elements of the newly created mental health website, WUWE, young people would find trustworthy or untrustworthy, by using TrustScapes (refer to the following sections for details). The WUWE project is a campaign developed to promote mental health literacy among young people. It comprised a series of 5 short, animated films and a companion mental health website created with and for young people, in partnership with multiple award–winning independent animation studio, Aardman Animations.

The website contains information on 5 mental health–related issues commonly experienced by young people (ie, perfectionism, loneliness and isolation, independence, social media, and competitiveness), including the short films corresponding to these themes. The website also contains sections about where to seek help and about the creators, privacy policy, and links to social media. Full details of the project and context can be accessed on the website [24] and the UK Research and Innovation announcement [25]. In a 4-month media campaign following the launch of WUWE on February 8, 2021, the companion website alone attracted 33,100 users, with 44,000 sessions; 101,000 unique page views; 994,000 total page views; and 4500 returning users.

Stakeholder Engagement

The stakeholders (ie, young people) were actively involved throughout the project, contributing to the development, production, implementation, and evaluation, as recommended by Jirotka et al [26]. Details of how we cocreated the animated films with young people, including how the 5 themes were generated, are described in another paper [27]. Regarding the companion website, 5 mental health professionals with background in psychology or mental health nursing first wrote the text to describe each of the 5 issues. Young people who had been involved in the animation production process then provided feedback on the text (eg, length and complexity) during workshops. The text was revised multiple times until it was fully approved. Young people also provided feedback on the design and layout of the website to the Aardman production team. In the implementation stage, the young cocreators provided suggestions on public relations strategy, such as which influencers are suited for the media campaign. This paper reports the methods and results of the final evaluation stage, focusing on the findings from the TrustScapes focus groups [FGs].

TrustScapes

The TrustScapes Toolkit was used during FGs to capture young people's trust-related concerns about and recommendations for the WUWE website. TrustScapes are part of the Fairness Toolkit, which was developed for the UnBias project, funded by the Engineering and Physical Sciences Research Council [28,29]. The Fairness Toolkit is aimed to raise awareness and facilitate a public civic dialogue about how algorithms shape web experiences and to reflect on possible changes to address issues of web fairness [30].

The TrustScapes Toolkit includes a worksheet (Multimedia Appendix 1), keywords (Multimedia Appendix 2), and sketches (Multimedia Appendix 3). The worksheet was cocreated by designers, stakeholders, and researchers through a series of workshops for the stakeholders, including the end users, to visualize their perceptions about data protection, web safety, and algorithmic bias and what they would like to see changed to make the web-based world fair and trustworthy. It is designed to capture both their feelings about the current situation and their dreams and ideals about what the digital world could or should be, in a dynamic and visual manner.

The following four questions are included in the worksheet:

1. Describe an experience of untrustworthiness you are concerned about.
2. Illustrate what is important to you about this experience.
3. How do you think these issues should be addressed by us?
4. Ideally, what would you like to see done?

Keywords and sketches, which had also been cocreated by designers, stakeholders, and researchers, were provided to inspire participants to complete the worksheet and prompt them to contribute their own drawings and insight. TrustScapes are highly interactive and used during FGs to discuss a specific topic (eg, algorithmic systems in mental health).

Procedure

Young people who confirmed their willingness to participate in the study received and submitted the consent form and a demographic questionnaire via email. A week before the TrustScapes FGs, participants received an email containing the TrustScapes worksheet, keywords, sketches, and the link to the WUWE website and were asked to familiarize themselves with the 3 documents and browse the website before the session. Participants were also asked to have a large piece of paper and a pen for completing the TrustScapes questions.

All TrustScapes FGs were conducted through the web via Zoom, as the United Kingdom was under COVID-19 restrictions at the time of the study. The TrustScapes FGs were moderated by 2 authors who specialize in psychology and digital technologies, and one of them has had substantial experience in facilitating TrustScapes FGs. The TrustScapes FG started with an ice breaker exercise, followed by a description of the TrustScapes methodology and the aims of the study.

A moderator showed the WUWE website to the participants by sharing their computer screen. Participants were asked to discuss trust-related concerns about and recommendations for the website on each page (ie, cookie pop-up, home page, about us, privacy policy, theme pages such as perfectionism, seeking help, and social media such as Instagram). Following the group discussions, participants were asked to complete the TrustScapes worksheet independently. Once all participants completed the TrustScapes worksheet, each participant presented their ideas to the whole group. Participants emailed their TrustScapes worksheet to the moderator after the session. Each session lasted for 1.5 hours and was audio-recorded and transcribed verbatim. Participants were compensated financially for their participation. A total of 6 TrustScapes FGs were conducted, and each session consisted of 3 to 7 participants. Data saturation became apparent,

with major trends being clear by the end of the sixth FG; thus, further sessions were not conducted.

Data Analysis

Thematic analysis was used to analyze the data collected from the TrustScapes worksheets and transcripts [31]. Thematic analysis has been used to understand young people's opinions about a newly developed digital intervention [27,32]. Following the six phases of thematic analysis by Braun and Clarke [31,33], an author (1) became familiar with the content by reading and rereading the transcripts, (2) generated initial codes, (3) searched for themes, (4) reviewed the themes, (5) defined and named the themes, and later, (6) produced the report. Throughout the process, research meetings were conducted for debriefing and discussion among the authors.

Results

Participant Characteristics

In total, 30 young people (mean age 19, SD 1.509; range 17-21 years) participated in the study. More than 70% of the participants were women. Participants were from diverse ethnic backgrounds, with White British being the majority (7/30, 23%). Approximately half of the participants (14/30, 47%) had level-3 qualification as their highest level of education, and 30% (9/30) had level-2 qualification. Complete demographic information is presented in Table 1.

Table 1. Demographics of participants (N=30).

Characteristics	Participants, n (%)
Sex	
Female	22 (73)
Male	7 (23)
Prefer not to say	1 (3)
Ethnic background	
White	
English, Welsh, Scottish, Northern Irish, or British	7 (23)
Irish	1 (3)
Other	3 (10)
Multiple ethnic groups	
White and Black Caribbean	1 (3)
White and Black African	1 (3)
Other	1 (3)
Asian	
Indian	5 (17)
Pakistani	1 (3)
Bangladeshi	2 (7)
Other	2 (7)
Black, African, Caribbean, or Black British	
African	2 (7)
Other	1 (3)
Prefer not to say	3 (10)
Religion	
Christian	6 (20)
Hindu	2 (7)
Muslim	4 (13)
Sikh	1 (3)
Prefer not to say	6 (20)
Other	1 (3)
None	10 (33)
Highest level of qualification[a] (participant)	
None	0 (0)
Level 1[b]	2 (7)
Level 2[c]	9 (30)
Level 3[d]	14 (47)
Level 4 or above[e]	2 (7)
Other[f]	1 (3)
Prefer not to say	2 (7)
Highest level of qualification[a] (parent, guardian, or carer)	
None	2 (7)

Characteristics	Participants, n (%)
Level 1[b]	1 (3)
Level 2[c]	4 (13)
Level 3[d]	3 (10)
Level 4 or above[e]	8 (27)
Other[f]	4 (13)
Prefer not to say	8 (27)

[a]Qualification levels (UK census).

[b]Includes O levels 1-4, Certificate of Secondary Education (CSEs), or General CSE (any grade); entry level; Foundation Diploma; National Vocational Qualification (NVQ) level 1; Foundation General NVQ (GNVQ); and basic skills.

[c]Includes O level ≥5 (pass), CSEs (grade 1), or General CSEs (grades A*-C); school certificate; A level 1, Advanced Supplementary level 2-3, or Victoria Certificate of Education; higher diploma; NVQ level 2; Intermediate GNVQ; City and Guilds Craft; Business and Technology Education Council (BTEC) First or General Diploma; and Royal Society of Art (RSA) Diploma.

[d]Includes A level ≥2 or Victoria Certificate of Education, Advanced Supplementary level ≥4, higher school certificate, Progression or Advanced Diploma, NVQ level 3, Advanced GNVQ, City and Guilds Advanced Craft, Ordinary National Certificate, Ordinary National Diploma, BTEC National, and RSA Advanced Diploma.

[e]Includes degree (eg, BA and BSc), higher degree (eg, MA, PhD, and Postgraduate Certificate in Education), NVQ level 4-5, Higher National Certificate, Higher National Diploma, RSA Higher Diploma, BTEC higher level, and professional qualifications (eg, teaching, nursing, and accountancy).

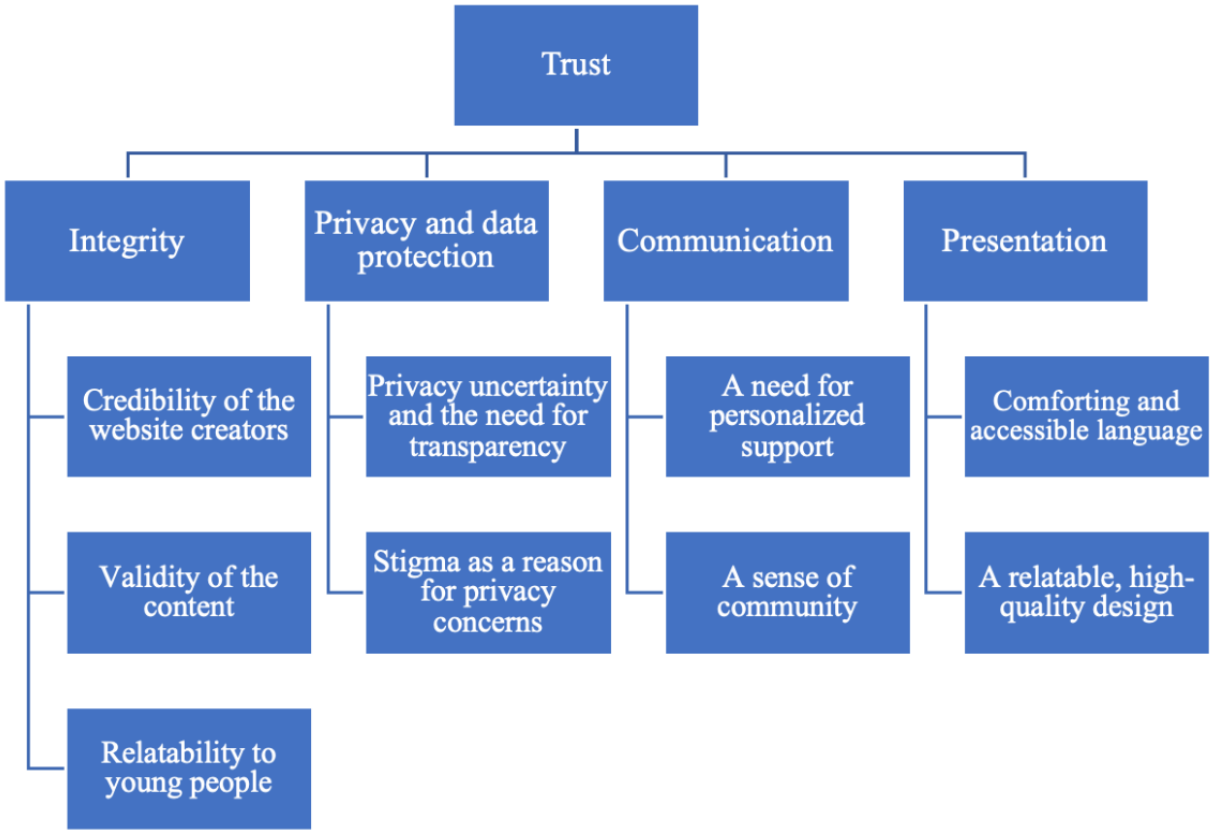
[f]Includes other vocational or work-related qualifications and foreign qualifications.

Themes and Subthemes

Overview

After applying thematic analysis across the 2 data sets (TrustScapes worksheets and FG transcripts), 4 main themes and 2 to 3 subthemes for each theme were generated (Figure 1). The 4 main themes were integrity, privacy and data protection, communication, and presentation. Each subtheme is described below, with supporting quotes from participants' TrustScapes worksheets and FGs. The quotes from the 2 data sets were complementary, as the same themes and subthemes were generated from the 2 separate analyses.

Figure 1. Themes related to young people's trust in the cocreated mental health website.



Integrity

The theme of integrity highlights that the credibility of the website creators, validity of the website content, and relatability to young people are important aspects to enhance trust for the end users.

Credibility of the Website Creators

A major factor that created a sense of trust among participants was the involvement of established institutions and organizations in the development of the WUWE website. Many participants felt that the inclusion of logos helped reassure them that the website had good intentions and the information on it could be trusted:

> *Trusting the website is the most important part as someone experiencing issues may not believe that the advice this website is offering is true or valid and having that trust, which can be gained from the organisations involved.* [Participant 8; FG2; worksheet]

However, some suggested the link to the *about us* page, which contains the logos, needs to be more salient (eg, by moving the button to the top of the landing page and making the button large) for easy access of information. Another suggestion was made to move all the information on the *about us* page to the landing page so that users could see that the WUWE website was legitimate as soon as they access the website:

> *[About Us] should be incorporated into the landing page so users know what to expect or gain from the website. Otherwise, there might be a lot of anxiety around the purpose of the website.* [Participant 10; FG2; worksheet]

Validity of the Content

Similarly, some participants recommended to include the source of the information on the website so that the end users can trust the content. In particular, it was suggested to indicate whether the information on the website had been proven to be significant or effective:

> *As a student I'm more trusting of ideas/findings that have been established as effective/significant in research studies. On the pages for each theme, if an idea/suggestion for how to deal with an issue is presented it'd be nice for a word in the sentence to be highlighted and have a hyperlink to the original study that this idea came from.* [Participant 22; FG5; worksheet]

Relatability to Young People

Some participants indicated that the information that young people were involved and considered in the creation of the website helps them to trust the content of the website. As this information was written only briefly on the *about us* page, participants recommended to highlight the information:

> *I think [knowing that WUWE was cocreated with young people] increases trust because it shows that you really care about the website because it is not just like adults only trying to make something for young people, you have actually interacted with young people to produce something that they want so I think that...it brings like reliability almost.* [Participant 28; FG6; transcript]

Privacy and Data Protection

The following 2 subthemes highlight that privacy is one of the major concerns for young people when they access web-based mental health resources.

Privacy Uncertainty and the Need for Transparency

Many participants were concerned about privacy. As they would not read the privacy policy owing to its length and complexity, they felt that they may be consenting to something they do not understand. They were particularly worried about their data being sold to third parties without their knowledge:

> *The privacy policy is quite long, you may consent to something you didn't read.* [Participant 17; FG4; worksheet]

> *Understanding the information easily [is important to me] - not feeling like I'm drowning. Illustration of a person drowning.* [Participant 18; FG4; worksheet]

> *I am worried the website would access features such as my location. The website may decide to sell my data in the future. Original illustration of a person and a computer.* [Participant 16; FG 4; worksheet]

Participants highlighted that these issues of distrust could be solved by increasing the transparency between the website and end users. They wanted to be reassured that data were not being sold to third parties and wanted to know whether their data were being stored:

> *Make clear what information will be taken whether it will be used by any third parties. TrustScapes illustrations of exploiting data and leaking data.* [Participant 9; FG2; worksheet]

> *TrustScapes illustrations of data mining, digital surveillance, and being watched.* [Participant 27; FG6; worksheet]

Similarly, many participants felt that the website needed more clarity about which cookies are being used and how exactly a user's data are managed and used. They recommended that users should have more control over the cookies that are being used. The WUWE website has only 2 cookie options: *allow cookies* and *decline*. Young people proposed that users should be provided with the option to select only the essential cookies or those they wanted to activate, so that the users have more control over their data:

> *Make it clear and transparent if cookies are/are not being used, what type of cookies, option to accept/reject specific cookies.* [Participant 1; FG1; worksheet]

> *Give a selection of cookies to accept rather than Accept All.* [Participant 16; FG4; worksheet]

At the time of the TrustScapes FGs, one of the cookie options, *allow cookies*, was more visually salient (ie, in a black box) than *decline* (ie, in a white box), and thus, the users' attention

was more likely to be drawn to *allow cookies*. Participants urged for the design to be changed to present the 2 options in a similar manner. The design has since been modified after reviewing the participants' recommendation:

> *The Cookie pop-up at the start is quite leading towards accept rather than decline. Make both accept and decline bold boxes.* [Participant 8; FG2; worksheet]

Stigma as a Reason for Privacy Concerns

Young people voiced the importance of privacy and need for anonymity, in part owing to the stigma around mental health issues. Some participants recommended to include a function that would allow users to quickly exit or hide the website:

> *Some people are quite defensive about their mental health issues, and so may not want those around them to know about this.* [Participant 7; FG2; worksheet]

> *Having a way to quickly hide the page, so that the user can cover what they're looking at, but can quickly go back to it, if they wish to –may want to hide the page from family walking past, for example, if they're not wanting to share their feelings with their family at that moment in time.* [Participant 4; FG2; worksheet]

> *It could be a concern for a young person struggling to talk about their mental health that if they accept cookies on say a family computer, WUWE may follow them to other sites and expose their search history that they're not comfortable sharing.* [Participant 27; FG6; worksheet]

Communication

The following 2 subthemes involve participants highlighting the need to be able to communicate with others to not feel alone on the internet.

A Need for Personalized Support

A concern raised about the WUWE website was the lack of personalized help and responses. Some participants felt that it would be easy to trust a website if live chat features were included for instant support:

> *Install and set up an online chat bot system—would provide instant help.* [Participant 3; FG1; worksheet]

This would allow users to have someone to talk to, and they will not feel alone. Participants highlighted that having professional support, such as therapists, would be a helpful inclusion and ultimately help the user to trust the remote resource:

> *An online 24/7 Live Chat facility in order to feel that there are people behind the organisation to make it more trustworthy, and that the person using the website is being considered at every touchpoint.* [Participant 21; FG5; worksheet]

A Sense of Community

Similarly, some participants reported that it would be helpful to have a function on the platform to interact anonymously with other young people about their mental health. Ideas for how this could be achieved included adding a space to leave comments. It was also highlighted that the inclusion of social media makes the WUWE campaign more accessible to young people, thus helping to create a positive environment:

> *Have a place to let others put comments and issues they have and allow everyone to interact with each other.* [Participant 24; FG5; worksheet]

> *I think [social media] just makes it a lot more accessible to people and I think that erm it is definitely something positive that people can see.* [Participant 1; FG1; transcript]

Suggestions were made to include quotes from young people who have benefited from the website. Trust is increased if the quotes include the person's identity, such as their name, photo, or avatar:

> *I think that giving names or posting photos from those helped by this service may be beneficial.* [Participant 25; FG5; worksheet]

Presentation

The following 2 subthemes highlight how the presentation of a website influences whether a young person decides to trust the website.

Use of Comforting and Accessible Language

The language used on the WUWE website played an important role in whether users trusted the web-based mental health resource. Some participants felt that the repeated use of the word *help* was overwhelming and that the website needed more comforting language to create a safe web-based environment:

> *Change "help" to "support" or have a heading like "Where can I go"; Change "About us" to "Come and meet us" or "Who we are."* [Participant 3; FG1; worksheet]

Related to the subtheme, privacy uncertainty and the need for transparency, many participants mentioned that the language used for the privacy policy was very difficult to understand. It was suggested that WUWE should consider providing a version with more accessible language or presenting it using an animated video to help more users understand and feel comfortable with what they are agreeing to:

> *I think it is important the privacy policy is less automated and includes accessible language that everyone is able to understand so they know exactly how their data is being stored. Include an animated video about the privacy policy, create a separate privacy policy with easy-to-understand language.* [Participant 19; FG4; worksheet]

> *Shorten the privacy policy and have 2 versions; short and long.* [Participant 16; FG4; worksheet]

A Relatable, High-Quality Design

Many participants felt that the design of the website was relatable to young people and had high production quality. In particular, the color scheme and use of animations made the

experience more comforting. They felt this was a good way to present information in a light-hearted and approachable manner:

> *I found the colour scheme particularly trustworthy because I feel like this made the experience more relatable to a young audience, also the use of the animations and doodles around the pages made the pages feel comfortable and easy to read.* [Participant 28; FG6; worksheet]

> *I think this is more like almost a light-hearted way of laying everything out.* [Participant 2; FG1; transcript]

Related to the subtheme, relatability to young people, many participants reported that the cocreation of the website with young people helped the design to be relatable to young people:

> *I think the layout kind of...and the colour scheme erm and you know like the fact that there are videos voiced by you know young people and things like that, I think that kind of tells you [young people were involved].* [Participant 3; FG1; transcript]

Many participants felt that the high-quality design of the website was a strong trust enabler, as it shows users that a lot of thought has been put into the creation of this website. This makes the website appear more professional and, therefore, more trustworthy:

> *I think it makes it a bit more trustworthy because it shows that more thought has gone into it...the fact that it is so professionally done, it makes it seem more legitimate.* [Participant 27; FG6; transcript]

Discussion

Principal Findings

This study explored young people's trust-related concerns about the mental health website. Qualitative analysis revealed that the WUWE website contains elements perceived to be both trustworthy (eg, relatable, high-quality design and creators' credibility) and untrustworthy (eg, privacy policy and cookie setting) by young people. Although some of the factors influencing trust in this study were complementary to those in previous studies, TrustScapes allowed us to gain deep insight into young people's perceptions about trust in mental health resources, including how to improve the trustworthiness of the website.

The fact that the website was created through collaborations among the established institutions and organizations, young people, and a professional creative studio contributed to increased trust in the website. Young people perceived that the high production quality and colorful and comforting design of the website increased trust. The relatable, high-quality design was achieved by collaborating with the end users and a team of design professionals. Stakeholder engagement is one of the pillars of responsible research and innovation and is crucial for creating new media interventions relatable to end users [26]. This study highlights the importance of stakeholder engagement and interdisciplinary collaborations with industrial partners.

Consistent with previous studies [9,16,21], the creators' credibility positively affected trust in digital health. In this study, although the website includes the information about the established institutions' involvement in the study, the information was recommended to be more salient. According to the Prominence-Interpretation Theory, it is not possible for users to notice all elements of a website, and as a result, not all elements are evaluated for credibility [34]. Therefore, website creators should ensure that the elements found to increase trust, such as logos of established institutions, are salient for users.

Consistent with previous studies investigating trust in digital health, privacy is of particular concern for young people [16,22] and transparent privacy policies are recommended to increase trust in a cocreated web-based platform [35]. Some suggestions were made to reduce privacy-related concerns and thus increase trust in the mental health website. One suggestion was to make the privacy policy easy to comprehend for young people. Inspired by this project, a proposal was made to use animations to explain complex privacy policies in clear language. Previous studies support the idea that animation is effective in communicating complex information in an understandable manner [36-39]. Thus, future web service providers should consider using animations to explain privacy policies.

Another way to increase trust related to privacy is to provide cookie settings that give more control to users. Users should be given the option to select the cookies that they wish to activate rather than only limited options (eg, *allow cookies* and *decline*). Moreover, cookie settings should not be designed to lead users to one of the options (ie, *allow cookies*) over the other (eg, *decline*). This is consistent with the age-appropriate design code, which requires web service providers to not use *nudge techniques* that encourage children to turn off privacy protections [40]. Although the code only covers web-based services that are likely to be accessed by children, this practice should be extended to those targeted at young people.

Stigma related to mental ill-health appears to be one of the reasons why privacy is an important factor associated with trust in digital mental health. Stigma is one of the most significant barriers to in-person mental health help seeking among young people [41]. Digital health interventions have the advantage of anonymity if their privacy is well protected [9]. In addition to the methods listed previously to protect users' privacy, young people proposed to have a function to hide the mental health website page from their family or friends, as some users may not be ready to reveal to others that they have mental health concerns. Although we should aim to reduce the stigma itself by promoting interventions such as the WUWE animations and website, functions to reassure young people of their privacy and anonymity should be incorporated.

Recommendations were made to include functions on the platform that enable users to communicate with other young people or mental health professionals, as shown in previous studies [16]. The WUWE website includes a *help seeking* page, which contains the details of a wide variety of mental health organizations with which young people can communicate via text or email, on the phone, or in person. However, this study revealed that young people prefer such communication tools integrated into the website, which allows them to feel more connected to the website and thus increase trust. Future creators

of digital health interventions may include communication functions on the platform by collaborating with mental health organizations that are already providing such services.

Finally, it is important to young people that web-based mental health resources are *comforting*. Regarding the design, consistent with the previous studies on youth digital interventions [42,43], young people reported liking the pleasant and comforting design. Similarly, our participants suggested that the language used on the website needed to be more comforting. For example, the use of the word *help* is perceived to be overwhelming and thus needs to be reconsidered. The word *help* is frequently used in mental health. *Help seeking* is one of the components of mental health literacy [44]. For young people to feel more comfortable and safe, especially when they are distressed, mental health professionals may consider using the word *support seeking* instead.

Limitations and Future Studies

Several limitations of the study should be noted. First, the sample consisted mostly of female participants. Although we aimed to recruit more male participants by exclusively recruiting them through a recruitment website, we gained only limited interest from men. Second, because of the COVID-19 restrictions during the study, we conducted all TrustScapes FGs via the web. Although 96% of all households in the United Kingdom have internet access and 99% of young people in the United Kingdom have a smartphone [45,46], it is possible that some young people with limited access to technology were prevented from participating in the study. To be more inclusive, future researchers should conduct FGs both in person and via the web.

Third, owing to time constraint, the TrustScapes FGs were conducted only after the website was promoted to the public. Although our young cocreators (ie, young people who cocreated the animations and website) provided valuable feedback and some adjustments were made accordingly before the launch, it is apparent that the recommendations from the independent sample of young people (ie, the TrustScapes participants) would have made the website even more trustworthy. We were able to make a few changes to the WUWE website based on the findings from TrustScapes (eg, cookie design) so far, but we recommend that future projects should allocate sufficient time before the public launch to adjust the intervention based on feedback from an independent group of end users (ie, those who were not involved in the cocreation process).

Conclusions

This study provided insight into young people's trust-related concerns about and recommendations for cocreated web-based mental health resources. Factors perceived to be trustworthy included the relatable, high-quality design and creators' credibility, whereas those perceived to be untrustworthy included the privacy policy and cookie settings. The findings highlighted the significance of collaborating with end users and industrial partners and the importance of making the trust-enabling factors salient for users. We hope that these findings will inform future creators of web-based mental health resources to make them as trustworthy and effective as possible.

Acknowledgments

This study was funded by the Arts and Humanities Research Council (grant AH/T003804/1—What's Up With Alex? Animated Storytelling for Mental Health Literacy Among Young People). SIJ and EPV acknowledge the financial support provided by the National Institute for Health and Care Research Nottingham Biomedical Research Centre.

Data Availability

Anonymized focus group transcripts are available from the University of Nottingham data repository [47].

Conflicts of Interest

None declared.

Multimedia Appendix 1
TrustScapes worksheet.
[PDF File (Adobe PDF File), 90 KB - mental_v10i1e38346_app1.pdf]

Multimedia Appendix 2
TrustScapes keywords.
[PDF File (Adobe PDF File), 409 KB - mental_v10i1e38346_app2.pdf]

Multimedia Appendix 3
TrustScapes sketches.
[PDF File (Adobe PDF File), 6276 KB - mental_v10i1e38346_app3.pdf]

References

1. Mental health. World Health Organization. URL: https://www.who.int/health-topics/mental-health#tab=tab_2 [accessed 2022-01-14]
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
3. Mental health. World Health Organization. URL: https://www.who.int/health-topics/mental-health#tab=tab_1 [accessed 2022-01-14]
4. Kutcher S, Wei Y, Coniglio C. Mental health literacy: past, present, and future. Can J Psychiatry 2016 Mar;61(3):154-158 [FREE Full text] [doi: 10.1177/0706743715616609] [Medline: 27254090]
5. Ito-Jaeger S, Perez Vallejos E, Curran T, Spors V, Long Y, Liguori A, et al. Digital video interventions and mental health literacy among young people: a scoping review. J Ment Health 2021 May 19:1-11. [doi: 10.1080/09638237.2021.1922642] [Medline: 34006191]
6. Janoušková M, Tušková E, Weissová A, Trančík P, Pasz J, Evans-Lacko S, et al. Can video interventions be used to effectively destigmatize mental illness among young people? A systematic review. Eur Psychiatry 2017 Mar;41:1-9 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.09.008] [Medline: 28049074]
7. Tay J, Tay Y, Klainin-Yobas P. Effectiveness of information and communication technologies interventions to increase mental health literacy: a systematic review. Early Interv Psychiatry 2018 Dec;12(6):1024-1037 [FREE Full text] [doi: 10.1111/eip.12695] [Medline: 29897166]
8. Griffiths K, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87 [FREE Full text] [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
9. Liverpool S, Mota CP, Sales CM, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
10. Farrer L, Gulliver A, Chan JK, Batterham PJ, Reynolds J, Calear A, et al. Technology-based interventions for mental health in tertiary students: systematic review. J Med Internet Res 2013 May 27;15(5):e101 [FREE Full text] [doi: 10.2196/jmir.2639] [Medline: 23711740]
11. Ryan ML, Shochet IM, Stallman HM. Universal online interventions might engage psychologically distressed university students who are unlikely to seek formal help. Advan Mental Health 2014 Dec 17;9(1):73-83.
12. Montagni I, Tzourio C, Cousin T, Sagara JA, Bada-Alonzi J, Horgan A. Mental health-related digital use by university students: a systematic review. Telemed J E Health 2020 Feb;26(2):131-146. [doi: 10.1089/tmj.2018.0316] [Medline: 30888256]
13. Gorczynski P, Sims-schouten W, Hill D, Wilson JC. Examining mental health literacy, help seeking behaviours, and mental health outcomes in UK university students. J Mental Health Train Educ Pract 2017 Mar 13;12(2):111-120. [doi: 10.1108/jmhtep-05-2016-0027]
14. Digital health practices, social media use, and mental well-being among teens and young adults in the U.S. Providence St. Joseph Health Digital Commons. URL: https://digitalcommons.psjhealth.org/publications/1093/ [accessed 2022-03-29]
15. Tseng S, Fogg BJ. Credibility and computing technology. Commun ACM 1999 May;42(5):39-44. [doi: 10.1145/301353.301402]
16. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. J Med Internet Res 2018 Dec 13;20(12):e11254 [FREE Full text] [doi: 10.2196/11254] [Medline: 30545807]
17. Rowe R, Calnan M. Trust relations in health care: developing a theoretical framework for the "new" NHS. J Health Organ Manag 2006;20(5):376-396. [doi: 10.1108/14777260610701777] [Medline: 17087401]
18. MUIR BM. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 2007 May 31;37(11):1905-1922. [doi: 10.1080/00140139408964957]
19. Rowley J, Johnson F. Understanding trust formation in digital information sources: the case of Wikipedia. J Inform Sci 2013 Mar 06;39(4):494-508. [doi: 10.1177/0165551513477820]
20. Kitchens B, Harle C, Li S. Quality of health-related online search results. Decision Support Systems 2014 Jan;57:454-462 [FREE Full text] [doi: 10.1016/j.dss.2012.10.050]
21. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
22. Chan JK, Farrer LM, Gulliver A, Bennett K, Griffiths KM. University students' views on the perceived benefits and drawbacks of seeking help for mental health problems on the internet: a qualitative study. JMIR Hum Factors 2016 Jan 19;3(1):e3 [FREE Full text] [doi: 10.2196/humanfactors.4765] [Medline: 27026140]
23. Wies B, Landers C, Ienca M. Digital mental health for young people: a scoping review of ethical promises and challenges. Front Digit Health 2021 Sep 6;3:697072 [FREE Full text] [doi: 10.3389/fdgth.2021.697072] [Medline: 34713173]
24. What's up with everyone? homepage. What's up with everyone?. URL: https://whatsupwitheveryone.com/ [accessed 2021-02-08]
25. Academics join forces with Aardman to tackle mental health crisis. UK Research and Innovation. URL: https://www.ukri.org/news/academics-join-forces-with-aardman-to-tackle-mental-health-crisis/ [accessed 2021-11-01]

26. Jirotka M, Grimpe B, Stahl B, Eden G, Hartswood M. Responsible research and innovation in the digital age. Commun ACM 2017 Apr 24;60(5):62-68.
27. Ito-Jaeger S, Perez Vallejos E, Curran T, Crawford P. What's Up With Everyone? A qualitative study on young people's perceptions of cocreated online animations to promote mental health literacy. Health Expect 2022 Aug;25(4):1633-1642 [FREE Full text] [doi: 10.1111/hex.13507] [Medline: 35507731]
28. UnBias: emancipating users against algorithmic biases for a trusted digital economy. UnBias. URL: https://unbias.wp.horizon.ac.uk/ [accessed 2022-12-08]
29. Rovatsos M, Mittelstadt B, Koene A. Landscape Summary: Bias in Algorithmic Decision-Making: what is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it? UK Government. 2019. URL: https://www.research.ed.ac.uk/en/publications/landscape-summary-bias-in-algorithmic-decision-making-what-is-bia [accessed 2022-12-03]
30. Fairness toolkit. Unbias. URL: https://unbias.wp.horizon.ac.uk/fairness-toolkit/ [accessed 2022-12-08]
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. Syan A, Lam JY, Huang CG, Smith MS, Darnay K, Hawke LD, et al. The Wellness Quest: a health literacy and self-advocacy tool developed by youth for youth mental health. Health Expect 2021 Apr;24(2):659-669 [FREE Full text] [doi: 10.1111/hex.13214] [Medline: 33636052]
33. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
34. Fogg B. Prominence-interpretation theory: explaining how people assess credibility online. In: Proceedings of the CHI '03 Extended Abstracts on Human Factors in Computing Systems. 2003 Presented at: CHI03: Human Factors in Computing Systems; Apr 5 - 10, 2003; Ft. Lauderdale Florida USA. [doi: 10.1145/765891.765951]
35. Grant C, Widnall E, Cross L, Simonoff E, Downs J. Informing the development of an E-platform for monitoring wellbeing in schools: involving young people in a co-design process. Res Involv Engagem 2020;6:51 [FREE Full text] [doi: 10.1186/s40900-020-00219-0] [Medline: 32908677]
36. Calderón JL, Shaheen M, Hays RD, Fleming ES, Norris KC, Baker RS. Improving diabetes health literacy by animation. Diabetes Educ 2014 May;40(3):361-372 [FREE Full text] [doi: 10.1177/0145721714527518] [Medline: 24676274]
37. George S, Moran E, Duran N, Jenders RA. Using animation as an information tool to advance health research literacy among minority participants. AMIA Annu Symp Proc 2013;2013:475-484 [FREE Full text] [Medline: 24551351]
38. Mayer RE, Moreno R. Animation as an aid to multimedia learning. Educ Psychol Rev 2002;14:87-99. [doi: 10.1023/A:1013184611077]
39. Meppelink CS, van Weert JC, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015 Jan 13;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]
40. Age appropriate design: a code of practice for online services. Information Commisioner's Office. 2017. URL: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services-2-1.pdf [accessed 2022-03-29]
41. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
42. Monshat K, Vella-Brodrick D, Burns J, Herrman H. Mental health promotion in the Internet age: a consultation with Australian young people to inform the design of an online mindfulness training programme. Health Promot Int 2012 Jun;27(2):177-186. [doi: 10.1093/heapro/dar017] [Medline: 21398335]
43. Syan A, Lam JYY, Huang CGL, Smith MSM, Darnay K, Hawke LD, et al. The Wellness Quest: A health literacy and self-advocacy tool developed by youth for youth mental health. Health Expect 2021 Apr;24(2):659-669 [FREE Full text] [doi: 10.1111/hex.13214] [Medline: 33636052]
44. Kutcher S, Wei Y, Coniglio C. Mental health literacy: past, present, and future. Can J Psychiatry 2016 Mar;61(3):154-158 [FREE Full text] [doi: 10.1177/0706743715616609] [Medline: 27254090]
45. Smartphone ownership penetration in the United Kingdom (UK) in 2012-2022, by age. Statista. 2022 Oct 18. URL: https://www.statista.com/statistics/271851/smartphone-owners-in-the-united-kingdom-uk-by-age/ [accessed 2021-11-01]
46. Internet access – households and individuals, Great Britain: 2020. Office for National Statistics. 2020 Aug 7. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2020 [accessed 2022-03-29]
47. TrustScapes transcripts: exploring young people's trust-related concerns about co-created online resources to promote mental health literacy. University of Nottingham. URL: https://doi.org/10.17639/nott.7250 [accessed 2022-12-22]

Abbreviations

FG: focus group
WUWE: What's Up With Everyone

Edited by J Torous; submitted 30.03.22; peer-reviewed by J Rowley, JJ Wu; comments to author 10.05.22; revised version received 04.10.22; accepted 26.10.22; published 09.01.23.

Please cite as:
Ito-Jaeger S, Perez Vallejos E, Logathasan S, Curran T, Crawford P
Young People's Trust in Cocreated Web-Based Resources to Promote Mental Health Literacy: Focus Group Study
JMIR Ment Health 2023;10:e38346
URL: https://mental.jmir.org/2023/1/e38346
doi:10.2196/38346
PMID:36622752

Viewpoint

From Social Network to Peer Support Network: Opportunities to Explore Mechanisms of Online Peer Support for Mental Health

Amy Rayland[1], MSc; Jacob Andrews[1,2,3], PhD

[1]Academic Unit of Mental Health and Clinical Neuroscience, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[2]National Institute for Health and Care Research Mindtech Medtech Co-operative, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[3]National Institute for Health and Care Research Nottingham Biomedical Research Centre, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
Jacob Andrews, PhD
National Institute for Health and Care Research Mindtech Medtech Co-operative
Institute of Mental Health
University of Nottingham
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 1157484 218
Email: jacob.andrews@nottingham.ac.uk

Abstract

An increasing number of psychological interventions are shifting to online modes of delivery. One such intervention is peer-to-peer support, which in this context may provide internet users living with mental health disorders an opportunity to connect with and support others living with similar conditions. This paper presents a call for further research into how platforms such as Facebook could be used as channels for peer support and the mechanisms that may underlie their effectiveness. We discuss the background of peer support, how it has transitioned online, and consider theories and models that may have relevance. We also consider the importance of moderation within online peer support and the development of specific social network–based online interventions. We conclude that for social network sites to be used as peer-to-peer support interventions, more research is needed to understand their effectiveness, the role of moderation in these communities, and the mechanisms that produce the benefits experienced by users.

(JMIR Ment Health 2023;10:e41855) doi:10.2196/41855

KEYWORDS

peer-to-peer support; Facebook; social networking sites; mental health; moderation

Introduction

Mental illness can be described as "health conditions that involve changes in emotions, thinking or behavior (or a combination of these)" [1]. Mental illnesses range from "common mental disorders" (CMDs), such as depression and anxiety, to psychotic and personality disorders [2]. Research indicates that 1 in 6 adults in England had a CMD in 2014 [3], and recent evidence suggests that the effects of the COVID-19 pandemic have exacerbated mental illness, with almost 1 in 5 adults in the United Kingdom experiencing some form of depression in 2021 [4]. Furthermore, 16% of children younger than 16 years were identified as having a probable mental health disorder in the same period [5].

Mental illnesses have a significant impact on quality of life, social functioning, work, and family activities. Many of those living with mental illness have educational difficulties, being less likely to finish school or enter college [6]. This is compounded by social and relationship problems, vulnerability to abuse, and significant social stigma and discrimination. Aside from the individuals themselves, family members of those with a mental illness often take on caring responsibilities and experience chronic stress from the emotional burden of care. It is estimated that mental illnesses cost the UK economy at least £117.9 billion (US $141 billion) annually (amounting to around 5% of the United Kingdom's gross domestic product), with almost three-quarters of this being attributable to the loss of productivity and the costs associated with unpaid informal care [7].

As the demands for mental health services exceed physical resources, it is more important now than ever that individuals are able to access free and widely available support. The use of social media is widespread across the world, with use in developing countries exceeding that in developed countries [8]. The growth of social media has seen it being used for many different reasons, including communication, marketing, advertising, media sharing, and entertainment. More recently, as the widespread accessibility of the internet for clinical interventions has been noticed, there has been an increased interest in using social media to deliver online interventions for mental health. Facebook, as the most popular social media platform [9], provides all the necessary tools to make it a suitable location for peer-to-peer support groups. However, as this is an emerging area of research, little is known about the effectiveness of Facebook-mediated peer support or what potential mechanisms may underlie any behavioral and attitudinal changes observed as a result of its use.

This viewpoint paper considers the background to peer support, how it is currently being used online, and the scope for Facebook to be used in this area, as well as suggesting some models and theories that may help to explain the mechanisms underlying effective online peer support. We suggest next steps in research to enable the development and rollout of online peer-to-peer support interventions for the benefit of those living with mental illness.

Traditional Peer Support

Peer support within mental health can be defined as support or services provided to individuals experiencing mental health problems by others who have experienced similar problems [10]. Peer support is not based on psychiatric models or diagnostic criteria but can be understood as an extension of the natural human tendency to respond compassionately to shared difficulty [11]. Peers, seen as equals, use their lived experiences to provide “been there” empathy, insight, encouragement, and assistance and to inculcate hope in a reciprocal relationship [11,12].

The roots of peer support services lie in long-established groups, such as Alcoholics Anonymous [13], designed to build upon peer support that occurs naturally between people [14]. Mental health service providers recognized the benefits of these groups and have since introduced peer support programs across the United Kingdom [13], with peer workers employed in various roles, including facilitating mutual support groups, providing one-to-one support, or running alternative services [15]. As services are provided by nonprofessionals, they can be available in a community setting at a relatively low cost compared to professional services.

The effectiveness of peer support has been examined in relation to multiple outcome measures. A recent meta-analysis of 28 randomized controlled trials investigating the effectiveness of peer support for individuals with a range of mental illnesses demonstrated significant improvements in clinical recovery (ie, measures of psychiatric symptoms) and personal recovery (ie, measures of perceived recovery, sense of purpose, and personal agency) but not functional recovery (ie, measures of quality of life and the degree of vocational and social functioning) [16]. Functional recovery did, however, show significant improvements for those with serious mental disorders. Peer support also has a positive impact on psychosocial and recovery outcomes [15], which include increased community integration, sense of control, social functioning, and social support; peer support also promotes feelings of empowerment and hope [17-20]. Furthermore, peer support can improve self-esteem, self-efficacy, and self-management of difficulties [21].

Despite evidence supporting their effectiveness, these formal services are usually only available through mental health services, which are facing pressures due to cuts in funding [22] and demand exceeding available National Health Service resources [23].

Online Peer Support

Recent figures suggest that 99% of people aged 16 to 44 years in the United Kingdom are internet users, with 89% of UK adults reporting using the internet daily [24]. There is a growing body of evidence demonstrating the benefits of internet interventions for mental health conditions, including depression and anxiety [25]. Interventions delivered online eliminate geographical barriers, can be accessed and used at any time, and may be more cost-effective than traditional services for young people aged 12 to 25 years [26].

Peer support can be provided informally online. This support is more informal in that it is not tied to services and can be accessed by anyone on the internet. Asynchronous platforms, such as forums, discussion groups, and bulletin boards, allow users to post topics and engage in discussions, using these platforms to exchange knowledge, ask for help, discuss problems, and share stories [27]. Synchronous platforms, which can be text-based chat rooms or purposely created virtual worlds, host real-time communication between users [28].

A systematic review indicated that internet support groups (ISGs) are most often used for emotional and informational support (more often experiential knowledge than “second-hand” professional knowledge), which creates a sense of social companionship [29]. This review found that 10 of 16 single-intervention or cross-sectional trials showed a positive effect of ISGs on depression symptoms, although study quality was rated as low, and the studies covered a range of mental and physical health conditions [30]. In one of these studies, which specifically examined individuals with depression, those who were more frequent users of the ISG (defined as using the group for 5 or more hours per week) showed a higher likelihood of depression resolution than less frequent users [30].

A review of online peer support for individuals with depression suggested several benefits, including increased empowerment and self-efficacy, enhanced coping strategies, and reduced social isolation [31]. However, there was limited empirical evidence to validate these descriptive findings. In a systematic review of online peer support for adolescents and young adults, Ali and colleagues [32] found mixed results, with only 2 of the 6 studies included in the review showing evidence of a positive effect. These 2 studies demonstrated reductions in anxiety [33] and

higher abstinence from smoking [28] after using online peer support. A further review of digital peer support for mental health found early evidence of feasibility, acceptability, and effectiveness among all studies, though it reported that the studies were lacking in measurements of fidelity, limiting understanding of underlying mechanisms [34].

Qualitatively, many advantages of online synchronous peer support platforms have been identified, including emotional support, the availability of advice, enablement of positive personal changes, engagement in valuable social interactions, and the opportunity to disclose and express feelings and views [35,36]. Furthermore, it has been suggested that online forums can function as platforms for peer education—being "experiential experts," peers can provide others with health-related information, such as potential interventions or treatments or how to behave or cope in the situation they are facing, and a greater understanding of how others experience the same illness. This is often information that individuals don't have access to offline [27].

However, users have also identified disadvantages, including the potential for others' experiences to cause personal distress, restrictive rules and moderation, and unhelpful interactions with others [36]. A survey study exploring adverse events in the use of online peer support found social exclusion and emotional contagion to be other potential risks [37]. Research has also raised the need for guidelines to safeguard vulnerable people, both as users and as moderators of online mental health peer support groups [38]. Furthermore, certain groups, including those with more serious mental illness, may be less likely to engage with the internet, and so may be less inclined to use these groups, missing out on any benefits they provide [39].

Social Media for Peer Support

Interventions for mental health, such as online peer-to-peer support, have the potential to exploit the increasing use of social media [40,41]. Social media provides users the ability to connect with similar others without fear of stigma or judgement, to create supportive networks and communities to discuss mental health, and to control their own actions, choosing how much to post and interact with others [40,42,43]. Evidence suggests that social media is already being used by individuals with mental illness for support—in a survey of young adults (aged 18-35 years) with a self-identified mental illness, social media was used for sharing personal experiences (68%), connecting with others with a mental illness (66%), learning coping strategies (50%), and learning about mental illness from others (42%) [38].

The most popular social media platform in the United Kingdom is Facebook [9], with 23% of users aged 18 to 24 years and 31% of users aged 25 to 34 years [44]. Recent evidence suggests that Facebook is used by around 90% of people aged 12 to 34 years who identify as having a mental illness [44,45]. Facebook is a free-to-use platform that has similar features to both synchronous and asynchronous platforms, with users able to post content and engage in discussions with others, and the messenger tool allows personal communication with others when group users may not want to share personal stories or information on the main feed. Therefore, Facebook is being increasingly used to host peer-to-peer support groups. These can be private or public, with anyone able to read and post content on public groups, but only those granted access by an individual with administration rights can see content in private groups. Moderators also on occasion set up live chat rooms where users can meet and speak live, on or off camera, with others in the group. Facebook peer support groups for mental health are typically self-organized in an informal way, rather than being organized, run, or moderated by medical professionals.

Currently, there is little research into the use of Facebook groups for mental health peer support. In a content and thematic analysis of public Facebook groups for mental health peer-to-peer support, Prescott and colleagues [46] found that groups are used to share personal experiences, to request advice, to find informational support such as signposting to services, and to seek emotional support. However, this was a passive analysis and did not involve gaining any personal insights, personal information, or confirmation of diagnoses from participants. In a qualitative evaluation of a private Facebook group, Watkins and colleagues [47] found the group to be well-liked by participants, who found it educational and thought-provoking and found that it enabled them to build relationships with others and have conversations they may not have felt comfortable having face-to-face.

Social Network Site–Based Interventions

As social media has developed as a platform for delivering peer support, some researchers have taken this further to develop a novel form of digital intervention called social network site (SNS)-based interventions. These interventions have an SNS format, combining individually tailored therapy with clinical and peer moderation within a peer-to-peer social support network [48,49]. While initial pilot and small-scale studies have demonstrated positive outcomes following the use of these interventions, further research is required to support their efficacy and to understand how the peer support element can best be utilized to support users.

Moderation

Many online peer support groups and SNS-based interventions involve moderation, which is important for building a safe and positive community [50]. Moderation within online communities can improve intention to participate [51] and the quality of users' contributions [52]. Furthermore, moderation has been identified as a key component for the success of online peer-to-peer interventions [53], with the most favorable being those guided by moderators, who are perceived as caring, supportive, and friendly by users.

Previous studies of moderators of various health-related online support groups for both mental and physical health conditions have found that most moderators are individuals living with the condition themselves. Some moderators described how moderating online communities fulfilled their own support needs and reported that it was empowering to be able to help others. Other motivations for setting up groups included addressing a

lack of good existing support groups, reducing isolation, and having a place to exchange support and information. Moderators described various roles they took within these groups, such as circulating health information; moderating messages; providing support, encouragement, and advice; making announcements; and performing administrative tasks, such as responding to requests, banning harmful users, and organizing discussion threads [54-56].

In a study specifically examining moderation within mental health communities on Reddit [57], moderators' motivations for starting Reddit threads included feeling good for helping others, taking the opportunity for leadership, improving the community, and spreading awareness about treatment. Their roles and responsibilities included clearing up spam and troll posts, handling posts on suicide or self-harm, and providing social support. In the same study, Saha and colleagues [57] found that the involvement of medical professionals as moderators or group members could be helpful to answer questions that other group members may have, but it was considered inappropriate to have medical professionals diagnose others via the group.

While moderators have many critical roles within these groups, there may be some inherent risks given they are often vulnerable individuals themselves. Moderation may become overwhelming, taking up much time and energy, and moderators must set boundaries to focus on their own health. Moderators are required to make decisions on censoring potentially harmful content, to which posters may react angrily [36,37]. It is possible that moderators may become stressed with the responsibility of looking after members. Despite these initial studies investigating moderation within online health communities, research has not yet considered the roles and experiences of group moderators within peer-to-peer support groups specifically mediated via Facebook.

Models and Theories

Several models and theories have been proposed for the mechanisms underlying offline peer support. Gillard and colleagues [58] proposed a change model for peer support worker interventions. Change models provide an understanding of how processes within an intervention are associated with outcomes. The model suggests the primary mechanism is building trusting relationships based on peer workers sharing lived experiences and understanding service users' experiences. Two parallel mechanisms flow from this: (1) peer workers role-modeling recovery and social functioning and (2) bridging the gap to professional services and the community. The model suggests these mechanisms lead to changes in outcomes such as hope, empowerment, social functioning, self-care, engagement with services, and strength of social networks. Chinman and colleagues [59] suggested these mechanisms show face validity, but no studies have explored their explanatory power. Moreover, no research has investigated whether these mechanisms can be applied to online peer support environments.

It is as yet unclear whether models like these transfer to online peer support, and there is a gap in research in this area [14,59,60]. However, some researchers have set out new models describing what benefits people seek from online networks in particular, and how these benefits may be achieved. Naslund and colleagues [61] proposed a theoretical model for online peer support, incorporating 3 main opportunities provided by such networks. The first, challenging stigma, relates to individuals connecting with similar others and feeling more comfortable expressing themselves due to the anonymity afforded by online peer support networks. The second, increasing consumer activation, suggests online settings enable individuals to learn what to expect from a condition, how to cope with it, and how to approach important health care decisions. The final opportunity suggests that online networks facilitate access to other (online and offline) interventions. No studies have so far empirically tested whether this model is supported by users of Facebook groups for peer-to-peer support.

Aside from these models, various theories from across the discipline of psychology have been suggested to underlie peer support, including social learning theory, social comparison theory, and helper therapy [62]. Social learning theory, in the context of peer support, suggests that peers with a history of or current mental illness act as role models for peers with similar mental illness [63]. Interacting with a peer perceived to be successfully coping with their illness is suggested to be more likely to result in positive behavior change. Barton and Henderson [64] suggested observing another peer demonstrating desired behaviors with visible consequences motivates others to replicate this behavior.

Social comparison theory suggests that people seek out others they perceive as having things in common with, such as a similar mental illness, to establish a sense of normalcy and identity [65]. Interacting with peers who have successfully managed their illness may lead to aspirations of positive behavior change [66]. Upward comparisons, to those perceived to be successfully managing their illness, can provide hope and promote self-improvement. Downward comparisons, to those believed to be worse off, put into perspective how bad things could be and are proposed to be self-enhancing.

Helper therapy proposes that individuals can benefit themselves through helping others [67]. Suggested benefits to the helper include enhanced self-image, development of their abilities, and the status of the helper role influencing the way they are treated [67,68]. Research has highlighted other benefits, including feeling useful to others, reducing internal stigma, feeling looked up to, and having a sense of achievement and competence [69-72]. Helper therapy has been identified as one of the mechanisms underpinning positive behavior changes with peer support-worker interventions [73].

In a study examining the mechanisms of peer support alongside a web-based psychoeducational program, Proudfoot and colleagues [74] found evidence for social comparison theory, which promotes hope, motivation, and faith in treatments, and helper therapy, which helps provide others awareness of how to manage their illness and promotes feelings of competence and a sense of connectedness to the mental health system. However, this study investigated peer support alongside another intervention, and research has yet to explore the mechanisms

underpinning behavior and attitude changes in peer support when it is delivered alone in online peer support settings.

When considering how these models and theories may be applied to online peer support, it will also be important to consider the extent to which users engage with and contribute to online groups. A common measure is frequency of posts, which has been suggested to broadly fit the 90-9-1 principle, stating that a majority of content will be posted by a small proportion of users (1%), while the vast majority of users (90%) choose to observe rather than participate actively [75]. However, this is a restricted, unidimensional measure, and reviews of multidisciplinary participatory styles have suggested there are up to 41 different styles of engagement found across various online health communities [76]. Further research is required to understand engagement categories specifically within mental health support groups on Facebook and how this engagement ties into the potential mechanisms of online peer support that have been proposed.

Research Perspectives

As we emerge into the aftermath of the coronavirus pandemic, it is important now to understand and act on the toll that the pandemic has taken on mental health. There are indications that COVID-19 can have a direct impact on mental well-being through mental and neurological manifestations, with many of those hospitalized with COVID-19 experiencing anxiety or depression [77]. The pandemic also impacted on mental well-being indirectly through grief for those dying as a result of contracting the virus, heightened loneliness, and isolation from social distancing and lockdown measures, as well as anxiety and uncertainty about the future [78]. Financial insecurity will have contributed to poorer well-being for many, with the number of those claiming benefits due to unemployment rising [79], food insecurity quadrupling, and an estimated 1.1 million people at the end of 2020 facing poverty [78]. Young people in particular have seen their mental health significantly worsen over the course of the pandemic [80].

Research shows promise for the use of online SNSs, including Facebook, to provide forums where mental health peer support can take place, with a variety of theories that could plausibly underpin their use. Novel online peer support platforms making use of social network functions may therefore offer a useful way to support people with mental health difficulties. However, it is currently difficult to use or promote online peer-to-peer support as an intervention when we do not have a quantification of the effectiveness of peer support provided in this way or a clear understanding of the benefits its users experience. We argue that in order for existing social network platforms like Facebook to be recommended as peer-to-peer support interventions and for separate peer support interventions to be developed based on SNSs, more research is needed to understand their effectiveness, making use of both quantitative and qualitative methods.

We suggest that the effectiveness of novel social network–based peer support interventions should be evaluated using similar experimental designs as applied in prior research into offline peer support [16,32-34] with a primary focus on symptomology (using both clinical and self-reported outcomes) and psychosocial outcomes, such as self-efficacy in managing disorders, quality of life, hope, knowledge, empowerment, and social isolation. Qualitative evidence could usefully be gathered on existing peer support groups on Facebook to describe processes involved in online peer support, such as how groups are created, how they function, and how groups differ in terms of purpose and target audience.

It is important also to be mindful of the limitations and potential harms of online peer support. The use of peer support may result in adverse events, such as behavioral contagion, corumination, or unpleasant or negative interactions with other users [36,37,81-83]. There is a need to apply principles of responsible research and innovation to future initiatives to develop online peer support, to anticipate potential harms and consequences of these approaches, and to engage with users to surmount these issues wherever possible [84].

Finally, exploring the mechanisms that underlie effective online peer support is crucial to understanding why it is helpful, and how its utility can be maximized. Particularly, in order to develop new interventions that are based on sites like Facebook, that is, SNS-based interventions, it will be important to understand whether theories underpinning offline peer support also apply to online peer-to-peer support in this context. It is also crucial to understand the role of moderation within these communities, and the benefits that moderators have for the workings of online peer support, including safeguarding, to clearly demonstrate their role in future social network–based peer support interventions.

Deriving knowledge from existing online peer support groups (ie, those on Facebook) will be beneficial for the design of new online peer support interventions. Research in this area has the potential to provide a better understanding of why and how these networks help, and how they can best be organized and managed for maximum effect while reducing risk of harms.

Acknowledgments

This research was supported by the National Institute for Health and Care Research Biomedical Research Centre Nottingham and the National Institute for Health and Care Research MindTech MedTech Co-operative. The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health and Care Research, or the Department of Health and Social Care.

Conflicts of Interest

None declared.

References

1. What is Mental Illness? American Psychiatric Association. URL: https://psychiatry.org/patients-families/what-is-mental-illness [accessed 2023-02-16]
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Washington, DC: American Psychiatric Publishing; 2013.
3. McManus S, Bebbington P, Jenkins R, Brugha T. Adult Psychiatric Morbidity Survey: Survey of Mental Health and Wellbeing, England, 2014. National Health Service (NHS Digital). 2014. URL: https://tinyurl.com/3jnnpv8j [accessed 2023-02-16]
4. Coronavirus and depression in adults, Great Britain: July to August 2021. Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/coronavirusanddepressioninadultsgreatbritain/julytoaugust2021 [accessed 2023-02-16]
5. Mental Health of Children and Young People in England, 2020: Wave 1 follow up to the 2017 survey. National Health Service (NHS Digital). URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2020-wave-1-follow-up [accessed 2023-02-16]
6. Kessler RC, Foster CL, Saunders WB, Stang PE. Social consequences of psychiatric disorders, I: Educational attainment. Am J Psychiatry 1995 Jul;152(7):1026-1032. [doi: 10.1176/ajp.152.7.1026] [Medline: 7793438]
7. McDaid D, Park A, Davidson G, John A, Knifton L, McDaid S, et al. The economic case for investing in the prevention of mental health conditions in the UK. Mental Health Foundation. 2022 Feb. URL: https://www.mentalhealth.org.uk/sites/default/files/2022-06/MHF-Investing-in-Prevention-Full-Report.pdf [accessed 2023-02-16]
8. Poushter J, Stewart R. Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones. Pew Research Center. 2018. URL: https://www.pewresearch.org/global/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/ [accessed 2023-02-16]
9. Tankovska H. Social media usage in the United Kingdom (UK) - Statistics and facts. Statista. URL: https://www.statista.com/topics/3236/social-media-usage-in-the-uk/ [accessed 2021-06-22]
10. Davidson L, Chinman M, Sells D, Rowe M. Peer support among adults with serious mental illness: a report from the field. Schizophr Bull 2006 Jul;32(3):443-450 [FREE Full text] [doi: 10.1093/schbul/sbj043] [Medline: 16461576]
11. Penney D. Defining "peer support": Implications for policy, practice and research. Advocates for Human Potential. 2018. URL: https://www.ahpnet.com/AHPNet/media/AHPNetMediaLibrary/White%20Papers/DPenney_Defining_peer_support_2018_Final.pdf [accessed 2023-02-16]
12. Chinman M, George P, Dougherty RH, Daniels AS, Ghose SS, Swift A, et al. Peer support services for individuals with serious mental illnesses: assessing the evidence. Psychiatr Serv 2014 Apr 01;65(4):429-441. [doi: 10.1176/appi.ps.201300244] [Medline: 24549400]
13. Watson E. Chapter 2: What is peer support? history, evidence and values. In: Watson E, Meddings S, editors. Peer Support in Mental Health. London, UK: Macmillan Education UK; 2019:16-31.
14. Lloyd-Evans B, Mayo-Wilson E, Harrison B, Istead H, Brown E, Pilling S, et al. A systematic review and meta-analysis of randomised controlled trials of peer support for people with severe mental illness. BMC Psychiatry 2014 Feb 14;14:39 [FREE Full text] [doi: 10.1186/1471-244X-14-39] [Medline: 24528545]
15. White S, Foster R, Marks J, Morshead R, Goldsmith L, Barlow S, et al. The effectiveness of one-to-one peer support in mental health services: a systematic review and meta-analysis. BMC Psychiatry 2020 Nov 11;20:534 [FREE Full text] [doi: 10.1186/s12888-020-02923-3] [Medline: 33176729]
16. Smit D, Miguel C, Vrijsen JN, Groeneweg B, Spijker J, Cuijpers P. The effectiveness of peer support for individuals with mental illness: systematic review and meta-analysis. Psychol Med 2022 Sep 06:1-10 [FREE Full text] [doi: 10.1017/S0033291722002422] [Medline: 36066104]
17. Corrigan PW. Impact of consumer-operated services on empowerment and recovery of people with psychiatric disabilities. Psychiatr Serv 2006 Oct;57(10):1493-1496. [doi: 10.1176/ps.2006.57.10.1493] [Medline: 17035571]
18. Davidson L, Bellamy C, Guy K, Miller R. Peer support among persons with severe mental illnesses: a review of evidence and experience. World Psychiatry 2012 Jun;11(2):123-128 [FREE Full text] [doi: 10.1016/j.wpsyc.2012.05.009] [Medline: 22654945]
19. Ochocka J, Nelson G, Janzen R, Trainor J. A longitudinal study of mental health consumer/survivor initiatives: Part 3—A qualitative study of impacts of participation on new members. J Community Psychol 2006 May;34(3):273-283 [FREE Full text] [doi: 10.1002/jcop.20099]
20. Yanos PT, Primavera LH, Knight EL. Consumer-run service participation, recovery of social functioning, and the mediating role of psychological factors. Psychiatr Serv 2001 Apr;52(4):493-500. [doi: 10.1176/appi.ps.52.4.493] [Medline: 11274496]
21. King AJ, Simmons MB. A systematic review of the attributes and outcomes of peer work and guidelines for reporting studies of peer interventions. Psychiatr Serv 2018 Sep 01;69(9):961-977. [doi: 10.1176/appi.ps.201700564] [Medline: 29962310]
22. Gilburt H. Mental health under pressure. The King's Fund. 2015. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/mental-health-under-pressure-nov15_0.pdf [accessed 2023-02-16]

23. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265 [FREE Full text] [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
24. Internet access - Households and individuals, Great Britain: 2020. Office for National Statistics. 2020. URL: https://tinyurl.com/4tvzvype [accessed 2023-02-16]
25. Cunningham JA, Gulliver A, Farrer L, Bennett K, Carron-Arthur B. Internet interventions for mental health and addictions: current findings and future directions. Curr Psychiatry Rep 2014 Dec;16(12):521. [doi: 10.1007/s11920-014-0521-5] [Medline: 25308390]
26. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]
27. Tanis M. Health-related on-line forums: what's the big attraction? J Health Commun 2008;13(7):698-714. [doi: 10.1080/10810730802415316] [Medline: 18958781]
28. Woodruff SI, Conway TL, Edwards CC, Elliott SP, Crittenden J. Evaluation of an Internet virtual world chat room for adolescent smoking cessation. Addict Behav 2007 Sep;32(9):1769-1786. [doi: 10.1016/j.addbeh.2006.12.008] [Medline: 17250972]
29. Griffiths KM, Calear AL, Banfield M, Tam A. Systematic review on internet support groups (ISGs) and depression (2): What is known about depression ISGs? J Med Internet Res 2009 Sep 30;11(3):e41 [FREE Full text] [doi: 10.2196/jmir.1303] [Medline: 19793718]
30. Griffiths KM, Calear AL, Banfield M. Systematic review on internet support groups (ISGs) and depression (1): Do ISGs reduce depressive symptoms? J Med Internet Res 2009 Sep 30;11(3):e40 [FREE Full text] [doi: 10.2196/jmir.1270] [Medline: 19793719]
31. Melling B, Houguet-Pincham T. Online peer support for individuals with depression: a summary of current research and future considerations. Psychiatr Rehabil J 2011;34(3):252-254. [doi: 10.2975/34.3.2011.252.254] [Medline: 21208865]
32. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
33. Ellis L, Campbell A, Sethi S, O'Dea B. Comparative randomised trial of an online cognitive-behavioural therapy program and an online support group for depression and anxiety. Journal of Cyber Therapy & Rehabilitation 2011;4(4):461-467.
34. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
35. Setoyama Y, Yamazaki Y, Namayama K. Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters. J Med Internet Res 2011 Dec 29;13(4):e122 [FREE Full text] [doi: 10.2196/jmir.1696] [Medline: 22204869]
36. Griffiths KM, Reynolds J, Vassallo S. An online, moderated peer-to-peer support bulletin board for depression: user-perceived advantages and disadvantages. JMIR Ment Health 2015;2(2):e14 [FREE Full text] [doi: 10.2196/mental.4266] [Medline: 26543919]
37. Easton K, Diggle J, Ruethi-Davis M, Holmes M, Byron-Parker D, Nuttall J, et al. Qualitative exploration of the potential for adverse events when using an online peer support network for mental health: cross-sectional survey. JMIR Ment Health 2017 Oct 30;4(4):e49 [FREE Full text] [doi: 10.2196/mental.8168] [Medline: 29084710]
38. Strand M, Eng LS, Gammon D. Combining online and offline peer support groups in community mental health care settings: a qualitative study of service users' experiences. Int J Ment Health Syst 2020;14:39 [FREE Full text] [doi: 10.1186/s13033-020-00370-x] [Medline: 32514303]
39. Spanakis P, Heron P, Walker L, Crosland S, Wadman R, Newbronner E, et al. Use of the internet and digital devices among people with severe mental ill health during the COVID-19 pandemic restrictions. Front Psychiatry 2021;12:732735 [FREE Full text] [doi: 10.3389/fpsyt.2021.732735] [Medline: 34630183]
40. Sanger E. Social networking in mental health interventions for adolescents. Perspect Public Health 2022 Sep;142(5):261-262 [FREE Full text] [doi: 10.1177/1757913920924431] [Medline: 33000675]
41. Naslund JA, Aschbrenner KA, McHugo GJ, Unützer J, Marsch LA, Bartels SJ. Exploring opportunities to support mental health care using social media: A survey of social media users with mental illness. Early Interv Psychiatry 2019 Jun;13(3):405-413 [FREE Full text] [doi: 10.1111/eip.12496] [Medline: 29052947]
42. Chung JE. Social networking in online support groups for health: how online social networking benefits patients. J Health Commun 2014;19(6):639-659. [doi: 10.1080/10810730.2012.757396] [Medline: 23557148]
43. Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: qualitative study. J Med Internet Res 2010 Dec 19;12(5):e70 [FREE Full text] [doi: 10.2196/jmir.1550] [Medline: 21169176]
44. Dixon S. Distribution of Facebook users worldwide as of January 2022, by age and gender. Statista. 2022 Nov 10. URL: https://www.statista.com/statistics/376128/facebook-global-user-age-distribution/ [accessed 2023-02-16]
45. Birnbaum ML, Rizvi AF, Confino J, Correll CU, Kane JM. Role of social media and the Internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2017 Aug;11(4):290-295 [FREE Full text] [doi: 10.1111/eip.12237] [Medline: 25808317]

46. Prescott J, Rathbone AL, Brown G. Online peer to peer support: Qualitative analysis of UK and US open mental health Facebook groups. Digit Health 2020;6:2055207620979209 [FREE Full text] [doi: 10.1177/2055207620979209] [Medline: 33354335]
47. Watkins DC, Allen JO, Goodwill JR, Noel B. Strengths and weaknesses of the Young Black Men, Masculinities, and Mental Health (YBMen) Facebook project. Am J Orthopsychiatry 2017;87(4):392-401 [FREE Full text] [doi: 10.1037/ort0000229] [Medline: 27977287]
48. Gleeson JFM, Alvarez-Jimenez M, Lederman R. Moderated online social therapy for recovery from early psychosis. Psychiatr Serv 2012 Jul;63(7):719. [doi: 10.1176/appi.ps.20120p719] [Medline: 22752039]
49. Lederman R, Wadley G, Gleeson J, Bendall S, Álvarez-Jiménez M. Moderated online social therapy: Designing and evaluating technology for mental health. ACM Trans Comput Hum Interact 2014 Feb 01;21(1):1-26. [doi: 10.1145/2513179]
50. Webb M, Burns J, Collin P. Providing online support for young people with mental health difficulties: challenges and opportunities explored. Early Interv Psychiatry 2008 May;2(2):108-113. [doi: 10.1111/j.1751-7893.2008.00066.x] [Medline: 21352141]
51. Wise K, Hamman B, Thorson K. Moderation, response rate, and message interactivity: features of online communities and their effects on intent to participate. J Comp Mediated Comm 2006 Oct;12(1):24-41 [FREE Full text] [doi: 10.1111/j.1083-6101.2006.00313.x]
52. Chen J, Xu H, Whinston AB. Moderated online communities and quality of user-generated content. J Manag Inf Syst 2014 Dec 08;28(2):237-268. [doi: 10.2753/mis0742-1222280209]
53. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244 [FREE Full text] [doi: 10.2196/12244] [Medline: 30563811]
54. Coulson N, Shaw R. Nurturing health-related online support groups: Exploring the experiences of patient moderators. Comput Human Behav 2013;29(4):1695-1701 [FREE Full text] [doi: 10.1016/j.chb.2013.02.003]
55. Smedley R, Coulson N. A thematic analysis of messages posted by moderators within health-related asynchronous online support forums. Patient Educ Couns 2017 Sep;100(9):1688-1693. [doi: 10.1016/j.pec.2017.04.008] [Medline: 28454663]
56. van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar MAFJ. Patient-initiated online support groups: motives for initiation, extent of success and success factors. J Telemed Telecare 2010;16(1):30-34. [doi: 10.1258/jtt.2009.001009] [Medline: 20086265]
57. Saha K, Ernala S, Dutta S, Sharma E, De Choudhury M. Understanding moderation in online mental health communities. In: Meiselwitz G, editor. Social Computing and Social Media Participation, User Experience, Consumer Experience, and Applications of Social Computing. HCII 2020. Lecture Notes in Computer Science, vol 12195. Cham, Switzerland: Springer; 2020.
58. Gillard S, Gibson SL, Holley J, Lucock M. Developing a change model for peer worker interventions in mental health services: a qualitative research study. Epidemiol Psychiatr Sci 2015 Oct;24(5):435-445 [FREE Full text] [doi: 10.1017/S2045796014000407] [Medline: 24992284]
59. Chinman M, McInnes DK, Eisen S, Ellison M, Farkas M, Armstrong M, et al. Establishing a research agenda for understanding the role and impact of mental health peer specialists. Psychiatr Serv 2017 Sep 01;68(9):955-957 [FREE Full text] [doi: 10.1176/appi.ps.201700054] [Medline: 28617205]
60. Rogers ES. Peer support services: state of the workforce-state of the field in the USA. Ment Health Soc Incl 2017 Jun 12;21(3):168-175. [doi: 10.1108/mhsi-03-2017-0015]
61. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
62. Salzer MS. Consumer-delivered services as a best practice in mental health care delivery and the development of practice guidelines. Psychiatr Rehabil Ski 2002 Dec;6(3):355-382. [doi: 10.1080/10973430208408443]
63. Bandura A. Social Learning Theory. New York, NY: General Learning Press; 1977.
64. Barton J, Henderson J. Peer support and youth recovery: a brief review of the theoretical underpinnings and evidence. CJFY 2016 Jan 27;8(1):1-17 [FREE Full text] [doi: 10.29173/cjfy27140]
65. Festinger L. A Theory of Social Comparison Processes. Hum Relat 1954;7(2):117-140. [doi: 10.1177/001872675400700202]
66. Glass TA, De Leon CFM, Bassuk SS, Berkman LF. Social engagement and depressive symptoms in late life: longitudinal findings. J Aging Health 2006 Aug 30;18(4):604-628. [doi: 10.1177/0898264306291017] [Medline: 16835392]
67. Riessman F. The "helper" therapy principle. Social Work 1965 Apr;10(2):27-32. [doi: 10.1093/sw/10.2.27]
68. Skovholt TM. The client as helper: a means to promote psychological growth. Couns Psychol 1974;4(3):58-64. [doi: 10.1177/001100007400400308]
69. Austin E, Ramakrishnan A, Hopper K. Embodying recovery: a qualitative study of peer work in a consumer-run service setting. Community Ment Health J 2014 Nov;50(8):879-885 [FREE Full text] [doi: 10.1007/s10597-014-9693-z] [Medline: 24464104]
70. Moran GS, Russinova Z, Gidugu V, Yim JY, Sprague C. Benefits and mechanisms of recovery among peer providers with psychiatric illnesses. Qual Health Res 2012 Mar;22(3):304-319. [doi: 10.1177/1049732311420578] [Medline: 21900694]

71. Mourra S, Sledge W, Sells D, Lawless M, Davidson L. Pushing, patience, and persistence: peer providers' perspectives on supportive relationships. Am J Psychiatr Rehabil 2014 Dec 13;17(4):307-328. [doi: 10.1080/15487768.2014.967601]
72. Rebeiro Gruhl KL, LaCarte S, Calixte S. Authentic peer support work: challenges and opportunities for an evolving occupation. J Ment Health 2016;25(1):78-86. [doi: 10.3109/09638237.2015.1057322] [Medline: 26397981]
73. Watson E. The mechanisms underpinning peer support: a literature review. J Ment Health 2019 Dec;28(6):677-688. [doi: 10.1080/09638237.2017.1417559] [Medline: 29260930]
74. Proudfoot JG, Jayawant A, Whitton AE, Parker G, Manicavasagar V, Smith M, et al. Mechanisms underpinning effective peer support: a qualitative analysis of interactions between expert peers and patients newly-diagnosed with bipolar disorder. BMC Psychiatry 2012 Nov 09;12:196 [FREE Full text] [doi: 10.1186/1471-244X-12-196] [Medline: 23140497]
75. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014 Feb 04;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
76. Carron-Arthur B, Ali K, Cunningham JA, Griffiths KM. From help-seekers to influential users: a systematic review of participation styles in online health communities. J Med Internet Res 2015 Dec 01;17(12):e271 [FREE Full text] [doi: 10.2196/jmir.4705] [Medline: 26627369]
77. Mao L, Jin H, Wang M, Hu Y, Chen S, He Q, et al. Neurologic manifestations of hospitalized patients with Coronavirus Disease 2019 in Wuhan, China. JAMA Neurol 2020 Jun 01;77(6):683-690 [FREE Full text] [doi: 10.1001/jamaneurol.2020.1127] [Medline: 32275288]
78. Jones L, Lewis C, Ubido J, Taylor C, Hefferon C, Baird S, et al. Direct and indirect impacts of COVID-19 on health and wellbeing: Rapid evidence review. Psychological Professions Network. 2020 Jul 30. URL: https://ppn.nhs.uk/attachments/article/2797/2020-07-direct-and-indirect-impacts-of-covid19-on-health-and-wellbeing.pdf [accessed 2023-02-16]
79. McCurdy C. Local Differences: Responding to the local economic impact of coronavirus. Resolution Foundation. 2020 Jun 14. URL: https://www.resolutionfoundation.org/app/uploads/2020/06/Local-differences.pdf [accessed 2023-02-16]
80. Takino S, Hewlett E, Nishina Y, Prinz C. Supporting young people's mental health through the COVID-19 crisis. Organisation for Economic Co-operation and Development. 2021 May 12. URL: https://www.oecd.org/coronavirus/policy-responses/supporting-young-people-s-mental-health-through-the-covid-19-crisis-84e143e5/ [accessed 2023-02-16]
81. Stevens EA, Prinstein MJ. Peer contagion of depressogenic attributional styles among adolescents: a longitudinal study. J Abnorm Child Psychol 2005 Feb;33(1):25-37. [doi: 10.1007/s10802-005-0931-2] [Medline: 15759589]
82. Mueller AS, Abrutyn S. Suicidal disclosures among friends: using social network data to understand suicide contagion. J Health Soc Behav 2015 Mar;56(1):131-148 [FREE Full text] [doi: 10.1177/0022146514568793] [Medline: 25722129]
83. Hankin BL, Stone L, Wright PA. Corumination, interpersonal stress generation, and internalizing symptoms: accumulating effects and transactional influences in a multiwave study of adolescents. Dev Psychopathol 2010;22(1):217-235 [FREE Full text] [doi: 10.1017/S0954579409990368] [Medline: 20102657]
84. Jirotka M, Grimpe B, Stahl B, Hartswood M, Eden G. Responsible research and innovation in the digital age. Commun ACM 2017;60(5):62-68. [doi: 10.1145/3064940]

Abbreviations

CMD: common mental disorder
ISG: internet support group
SNS: social network site

Edited by J Torous; submitted 12.08.22; peer-reviewed by P Martin, D Smit; comments to author 04.10.22; revised version received 14.12.22; accepted 02.01.23; published 28.02.23.

Please cite as:
Rayland A, Andrews J
From Social Network to Peer Support Network: Opportunities to Explore Mechanisms of Online Peer Support for Mental Health
JMIR Ment Health 2023;10:e41855
URL: https://mental.jmir.org/2023/1/e41855
doi:10.2196/41855
PMID:36853738

Viewpoint

Technology Implementation for Mental Health End Users: A Model to Guide Digital Transformation for Inpatient Mental Health Professionals

Jessa Lin Westheimer[1], BSA; Nidal Moukaddam[2,3], MD, PhD; Jan A Lindsay[2,4], PhD; Ashutosh Sabharwal[5], MS, PhD; Bijan Najafi[6], MSc, PhD; Peter A Iacobelli[1], BSc; Robert J Boland[2,7], MD; Michelle A Patriquin[1,2], ABPP, PhD

[1]Research Department, The Menninger Clinic, Houston, TX, United States

[2]Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, United States

[3]Adult Outpatient Services, Ben Taub Hospital, Houston, TX, United States

[4]Michael E DeBakey VA Medical Center, Houston, TX, United States

[5]Department of Electrical and Computer Engineering, Rice University, Houston, TX, United States

[6]Michael E DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX, United States

[7]The Menninger Clinic, Houston, TX, United States

Corresponding Author:
Jessa Lin Westheimer, BSA
Research Department
The Menninger Clinic
12301 S Main St
Houston, TX, 77035
United States
Phone: 1 832 738 8300
Email: jwestheimer@menninger.edu

Abstract

Digital transformation is the adoption of digital technologies by an entity in an effort to increase operational efficiency. In mental health care, digital transformation entails technology implementation to improve the quality of care and mental health outcomes. Most psychiatric hospitals rely heavily on “high-touch” interventions or those that require in-person, face-to-face interaction with the patient. Those that are exploring digital mental health care interventions, particularly for outpatient care, often copiously commit to the “high-tech” model, losing the crucial human element. The process of digital transformation, especially within acute psychiatric treatment settings, is in its infancy. Existing implementation models outline the development of patient-facing treatment interventions within the primary care system; however, to our knowledge, there is no proposed or established model for implementing a new provider-facing ministration tool within an acute inpatient psychiatric setting. Solving the complex challenges within mental health care demands that new mental health technology is developed in concert with a use protocol by and for the inpatient mental health professional (IMHP; the end user), allowing the “high-touch” to inform the “high-tech” and vice versa. Therefore, in this viewpoint article, we propose the Technology Implementation for Mental-Health End-Users framework, which outlines the process for developing a prototype of an IMHP-facing digital intervention tool in parallel with a protocol for the IMHP end user to deliver the intervention. By balancing the design of the digital mental health care intervention tool with IMHP end user resource development, we can significantly improve mental health outcomes and pioneer digital transformation nationwide.

(JMIR Ment Health 2023;10:e40429) doi:10.2196/40429

KEYWORDS

digital transformation; user-centered design; innovation; implementation science; user acceptability; wearables; mental health; implementation; technology implementation

Background

Digital transformation in mental health care has reshaped care delivery over the last few years and has been catalyzed particularly due to the global COVID-19 pandemic. Digital transformation is the adoption of digital technologies by an entity in an effort to increase operational efficiency [1]. In mental health care, digital transformation entails technology

implementation to improve the quality of care and clinical outcomes. Digital transformation makes possible an expanded "full continuum of care" for inpatient psychiatry, which is desperately needed.

Earlier studies have shown that the success of a digital transformation initiative is predicated on transparency, user-centered design, and iteration [2]. Additionally, it is important to balance the design of the digital health care intervention with end user resource development [1,3,4]. Other implementation models for digital transformation have been proposed [4-8]. For example, the Accelerated Creation-to-Sustainment (ACTS) model outlines a framework for implementing a sustainable digital mental health intervention in a real-world treatment setting that emphasizes rapid, iterative, user-centered design [5]. This model describes an end-to-end process for accelerating the research-to-practice pipeline for digital mental health care innovations. While the ACTS model importantly incorporates user-centered design and sustainable implementation, this model is intended to develop a patient-facing treatment intervention within the primary care system. To our knowledge, there is not yet a comparable model for implementing a new provider-facing ministration tool within an acute inpatient psychiatric setting.

Inpatient mental health professionals (IMHPs) are vulnerable to experiencing burnout [9,10]. Sources of stress and burnout unique to the roles of IMHPs include demanding therapeutic relationships, difficult and sometimes violent patients, and patient safety management (where the risk for suicide and self-harm is constant) [9,10]. Because of the inpatient setting's unpredictability, the adaptability of an experienced IMHP is relied on heavily to ensure quality patient care. Digital transformation offers the opportunity to offset the stress and workload of IMHPs by leveraging the continuous availability and predictive power of technology. Importantly, the unique demands for an IMHP demand a unique approach to digitally transforming clinician-facing tools within an inpatient psychiatric setting to ensure their successful implementation and that the tools reduce, not increase, stress and burnout.

Therefore, we propose the Technology Implementation for Mental-Health End-Users (TIME) framework in inpatient mental health care (Figure 1), which outlines the process of developing a prototype of an IMHP-facing intervention (Figure 1A) in parallel with a protocol for an IMHP to deliver the intervention (Figure 1B). Similar to the ACTS model, the TIME framework is grounded in well-defined frameworks for user-centered design, software development, and implementation science. The TIME model also emphasizes participatory design—the active involvement of the end user in the design process—so the ultimate deliverable has inherent buy-in and utility [11,12].

An important feature of the TIME framework is its balance of the "high-tech" prototype with the "high-touch" protocol. Most inpatient psychiatric hospitals rely heavily on "high-touch" interventions, that is, interventions that require in-person, face-to-face interaction with the patient [13]. Conversely, those that are developing digital mental health care interventions often fully commit to the "high-tech" model, losing the crucial human element. By leveraging the continuous availability and predictive power of technology in concert with the adaptability and experience of clinical care providers, we can significantly improve mental health outcomes and pioneer digital transformation. While the term "blended care" refers to the blending of technology and face-to-face treatment for a patient-facing intervention [14], the TIME framework produces a form of "blended ministration" where an IMHP-facing technology is blended with an IMHP-facing implementation blueprint (Figure 2).

Figure 1. The Technology Implementation for Mental-Health End-Users framework for inpatient mental health care. (A) "High-tech prototype" refers to a technology-based, inpatient mental health professional (IMHP)–facing mental health care innovation or intervention. The top steps outline the development of a high-tech prototype. (B) "High-touch protocol" refers to the human side of mental health care (how the prototype will be used by the IMHPs). The bottom steps outline the development of a high-touch protocol.

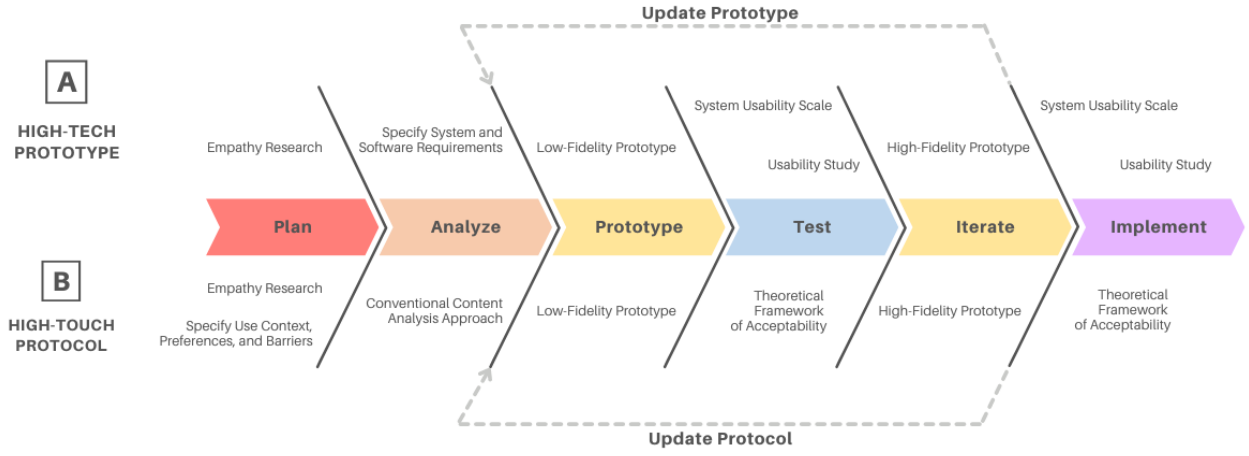
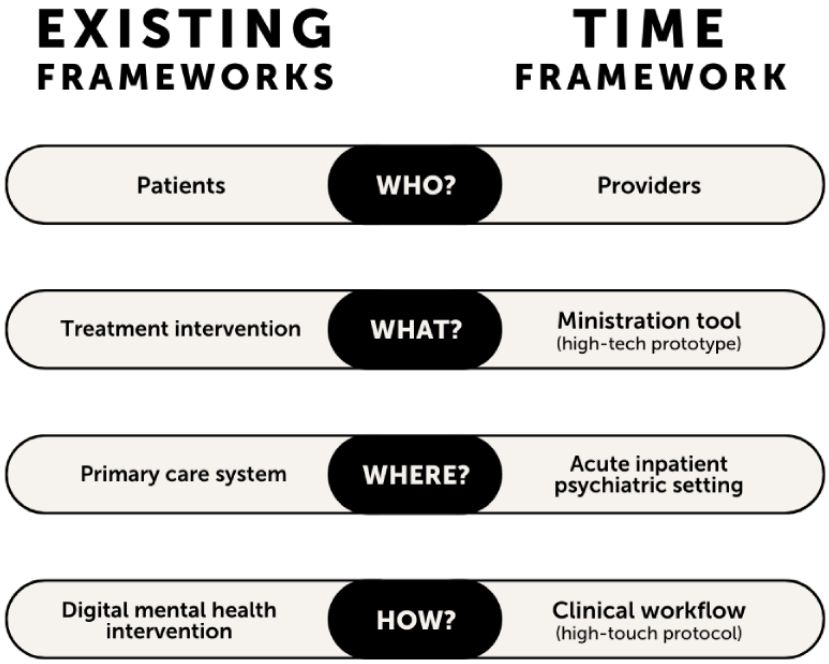


Figure 2. Comparing the Technology Implementation for Mental-Health End-Users (TIME) framework with existing frameworks, such as the Accelerated Creation-to-Sustainment (ACTS) model.



Operational Definitions

Operational definitions are provided in Table 1.

Table 1. Operational definitions.

Term	Definition
High-tech	Referring to technological innovations and interventions
High-touch	Referring to the human side of mental health care (clinical care staff interacting directly or face-to-face with the patient)
Wearable [15,16]	Electronic devices that can be worn on or close to the user's body to monitor vital signs
End user	The final user of the intervention and protocol (in the TIME[a] model, typically nursing or clinical care staff)
Empathy research [11,17]	The design process of empathizing with the end user of the project through interviews and contextual inquiry

[a]TIME: Technology Implementation for Mental-Health End-Users.

Challenges

Technology Acceptance

In the Technology Acceptance Model, an individual's acceptance of a new technology is determined by its perceived usefulness and perceived ease of use [18]. Therefore, perceived usefulness and perceived ease of use govern individuals' attitudes toward the new technology and, in turn, their use of it [18]. Therefore, the successful implementation of an IMHP-facing technology in inpatient psychiatric care must consider the perceived use of the IMHP ("will they trust this innovation?"), the perceived ease of use of the IMHP ("is this innovation technologically unambiguous?"), and the attitude of the IMHP ("will this innovation positively contribute to their treatment of patients?").

Technology acceptance was tested significantly during the global COVID-19 pandemic [19]. Quarantine meant an immediate and massive digital migration, where digital immigrants found themselves thrust into unavoidable technology and unprecedented connectivity [20,21]. By the same token, the demographic landscape of the mental health care system comprises predominantly digital immigrants—generally considered those who are older than 40 years and were born before the "proliferation of digital technologies." In 2022, the average age of a mental health care professional in the United States is 44 years, with 60% of the professionals being older than 40 years [22]. Thus, designers of digital interventions for mental health must carefully consider usability, inclusivity, and accessibility [23,24].

Perceived Dispensability of User-Centered Design

User-centered design is an approach to development wherein the intended users of the system (end users) are actively and intentionally involved in its development. There are three major tenets of this framework, defined by Gould and Lewis [25] as follows:

1. Front-load with empathy: understanding the end users' needs and expectations from the project's inception ensures *customized solutions*.
2. Measure usability empirically: quantifying the end users' satisfaction throughout project development ensures *usability*.
3. Design and test iteratively: designing and testing the system iteratively with the end users' input ensures *acceptability*.

Despite its proven effectiveness in developing digital systems, there is still a widely held belief that applying user-centered design is superfluous to the development life cycle [26]. It is common for developer-driven needs to take priority over those of the user [26,27]. Additionally, there is a stigma that user-centered design is time-consuming and costly to the developer; on the contrary, by involving the intended user throughout the design process, usability problems are identified and resolved before the system is ever deployed, saving time and money [26-28]. By emphasizing empathy research, empirical measures of usability and iterative design based on user-centered design principles increase the likelihood for successful implementation and acceptability. Digital transformation in inpatient psychiatric care will require a shift to the user-driven development of novel technologies and interventions.

Health Care Disparities and Technology Implementation

Internet and smartphone availability are ubiquitous in the general population. However, this ubiquity is curtailed by the following two primary factors: socioeconomic disparities and mental illness symptoms. A closer look reveals that the first gap is in smartphone ownership and internet access. Of patients discharged from inpatient psychiatric units in the United States, 82% are estimated to have smartphones and internet access [29]; however, in high-income households, that number is close to 100% [30]. This digital divide correlates with socioeconomic and racial disparities. For example, low-income Hispanic and Black youth are less likely to have home or smartphone internet access [31]. So, while digital technologies promise to bring individualized high-tech monitoring, their effect could be limited by a return to underresourced environments upon discharge from an inpatient setting.

The second gap relates to the severity of mental illness symptoms that affect trust in and use of technology. In a survey of individuals with severe mental illness (SMI) on internet and smartphone use during the global COVID-19 pandemic, 61.6% of them were limited users or nonusers of the internet, mostly using phones to communicate with their families and friends [32]. A quarter of the surveyed sample—markedly among older individuals with schizophrenia—reported being weary of internet use for "security concerns." Thus, while most mobile health interventions report favorable feasibility outcomes, they typically do not have the same success rate in including patients with SMI. Studies that document significant web-based participation among individuals with SMI demonstrate a correlation with positive recovery and positive emotions toward technology [33]. These findings suggest that attitudes of people with SMI meaningfully affect successful digital transformation. There are gaps in our understanding of how delusions, paranoia, and other severe symptoms may impact the scalability of the TIME framework as well, and these are intriguing areas for continued research and development.

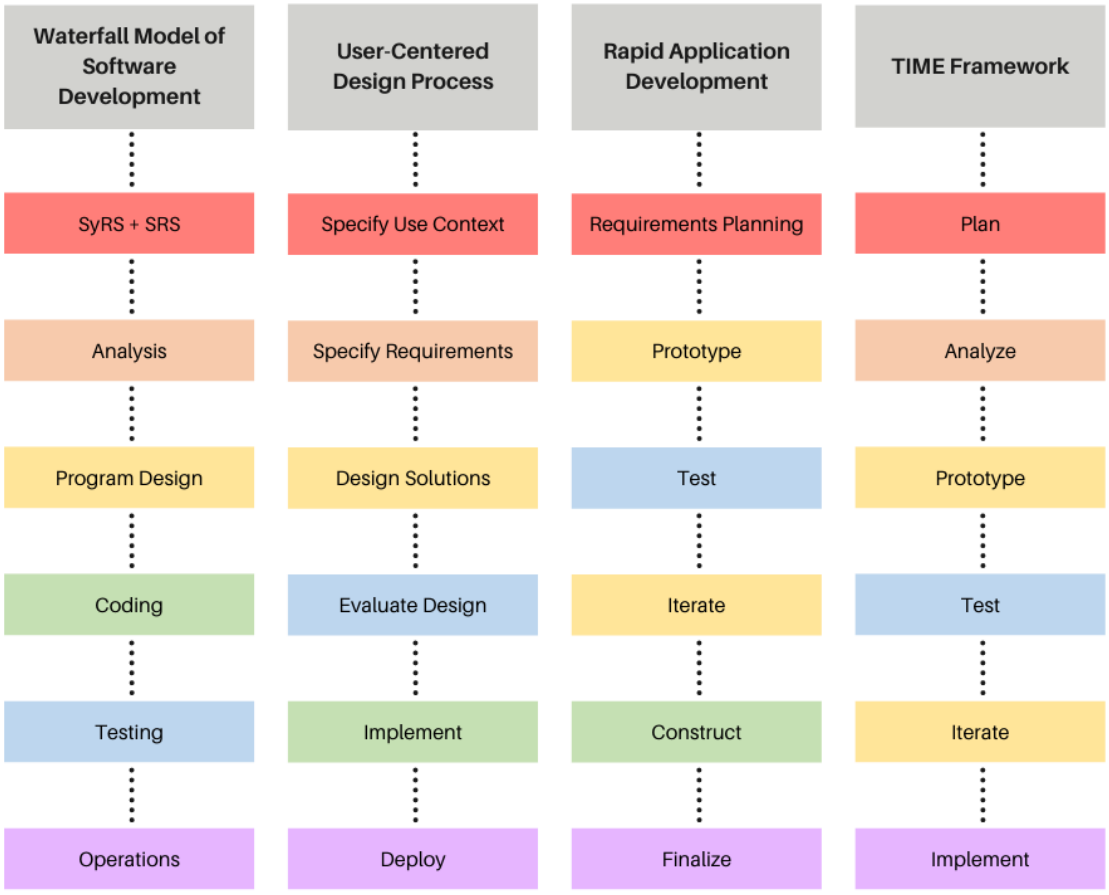
The TIME Framework

Overview

When we have more technology available than ever before, the entire onus of patient safety should not fall on the IMHP. An important feature of the TIME framework is its balance of the "high-tech" prototype with the "high-touch" protocol. Most psychiatric hospitals rely heavily on "high-touch" interventions, that is, interventions that require in-person, face-to-face interaction with the patient [13]. Those that are exploring digital mental health care interventions, particularly for outpatient care, often copiously commit to the "high-tech" model, losing the crucial human element [34-37]. Solving the complex challenges faced by mental health care providers demands that new mental health technology tools are developed in concert with a use protocol by and for the end user (mental health care provider), allowing the "high-touch" to inform the "high-tech" and vice versa.

High-tech interventions and their corresponding high-touch protocols must be developed alongside each other to ensure usability and end user acceptability. Thus, we have defined a model—the TIME framework—for developing novel IMHP-facing tools and their clinical workflows alongside each other (Figure 1). This model is grounded in established frameworks for user-centered design and digital technology development (Figure 3).

Figure 3. The Technology Implementation for Mental-Health End-Users (TIME) framework is grounded in design processes across disciplines. The TIME framework combines elements from the Waterfall Model of Software Development [38], the user-centered design process [11,17], and rapid application development [39]. Many of these processes contain corresponding steps and are color coded as follows: planning steps describing information gathering for the project, including empathy and user-based research (red); analysis steps describing how any information gathered in a planning step is further evaluated to define the project's scope and requirements (orange); prototype or iterate steps describing the iterative process of prototyping the design solution (yellow); construct steps describing the creation of a final deliverable (green); test steps describing the evaluation of the design solution and collection of feedback from end users (blue); and implementation of steps describing how the design solution is implemented (purple).



Phases of the TIME Framework

Plan Phase

At the core of the *Plan* phase of the TIME framework is empathy research [11,17]. During this phase, interviews are conducted with the IMHP end user of the mental health care tool. Prior to this phase, the clinical providers' problem has been identified. The goal of these interviews will be to understand the needs and expectations of the IMHP end user for a high-tech solution to the clinical problem. Importantly, this entails building an in-depth understanding of the current clinical workflows and characterizing the distinct needs of different types of IMHPs. For instance, a day shift mental health nurse would have different responsibilities from those of a night shift mental health nurse (eg, nighttime safety checks). Likewise, barriers and preferences for the use protocol will be defined, so careful consideration must be made for institutional policies and regulation (ie, crisis intervention protocols, such as seclusion and restraint, hospital documentation guidelines, and others). Empathizing at the onset operationalizes the end users' needs in the solutions we design.

Analyze Phase

The feedback from empathy research will be analyzed to specify prototype requirements. Understanding the functional and emotional needs of the end user is essential to human-centered design. In the *Analyze* phase, qualitative discourse is anatomized to scope both the technical requirements for the IMHP-facing digital intervention and the key features of its use protocol within the context of the inpatient psychiatric hospital setting. In applying the TIME framework, it is possible to have different IMHP end user groups. For example, in designing a new electronic health record system, the end users could be mental health nurses, social workers, psychologists, psychiatrists, researchers, physicians, administrative assistants, and others. Developing a firm understanding of both the unique and overlapping needs of each group during this phase will ensure a more functional prototype. Further, the needs of these mental health professionals differ significantly from those of general medical professionals [40,41]. A common misstep in establishing digital transformation in an inpatient psychiatric setting is the presumption that an intervention for physical health can be superficially adapted to mental health. For instance, among the chief concerns during the implementation of an electronic health record system within a mental health hospital in England was that the technology was "not fit for [a mental

health] purpose" and much customization would be required for the preexisting system to meet the needs of their setting [42].

Prototype Phase

Before resources are dedicated to developing sophisticated technologies, low-fidelity prototypes of initial designs will be created and evaluated in the *Prototype* phase of the TIME framework. The digital intervention prototype will capture the basic architecture and user flow of the technology. A major benefit of digital transformation is the potential to manage exponentially larger data sets. In an inpatient psychiatric setting, consideration of the end users' ultimate data needs will inform prototype configuration. Innovations developed in an inpatient psychiatric setting may serve discrete goals for different IMHP end users. For instance, in developing a new data management system for clinical outcomes, a clinician would need item-level responses to outcome measures, while a researcher may only need unit-wide average scores. The use protocol prototype will outline the procedure for using the technology within the context of the IMHP's role. Similar to the implementation blueprint in the ACTS model, this will be a critical part of facilitating the transition to a usable intervention [5].

Test Phase

During the *Test* phase, the low-fidelity prototypes will be evaluated by the end users through usability studies adapted for ascertaining qualitative and quantitative feedback to update the design. Key performance indicators (time on task, use of navigation vs search, conversion rate, error rate, and drop-off rate) and adapted system usability scales will be defined to guide quantitative evaluation. To evaluate the prototype's usability and the protocol's effectiveness, usability studies will be designed and carried out with simulated practical scenarios in the target inpatient setting. Additional moderated usability studies will be conducted to assess more qualitative feedback about the prototype and the clarity of its use protocol. Testing and feedback review with the IMHP end users are necessary for participatory design and ensuring the usability of the final solution.

Iterate Phase

Human-centered design is inherently iterative. Adapting each design iteration to tailor to the preferences and behaviors of the IMHP end user will produce better, more acceptable digital mental health care interventions. During design, developers maintain a high level of control over a prototype; therefore, the evaluation of the prototype's behavior in a real-world clinical care setting is necessary to make it truly functional in its final state. In the *Iterate* phase, the IMHP end user feedback gathered during the usability studies informs the development of the high-fidelity prototypes.

Implement Phase

Applying the iterative process of prototyping and testing, the high-fidelity prototypes are evaluated by way of a usability study again in the *Implement* phase. Repeated evaluations ensure that any changes made are improvements to the overall design and acceptability by the IMHP end user. Additionally, the ultimate goal is to deploy the prototype and protocol in a real-world inpatient treatment setting, which is inherently unpredictable and chaotic. Reevaluating updated designs is helpful for verifying that the IMHPs' feedback is being captured and the technology can stand up to a variable treatment environment.

The *Test→Iterate→Implement* sequence is repeated as many times as necessary to achieve an effective solution (digital intervention and its use protocol).

Hypothetical Example

To clarify the phases of our TIME model, we apply it as a framework for developing a nurse station monitoring dashboard to visualize wearable-based data in real time.

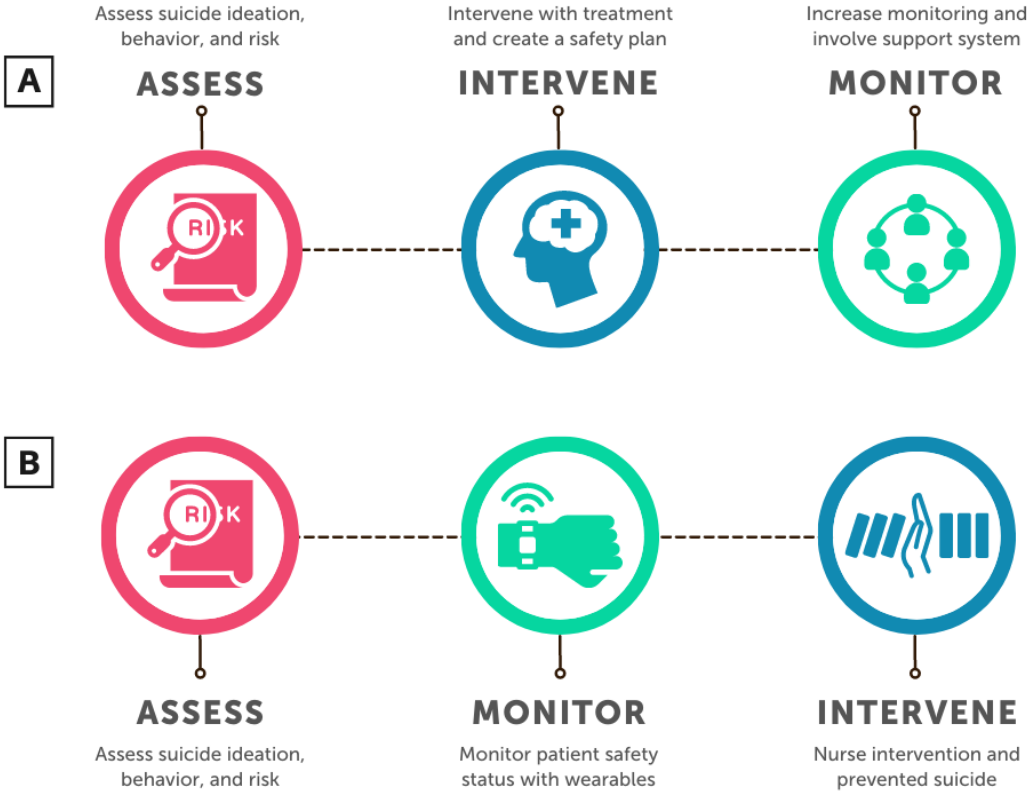
In an inpatient hospital setting, there is an average of 56.7 patient suicides a year [43]. Further, 74% of those occur while the patient is undergoing psychiatric treatment. The issue of suicide is pressing. The Assess, Intervene, and Monitor for Suicide Prevention model proposes a clinical management framework to *assess* suicide risk, *intervene* with coping strategies, and *monitor* high-risk patients [44] (Figure 4A). This is a high-touch solution. We propose that by leveraging the continuous availability and predictive power of technology through digital transformation, we may be able to rearrange the Assess, Intervene, and Monitor model as follows: (1) assess suicide risk, (2) monitor patient safety status by using wearable devices, and (3) intervene and prevent suicide every time (Figure 4B). This is a high-tech and high-touch solution. The capitalization on the continuous, noninvasive availability of technology has been described as "digital phenotyping" [13]. We can draw upon wearable technology to provide insight and predictive power for a patient's status without having to be present in front of them. Not only will this monitoring dashboard display a patient's current state (awake, asleep, walking, sitting, etc) as measured by the wearable device, but also it will serve as a "digital smoke alarm" for a critical change in the patient's status [13]. Feedback provided by IMHP end users will inform both the development of the dashboard's user interface and the crisis intervention protocol when a dashboard alert is issued.

Below, we outline how the TIME framework can be implemented to develop the high-tech remote patient monitoring dashboard in parallel with the high-touch dashboard alert protocol:

- *Plan:* we define the needs and scope for both the dashboard and the dashboard alert protocol by interviewing IMHP end users (mental health associates and nurses from each shift) from each inpatient unit to specify their needs and expectations.
- *Analyze:* we define the expected behavior and functionality of the monitoring dashboard and crisis intervention protocol based on what we learned from the IMHPs.
- *Prototype:* we create low-fidelity prototypes for both the dashboard and the dashboard alert protocol based on the requirements specified in the *Analyze* phase.
- *Test:* in a focus group, end users will interact with the monitoring dashboard and be trained on the dashboard alert protocol in a moderated usability study. The efficacy of the dashboard prototype and the implementability and instructional clarity of the dashboard alert protocol will be

evaluated with defined key performance indicators (ie, time on task, system usability scale, conversion rate, and error rate).
- *Iterate:* we will use the IMHPs' feedback from the *Test* phase to iterate high-fidelity prototypes for the monitoring dashboard and the dashboard alert protocol.
- *Implement:* we will again conduct focus groups to evaluate the monitoring dashboard and dashboard alert protocol in a simulated scenario within the target inpatient setting. The feedback will be used to inform further iteration or produce the final deliverables.

Figure 4. Combining high-tech and high-touch solutions to rearrange the Assess, Intervene, and Monitor for Suicide Prevention (AIM-SP) model. (A) The AIM-SP model [44]. (B) The digitally transformed AIM-SP model incorporating digital technology through remote monitoring can improve suicide prevention outcomes.



Conclusions

The last few years have exposed an ongoing public mental health crisis through isolation, health anxiety, job loss, etc [45,46]. To meet these mental health needs—branded by increased public safety concerns and mental health issues—incorporating technology into mental health treatment is critical. At present, there is no established model for facilitating digital transformation—developing technology alongside a protocol for its eventual use—for provider-facing technology tools in inpatient psychiatry.

We propose an innovative model for designing a high-tech IMHP tool in parallel with a standardized, high-touch clinical protocol. Emphasizing the active involvement of the IMHP end user at each phase of the process will result in the most usable and acceptable digital intervention. The successful implementation of the TIME framework will provide an avenue for digital transformation in mental health care. To our knowledge, this is the first established model for facilitating digital transformation in inpatient psychiatric settings. Meeting the mental health needs of a postpandemic world by incorporating technology carefully into mental health ministration, especially for our most acute and intensive settings, will be lifesaving.

Acknowledgments

We would like to recognize the members of the research team at the Menninger Clinic for their contributions to research and data collection; this includes Tina Holmes, Nadine Obeid, Anika Wiltgen, Ryan Smith, Julia Myerson, Jillian Thomas, Christopher Shephard, Cameron Johnson, Nesreen Mattar, Danna Ramirez, Chasee Boyd, Christopher Frazier, Tiffany Tran, Hayley Viter, PAI, Dr Hyuntaek Oh, and Dr Katrina Rufino. We would also like to thank the IT department at the Menninger Clinic for their continued technical support.

Conflicts of Interest

None declared.

References

1. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: Analyzing the current state-of-research. J Bus Res 2021 Feb;123:557-567. [doi: 10.1016/j.jbusres.2020.10.030]
2. Benjamin K, Potts H. Digital transformation in government: Lessons for digital health? Digit Health 2018;4:2055207618759168 [FREE Full text] [doi: 10.1177/2055207618759168] [Medline: 29942624]
3. Ifenthaler D, Egloffstein M. Development and implementation of a maturity model of digital transformation. TechTrends 2019 Nov 22;64(2):302-309. [doi: 10.1007/s11528-019-00457-4]
4. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4(1):50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
5. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
6. Curran G, Bauer M, Mittman B, Pyne J, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
7. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
8. Singh S, Germine L. Technology meets tradition: a hybrid model for implementing digital tools in neuropsychology. Int Rev Psychiatry 2021 Jun 25;33(4):382-393. [doi: 10.1080/09540261.2020.1835839] [Medline: 33236657]
9. Sørgaard KW, Ryan P, Hill R, Dawson I, OSCAR Group. Sources of stress and burnout in acute psychiatric care: inpatient vs. community staff. Soc Psychiatry Psychiatr Epidemiol 2007 Oct 13;42(10):794-802. [doi: 10.1007/s00127-007-0228-6] [Medline: 17700977]
10. O'Connor K, Muller Neff D, Pitman S. Burnout in mental health professionals: a systematic review and meta-analysis of prevalence and determinants. Eur Psychiatry 2018 Sep 26;53:74-99. [doi: 10.1016/j.eurpsy.2018.06.003] [Medline: 29957371]
11. Sanders E. From user-centered to participatory design approaches. In: Design and the Social Sciences. Boca Raton, FL: CRC Press; 2002:18-25.
12. Damodaran L. User involvement in the systems design process-a practical guide for users. Behav Inf Technol 1996 Jan;15(6):363-377. [doi: 10.1080/014492996120049]
13. Insel T. Bending the curve for mental health: technology for a public health approach. Am J Public Health 2019 Jun;109(S3):S168-S170. [doi: 10.2105/ajph.2019.305077]
14. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JEWC. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
15. Johnson C, Mathew SJ, Oh H, Rufino KA, Najafi B, Colombo AE, et al. Wearable technology: A promising opportunity to improve inpatient psychiatry safety and outcomes. J Psychiatr Res 2021 Mar;135:104-106. [doi: 10.1016/j.jpsychires.2021.01.003] [Medline: 33465608]
16. Hickey BA, Chalmers T, Newton P, Lin CT, Sibbritt D, McLachlan CS, et al. Smart Devices and Wearable Technologies to Detect and Monitor Mental Health Conditions and Stress: A Systematic Review. Sensors (Basel) 2021 May 16;21(10) [FREE Full text] [doi: 10.3390/s21103461] [Medline: 34065620]
17. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004:445-456.
18. Chuttur M. Overview of the technology acceptance model: Origins, developments and future directions. Sprouts: Working Papers on Information Systems. 2009. URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1289&context=sprouts_all [accessed 2023-03-13]
19. Wheatley L, Brent M. COVID-19 and technology: Trust and connectivity in a new digital world. Deloitte. 2021. URL: https://www2.deloitte.com/nz/en/blog/consulting/2021/covid-19-and-technology-trust-and-connectivity.html [accessed 2023-03-13]
20. Prensky M. Digital natives, digital immigrants part 2: do they really think differently? On the Horizon 2001 Nov;9(6):1-6. [doi: 10.1108/10748120110424843]
21. Amedo F. The rise of digital immigrants: beyond a pandemic. KPMG. 2021. URL: https://www.gnbcc.net/Pix/News/Files/Thought_leadership_2021_The_Rise_of_Digital_Immigrants_25.01.2021.pdf [accessed 2023-03-13]
22. Mental health professional demographics and statistics in the US. Zippa. URL: https://www.zippia.com/mental-health-professional-jobs/demographics/ [accessed 2023-03-13]

23. Seok S, DaCosta B, Hodges R. A systematic review of empirically based universal design for learning: implementation and effectiveness of universal design in education for students with and without disabilities at the postsecondary level. J Soc Sci 2018;06(05):171-189. [doi: 10.4236/jss.2018.65014]
24. Ostroff E. Universal design: an evolving paradigm. In: Universal design handbook. New York, NY: McGraw-Hill; 2011:1.3-2.0.
25. Gould JD, Lewis C. Designing for usability: key principles and what designers think. Commun ACM 1985 Mar;28(3):300-311. [doi: 10.1145/3166.3170]
26. De Vito Dabbs A, Myers B, Mc Curry KR, Dunbar-Jacob J, Hawkins R, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
27. Mayhew D, Bias R. Cost-justifying web usability. In: Human factors and web development. Boca Raton, FL: CRC Press; 2002:63-87.
28. Gustafson D, Robinson T, Ansley D, Adler L, Brennan P. Consumers and evaluation of interactive health communication applications. The Science Panel on Interactive Communication and Health. Am J Prev Med 1999 Jan;16(1):23-29. [doi: 10.1016/s0749-3797(98)00104-4] [Medline: 9894551]
29. Iliescu R, Kumaravel A, Smurawska L, Torous J, Keshavan M. Smartphone ownership and use of mental health applications by psychiatric inpatients. Psychiatry Res 2021 May;299:113806. [doi: 10.1016/j.psychres.2021.113806] [Medline: 33667947]
30. Percentage of U.S. adults who own a smartphone from 2011 to 2021. Statista. 2022. URL: https://www.statista.com/statistics/219865/percentage-of-us-adults-who-own-a-smartphone/ [accessed 2022-06-28]
31. Dolcini MM, Canchola JA, Catania JA, Song Mayeda MM, Dietz EL, Cotto-Negrón C, et al. National-level disparities in internet access among low-income and Black and Hispanic Youth: current population survey. J Med Internet Res 2021 Oct 12;23(10):e27723 [FREE Full text] [doi: 10.2196/27723] [Medline: 34636728]
32. Spanakis P, Heron P, Walker L, Crosland S, Wadman R, Newbronner E, et al. Use of the internet and digital devices among people with severe mental ill health during the COVID-19 pandemic restrictions. Front Psychiatry 2021 Sep 24;12:732735 [FREE Full text] [doi: 10.3389/fpsyt.2021.732735] [Medline: 34630183]
33. Shpigelman CN, Tal A, Zisman-Ilani Y. Digital community inclusion of individuals with serious mental illness: a national survey to map digital technology use and community participation patterns in the digital era. JMIR Ment Health 2021 Sep 21;8(9):e28123 [FREE Full text] [doi: 10.2196/28123] [Medline: 34546177]
34. Rodriguez-Villa E, Rauseo-Ricupero N, Camacho E, Wisniewski H, Keshavan M, Torous J. The digital clinic: implementing technology and augmenting care for mental health. Gen Hosp Psychiatry 2020;66:59-66 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.06.009] [Medline: 32688094]
35. Titov N, Hadjistavropoulos HD, Nielssen O, Mohr DC, Andersson G, Dear BF. From research to practice: ten lessons in delivering digital mental health services. J Clin Med 2019 Aug 17;8(8):1239 [FREE Full text] [doi: 10.3390/jcm8081239] [Medline: 31426460]
36. Taylor CB, Fitzsimmons-Craft EE, Graham AK. Digital technology can revolutionize mental health services delivery: the COVID-19 crisis as a catalyst for change. Int J Eat Disord 2020 Jul 25;53(7):1155-1157 [FREE Full text] [doi: 10.1002/eat.23300] [Medline: 32449523]
37. Molfenter T, Heitkamp T, Murphy AA, Tapscott S, Behlman S, Cody OJ. Use of telehealth in mental health (MH) services during and after COVID-19. Community Ment Health J 2021 Oct 24;57(7):1244-1251 [FREE Full text] [doi: 10.1007/s10597-021-00861-2] [Medline: 34165695]
38. Ruparelia NB. Software development lifecycle models. SIGSOFT Softw Eng Notes 2010 May 11;35(3):8-13. [doi: 10.1145/1764810.1764814]
39. Mackay H, Carne C, Beynon-Davies P, Tudhope D. Reconfiguring the user:: using rapid application development. Soc Stud Sci 2016 Jun 29;30(5):737-757. [doi: 10.1177/030631200030005004]
40. Hanrahan NP, Aiken LH, McClaine L, Hanlon AL. Relationship between psychiatric nurse work environments and nurse burnout in acute care general hospitals. Issues Ment Health Nurs 2010 Mar 09;31(3):198-207 [FREE Full text] [doi: 10.3109/01612840903200068] [Medline: 20144031]
41. Verhaeghe M, Bracke P, Bruynooghe K. Stigmatization in different mental health services: a comparison of psychiatric and general hospitals. J Behav Health Serv Res 2007 Apr 01;34(2):186-197. [doi: 10.1007/s11414-007-9056-4] [Medline: 17333402]
42. Takian A, Sheikh A, Barber N. We are bitter, but we are better off: case study of the implementation of an electronic health record system into a mental health hospital in England. BMC Health Serv Res 2012 Dec 31;12(1):484 [FREE Full text] [doi: 10.1186/1472-6963-12-484] [Medline: 23272770]
43. Williams SC, Schmaltz SP, Castro GM, Baker DW. Incidence and method of suicide in hospitals in the United States. Jt Comm J Qual Patient Saf 2018 Nov;44(11):643-650. [doi: 10.1016/j.jcjq.2018.08.002] [Medline: 30190221]
44. Brodsky BS, Spruch-Feiner A, Stanley B. The Zero Suicide model: applying evidence-based suicide prevention practices to clinical care. Front Psychiatry 2018 Feb 23;9:33 [FREE Full text] [doi: 10.3389/fpsyt.2018.00033] [Medline: 29527178]
45. Panchal N, Kamal R, Cox C, Garfield R. The implications of COVID-19 for mental health and substance use. Kaiser Family Foundation. 2021. URL: https://tinyurl.com/3jubn8hs [accessed 2023-03-13]

46. Taquet M, Luciano S, Geddes JR, Harrison PJ. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. Lancet Psychiatry 2021 Feb;8(2):130-140 [FREE Full text] [doi: 10.1016/S2215-0366(20)30462-4] [Medline: 33181098]

Abbreviations

ACTS: Accelerated Creation-to-Sustainment
IMHP: inpatient mental health professional
SMI: severe mental illness
TIME: Technology Implementation for Mental-Health End-Users

Edited by J Torous; submitted 30.06.22; peer-reviewed by A Kendall, S Cross, S Chen; comments to author 05.11.22; revised version received 16.01.23; accepted 26.01.23; published 06.04.23.

Please cite as:
Westheimer JL, Moukaddam N, Lindsay JA, Sabharwal A, Najafi B, Iacobelli PA, Boland RJ, Patriquin MA
Technology Implementation for Mental Health End Users: A Model to Guide Digital Transformation for Inpatient Mental Health Professionals
JMIR Ment Health 2023;10:e40429
URL: https://mental.jmir.org/2023/1/e40429
doi:10.2196/40429
PMID:37023415

Viewpoint

Assessing Diversity and Inclusivity is the Next Frontier in Mental Health Recovery Narrative Research and Practice

Yasuhiro Kotera[1], PhD; Stefan Rennick-Egglestone[1], PhD; Fiona Ng[1], PhD; Joy Llewellyn-Beardsley[1], MA; Yasmin Ali[1], MSc; Chris Newby[2], PhD; Caroline Fox[1], PhD; Emily Slade[3]; Simon Bradstreet[4], PhD; Julian Harrison[5], MA; Donna Franklin[5], BSc; Olamide Todowede[1], PhD; Mike Slade[1,6], PhD

[1]School of Health Sciences, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[2]School of Medicine, Queen's Medical Centre, University of Nottingham, Nottingham, United Kingdom

[3]Department of Computer Science, University of Oxford, Oxford, United Kingdom

[4]School of Health & Wellbeing, University of Glasgow, Glasgow, United Kingdom

[5]Narrative Experiences Online Lived Experience Advisory Panel, Nottingham, United Kingdom

[6]Faculty of Nursing and Health Sciences, Health and Community Participation Division, Nord University, Namsos, Norway

Corresponding Author:
Yasuhiro Kotera, PhD
School of Health Sciences
Institute of Mental Health
University of Nottingham
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 01158231294
Email: yasuhiro.kotera@nottingham.ac.uk

Abstract

Demand for digital health interventions is increasing in many countries. The use of recorded mental health recovery narratives in digital health interventions is becoming more widespread in clinical practice. Mental health recovery narratives are first-person lived experience accounts of recovery from mental health problems, including struggles and successes over time. Helpful impacts of recorded mental health recovery narratives include connectedness with the narrative and validation of experiences. Possible harms include feeling disconnected and excluded from others. Diverse narrative collections from many types of narrators and describing multiple ways to recover are important to maximize the opportunity for service users to benefit through connection and to minimize the likelihood of harm. Mental health clinicians need to know whether narrative collections are sufficiently diverse to recommend to service users. However, no method exists for assessing the diversity and inclusivity of existing or new narrative collections. We argue that assessing diversity and inclusivity is the next frontier in mental health recovery narrative research and practice. This is important, but methodologically and ethically complex. In this viewpoint, we propose and evaluate one diversity and two inclusivity assessment methods. The diversity assessment method involves use of the Simpson Diversity Index. The two inclusivity assessment methods are based on comparator demographic rates and arbitrary thresholds, respectively. These methods were applied to four narrative collections as a case study. Refinements are needed regarding a narrative assessment tool in terms of its practicality and cultural adaptation.

(JMIR Ment Health 2023;10:e44601) doi:10.2196/44601

KEYWORDS

recovery narrative; web-based mental health interventions; inclusivity; diversity; collective action; curation; mental health; digital health; telemedicine; clinical practice; narrative research; demographic

Introduction

Background

Demand for digital mental health interventions (DMHIs) has been increasing in many countries [1]. For example, 76% of the Australian general public indicated an interest in using DMHIs [2], and US $2.4 billion were invested in these interventions in 2020 alone in the United States [3]. A Singaporean mental health app, Intellect, is now used by over 3 million people in 20 countries, mostly in Asia [4]. A large-scale education program about ethics in DMHIs has been launched [5], and more active use of DMHIs is expected in Africa [6]. The advantages of DMHIs include accessibility, cost-effectiveness, and personalization [7], which can address key barriers for mental health recovery, such as low help-seeking and the stigma associated with mental health problems [8]. Effectiveness of DMHIs has been generally reported in diverse populations (eg, children, young people, older adults, university students, health care workers, people with neurodevelopmental disabilities) and on diverse mental health experiences [2,9-12]. DMHIs have become an important domain in medical internet research [13-15].

Mental health recorded recovery narratives (RRNs) have been used in DMHIs. Mental health recovery narratives can be defined as "first-person lived experience accounts of recovery from mental health problems, which refer to events or actions over a period of time, and which include elements of both adversity/struggle and of self-defined strengths/successes/survival" [16]. Although narratives can be shared in person, such as when a peer support worker tells their story [17], RRNs are those presented in recorded formats such as written text, audio, and video. RRNs are often disseminated online [18] and have been used in a range of clinical and societal interventions [19]. For example, the Narrative Story Bank was created by the Scottish Recovery Network to inspire hope and offer tools and technologies for recovery [20]. Content from the Narrative Story Bank helped to inform Scotland's mental health strategy [21]. In Scotland, the use of recovery narratives has been incorporated into clinical practice, such as the production of written RRNs to support self-reflection [21]. Similarly, RRNs have been used in clinical training to enhance communication skills and empathy of health care workers [22]. RRNs are used in national antistigma campaigns [23], which aim to reduce stigma associated with mental health problems [24].

The impact of RRNs on recipients has been evaluated. For example, students who listened to recovery narratives of people with anorexia nervosa showed more understanding of the mental illness and a less stigmatized view toward these individuals [25]. Eating disorder memoirs offered validation of personal experience and created a positive affective response among people with eating difficulties [26]. Helpful outcomes of accessing both live and recorded recovery narratives include connection with others, validation of own experiences, empowerment, hopefulness, gratitude, and stigma reduction [27]. Mechanisms of impact from RRNs include comparison with the narrative and narrator, learning about the experiences of others, and feeling empathy, all of which create a sense of connection [28]. Accessing RRNs can thus support personal recovery by increasing hope, meaning, and a sense of fulfillment [18,29].

RRNs are often grouped and presented as collections [19], which have been provided on bespoke websites [30] and in books composed of recovery narratives linked by a theme such as psychosis [31] or men's eating disorders [32]. The people who assemble RRN collections do so for a range of reasons, including providing education about mental health and recovery, supporting others in their recovery journey, and campaigning for change in health service provision [33].

Clinicians who use RRNs as part of treatment need to know the possible impact of the RRN collections they recommend to their service users [34]. Given their broad range of clinical applications, the ability to characterize a narrative collection is important because not all RRNs are helpful to all people. An interview study reported that recovery narratives can be harmful if recipients feel disconnected from the narrative, resulting in distancing themselves from the narrator [27]. For instance, if a "right" way to recover is communicated by the RRNs, this implicit message can cause a recipient to feel disconnected [35,36] if that particular pathway is in some way "outside" that person's scope, experience, and perspective. A feeling of being disconnected is detrimental because connection is the central mechanism supporting recovery after access to RRNs [28,37].

Diversity and Inclusivity

Providing diversity in RRN collections may be one approach to maximizing benefit and minimizing harms, enabling a sense of connection to people with different backgrounds and experiences. A diverse set of narratives will increase the opportunity for the recipient to feel connected to a narrative [27]. Further, a lack of diversity in RRN collections may give rise to stereotyping [38], leaving recipients feeling excluded and disconnected [18], or suggest that there is only one way or a "right" way to recover [35,39], which may not fit with their own beliefs, experiences, and needs. Recipient characteristics and narrative characteristics moderate the impact of a narrative on a sense of connection [28]. For example, people in ethnic minority groups felt a lower level of connection when an RRN collection marginalized these groups [40]. Diverse narratives are needed to ensure that recipients from different backgrounds and with different identities have the greatest opportunity to feel connected with both a narrative and a narrator, as both types of connection are known to maximize the positive effects of RRNs [38]. One exception might be collections of RRNs targeted at specific groups experiencing structural inequalities or marginalization (eg, refugees), where the selection of narrators who belong to those groups might be an appropriate strategy for maximizing beneficial impact.

Two types of variation in a narrative collection can be differentiated: diversity and inclusivity [41]. *Diversity* is defined as heterogeneity in narrative characteristics, such as a spread of narrator demographics and "protected characteristic" identities, as defined by the Equality Act 2010 (eg, age, gender, sexual orientation), or types of narrative content (eg, trajectory,

genre). *Inclusivity* is defined as representativeness of narrative characteristics in relation to a broader population [42].

Diversity and inclusivity concepts can be applied to a specific characteristic of an RRN collection, such as narrator age or a narrative focused on a certain demographic characteristic (eg, race, disability). Diversity is present when there are a meaningful number of narratives within each subcategory [43]. For example, diversity in narrator age is present when there is a spread of younger, middle-aged, and older narrators. Since diversity is a feature of the collection, a diversity metric will be a constant for a given collection. By contrast, inclusivity is present when the proportion of narratives within each subcategory is similar to the proportion in a wider comparison population at a given time [44]. For example, inclusivity in narrator age is present when there are similar proportions of each narrator age group in the collection and in a comparison population, such as other people on the caseload of the mental health service or in the general population. Since inclusivity is a function of the collection when used in a specific context, an inclusivity metric will vary based on context.

We argue that assessing diversity and inclusivity of narrative collections is the next frontier in mental health recovery narrative research and practice. A tool to characterize individual RRNs has been developed. The Inventory of Characteristics of Recovery Stories (INCRESE) is a standardized 77-item instrument characterizing narrative mode, narrator and narrative characteristics, content warnings, turning points, and narrative content [45]. Although INCRESE is used to characterize individual RRNs, no method currently exists to assess the diversity and inclusivity of an RRN collection.

We here discuss aspects of measuring diversity and inclusivity in recovery narrative collections. The Narrative Experiences Online (NEON) study used INCRESE to characterize a large collection of 687 recovery narratives. Because we had a large data set of narrative characteristics measured using INCRESE available, we used this data set for the present analysis.

Ethics Approval

The NEON study received approval from a UK National Health Service Research Ethics Committee (West London and GTAC, 18/LO/0991).

Identifying the Relevant Characteristics

To develop diversity and inclusivity metrics, characteristics for the evaluation of diversity and inclusivity need to be identified. To establish a theoretical and cross-culturally valid understanding of important diversity and inclusivity characteristics, we analyzed policy and research to identify characteristics that are internationally agreed as requiring protection from discrimination. Three data sources were used. First, national policy documents relating to equality, diversity, and inclusivity were reviewed to identify characteristics protected by law in each country. Policy documents were collated from a purposive sample of 20 predefined countries shown in Table 1, chosen for variation in (1) region; (2) income level as classified by the World Bank; and (3) status as a Western, Educated, Individualized, Rich and Democratic (WEIRD) versus non-WEIRD country [46] (see Multimedia Appendix 1 for income levels and WEIRD/non-WEIRD status). Policy documents were retrieved using Google searches with the terms “antidiscrimination [country]” and “human rights [country].” At least one source of information was identified for each country. Where the governmental information was not available in English (eg, Iran, Yemen, Morocco), online sources such as information websites (eg, The Academic Network of European Disability Experts, Human Rights Watch, European Commission, International Labor Organization) or reports (eg, Human Rights Committee Report, Human Rights Watch Report) written in English about protected characteristics in the country were reviewed. The identified characteristics were grouped, and Table 1 shows the frequency across the 20 countries of the 13 identified characteristics protected by law and policy.

Some characteristics collapse complex and contested components, where terms are used inconsistent internationally. For example, “Sex and gender” refers to both biological sex assigned at birth and the social construct gender, with subcategories including “female,” “male,” and “nonbinary” [47]. Therefore, the theme is categorized as “Sex and gender” [48].

The five characteristics with the highest international consensus are Sex and gender (eg, assigned sex at birth, socially constructed gender, female/male/nonbinary), Beliefs (eg, political, religious, philosophical), Origin (eg, race, ethnicity), Family (eg, marital status, carer responsibilities), and Disability (mental, physical, learning, and sensory).

To maximize cross-cultural validity, four multinational documents were reviewed, comprising two international human right treaties (Universal Declaration of Human Rights, Convention on the Rights of Persons with Disabilities) and two relevant systematic reviews about diversity and inclusivity [49,50]. The presence of each of the 13 identified characteristics in these four documents was tabulated to identify the most widely agreed characteristics relevant to diversity and inclusivity (Table 2). In both tables, the authors YK and FN independently reviewed the documents and discussed the rating until consensus was reached, which was then confirmed by the other authors.

Table 1. Characteristics (N=13) protected by law and policy in 20 countries.

Country	Sex and gender	Beliefs[a]	Origin[b]	Family[c]	Disability	Sexuality	Age	Economics[d]	Employment	Pregnancy	Education	Language	Military veteran
Australia	✓		✓		✓	✓	✓		✓				
Brazil	✓	✓	✓	✓	✓	✓	✓		✓	✓			
Cuba	✓	✓											
Greece	✓	✓	✓	✓					✓				
Guyana	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Iran	✓	✓	✓			✓						✓	
Ireland	✓	✓	✓	✓	✓	✓	✓	✓					
Italy	✓	✓	✓					✓				✓	
Japan	✓	✓	✓	✓				✓			✓		
Libya	✓	✓	✓										
Morocco	✓	✓		✓									
Netherlands	✓	✓	✓	✓	✓	✓	✓	✓	✓				
Norway	✓	✓	✓	✓	✓	✓	✓			✓			
Palestine	✓	✓									✓		
Spain	✓	✓	✓		✓								
Suriname	✓	✓	✓					✓			✓	✓	
Tunisia	✓		✓		✓			✓					
United Kingdom	✓	✓	✓	✓	✓	✓	✓			✓			
United States	✓	✓	✓	✓	✓	✓	✓			✓			✓
Yemen	✓	✓	✓	✓				✓					
Count	20	18	17	11	10	8	8	8	5	5	3	3	1

[a]Includes religious, political, and philosophical beliefs.

[b]Includes race, ethnicity, migration.

[c]Includes marriage status and carer responsibilities.

[d]Includes social class.

Table 2. Candidate diversity and inclusivity characteristics mapped against international treaties and systematic reviews.

Characteristic	UDHR[a] article	CRPD[b] article	Yadav and Lenka [49]	Manoharan and Singal [50]
Sex and gender	✓	✓		✓
Beliefs	✓	✓		
Origin	✓		✓	✓
Family	✓	✓		✓
Disability		✓		
Sexuality			✓	
Age		✓	✓	✓
Economics	✓	✓		
Employment	✓	✓		
Pregnancy				
Education	✓	✓		
Language	✓			✓
Military veteran				

[a]UDHR: Universal Declaration of Human Rights.

[b]CRPD: Convention on the Rights of Persons with Disabilities.

All characteristics apart from "Pregnancy" were identified by at least one of the four sources. Finally, to maximize relevance to mental health recovery narratives, a systematic review making recommendations for best practice in curating mental health lived experience narrative collections was assessed [38]. "Positioning" was added as a mental health narrative–specific characteristic to assess if a collection includes both positive and negative narratives about mental health services to capture whether a broad range of perspectives are included [38]. In total, these 14 characteristics were identified as relevant to the diversity and inclusivity of RRN collections.

Mapping Against a Narrative Characterization Tool

To enable an assessment of the appropriateness of INCRESE in assessing diversity and inclusivity, the 77 INCRESE items were mapped against the 14 diversity and inclusivity characteristics (Table 3).

Twenty-two INCRESE items were able to be mapped against the diversity and inclusivity characteristics. No INCRESE items were identified relevant to the characteristics of Language and Military veteran. Twelve characteristics, including all five of the most supported characteristics and Positioning, can be measured using the INCRESE items. Our INCRESE database enabled a preliminary investigation of diversity and inclusivity, despite INCRESE not being a perfect tool for assessing these metrics. The approach may be refined in the future by including items regarding language and military status.

Table 3. Inventory of Characteristics of Recovery Stories (INCRESE) items mapped against the diversity and inclusivity characteristics.

Characteristic	Corresponding INCRESE items
Sex and gender	11 Gender
Beliefs	71 Activism, 72 Spiritual/religious activities
Origin	13 Ethnicity, 15 Location
Family	49 Family, 53 Relationships, 74 Caring responsibilities, 75 Family experiences of mental health issues
Disability	17 Visual difficulties, 18 Hearing difficulties, 19 Mobility/stamina difficulties, 20 Cognitive difficulties, 21 Self-care difficulties
Sexuality	16 Sexuality
Age	12 Age
Economics	54 Income, 55 Housing
Employment	56 Work
Pregnancy	48 Pregnancy/birth
Education	51 Education
Language	Not applicable
Military veteran	Not applicable
Positioning	32 Positioning

Quantifying Each Characteristic

To quantify the diversity and inclusivity of RRN collections, each characteristic needs to be assessed using its subcategories (eg, for the Sex and gender characteristic, the INCRESE characterization choices of "Male," "Female," and "Other" may be the subcategories). Measuring diversity involves characterizing the spread of narratives across each subcategory. For example, an RRN collection that includes no narrator categorized as "Other" in the Sex and gender characteristic is less diverse than an RRN collection that does include such narrators. By contrast, measuring inclusivity involves establishing the same two parameters of characteristics and subcategories, and additionally identifying the comparison population. For example, presence of an Origin subcategory of "white" may not increase an inclusivity metric in the UK general population as much as it does in many other populations.

Measurement of Diversity and Inclusivity

Overview

We present one option for measuring diversity and two options for measuring inclusivity for recovery narrative collections. The two options for measuring inclusivity have different properties, enabling people assessing inclusivity to make a choice over which to use.

Measuring Diversity: Simpson Diversity Index

The Simpson Diversity Index (SDI) is an established index used in the natural sciences to assess biodiversity [43]. The SDI considers the number of species present and the abundance in each species to indicate the variance in species. The SDI is calculated by deducting the Simpson Index (SI) from 1, where $SI=\Sigma n(n-1)/N(N-1)$. When used to assess narrative collection diversity in relation to a particular characteristic (eg, "narrator gender"), n refers to the total number of narratives within each option of the characteristic (eg, "female narrator") and N refers to the total number of narratives across all options. See Multimedia Appendix 2 for example calculations. The SDI ranges from 0 (low diversity) to 1 (high diversity).

Measuring Inclusivity Option 1: Demographic Rates as Comparison Population

Inclusivity captures the extent to which minority groups in a comparison population (eg, a country population, a service user cohort at one mental health service) are included in a collection [51]. One approach to assess inclusivity is to identify the categories that are minoritized in the comparison population and compare their proportion in the narrative collection. For example, the Origin characteristic is measured by INCRESE item 13 "Ethnicity," with categories of "Not identifiable," "Asian," "Black/African/Caribbean," "Dual/multiple ethnic group," "Other ethnic group," and "white." When used in the United Kingdom, all choices apart from "white" are minority groups (a limitation of this categorization is that some "white" communities such as the Traveler, Gypsy, and Roma communities are also very socially excluded). In the United Kingdom, 13% of the population are nonwhite [52]. In a collection, among all narratives, if the ratio of nonwhite narratives is higher than 13%, the collection can be considered as inclusive with respect to ethnicity in the United Kingdom. If inclusivity is being assessed in a different comparison population such as a different country, then alternative choices for minority categories would be made.

Measuring Inclusivity Option 2: Arbitrary Threshold as Benchmark

A second approach to measuring inclusivity is to set an arbitrary benchmark. One approach is to decide that five narratives are sufficient to satisfy a benchmark for a certain characteristic. A stronger approach, as often used for external examination in the university sector [53,54], is to decide both a minimum rate (eg, 10%) and number (eg, 5), and choose whichever is greater.

Case Study

Data Set, Analysis, and Outcomes

The three methods above were applied to the NEON Collection as a case study. Four characteristics matching INCRESE items were considered: Sex and gender, Origin, Disability, and Positioning. The NEON Collection is a curated collection of mental health RRNs. All narratives are included in the NEON Collection with permission [55]. Each narrative is characterized using INCRESE by multiple raters [45].

The candidate approaches were applied to four groups: the entire NEON Collection, two of the larger collections chosen for difference in source, and the individual donations contained in the NEON Collection. In September 2022, the NEON Collection comprised 687 narratives compiled from 34 public collections and from individual donations. One of the two larger collections, which we here refer to as "statutory service" (78 narratives), was compiled by a statutory mental health service. The other, which we refer to as "ethnic minority book" (19 narratives), was published as a book focusing on the mental health of ethnic minority groups. Individual donations (n=29) comprised narratives collected directly from individuals as donations to the NEON Collection.

To evaluate diversity using the SDI, because the Origin and Disability characteristics consist of multiple INCRESE items, the mean SDI scores were calculated. The SDI scores for each collection group are shown in Table 4.

Table 4. Diversity scores (Simpson Diversity Index) for Sex and gender, Positioning, Origin, and Disability characteristics.

Characteristics and corresponding INCRESE[a] items	NEON[b] collection	Statutory service	Ethnic minority book	Donations
Sex and gender (INCRESE=Gender)	0.58	0.60	0.29	0.64
Positioning	0.61	0.56	0.56	0.53
Origin				
Location	0.70	0.50	0.20	0.57
Ethnicity	0.58	0.07	0.61	0.46
Mean[c]	0.64	0.37	0.51	0.52
Disability				
Visual	0.01	0.00	0.00	0.00
Hearing	0.01	0.00	0.00	0.00
Mobility	0.01	0.03	0.00	0.00
Cognitive	0.01	0.07	0.00	0.07
Self-care	0.05	0.05	0.11	0.00
Mean[c]	0.04	0.03	0.02	0.01

[a]INCREASE: Inventory of Characteristics of Recovery Stories.

[b]NEON: Narrative Experiences Online.

[c]Multiple INCRESE items are attached to one characteristic.

Of the four groups assessed, the NEON Collection is the most diverse in terms of Positioning, Origin, and Disability. Individual donations are the most diverse in terms of Sex and gender. Statutory service is the most diverse in the Disability subcategory of cognitive difficulties. Ethnic minority book is the most diverse in the Original subcategory ethnicity.

To evaluate inclusivity using option 1 (demographic rates), the comparison population used was the UK general population. The Positioning characteristic was excluded as there are no demographic data of narrative positioning available. Minority groups in each characteristic were identified and the proportion of narratives from each minority group was calculated. The proportions of minority groups in the comparison UK general population were obtained for gender [56], ethnicity [52], and each Disability component [57-61]. For location, the number of British nationals living outside Europe was identified and then the proportion against the UK population was calculated [62]. For self-care, the prevalence of self-neglect was identified [61]. The findings are shown in Table 5.

Table 5. Inclusivity option 1: minority group proportions compared to the UK general population.

Characteristic and INCRESE[a] item	Response categories		Proportion of minority narratives (%)				Minority proportion in UK general population (%)
	Nonminority	Minority	NEON[b] Collection	Statutory service	Ethnic minority book	Donations	
Sex and gender (INCREASE=Gender)	Not identifiable, Female, Male	Other	1	0	0	0	3
Origin							
Ethnicity	Not identifiable, white	Asian, Black/African/Caribbean, Dual/multiple ethnic group, Other ethnic group	10	0	*47* [c]	0	13
Location	Europe	Africa, Asia, Australasia, North America, South America	*36*	0	3	3	7
Disability							
Visual difficulties	Not identified	Yes	0.4	0	0	0	3
Hearing difficulties	Not identified	Yes	0.3	0	0	0	17
Mobility	Not identified	Yes	3	1	0	0	46
Cognitive difficulties	Not identified	Yes	*4*	*4*	0	*3*	2
Self-care	Not identified	Yes	*2*	*3*	*5*	0	0.2

[a]INCRESE: Inventory of Characteristics of Recovery Stories.

[b]NEON: Narrative Experiences Online.

[c]Numbers in italics indicate proportions above those of the UK general population.

The NEON Collection met the inclusivity benchmark of being above the UK general population for the location subcategory in the Origin characteristic and for two subcategories in the Disability characteristic: cognitive difficulties and self-care. Likewise, statutory service met the inclusivity benchmarks for cognitive difficulties and self-care. Ethnic minority book met the inclusivity benchmarks for ethnicity in the Origin characteristic and self-care in the Disability characteristic. Individual donations met the inclusivity benchmark for cognitive difficulties. No collections met the inclusivity benchmarks for the Sex and gender characteristic or the Disability subcategories of visual difficulties, hearing difficulties, and mobility. The largest inclusivity score was for ethnicity in the ethnic minority book, followed by location in the NEON Collection.

To evaluate inclusivity using option 2 (arbitrary threshold), benchmarks were set at a minimum proportion of 10% of the number of narratives in a collection and a minimum number of 5 narratives. Whichever was the higher number was used as the threshold, as shown in Table 6.

The NEON Collection met the inclusivity benchmarks for both ethnicity and location in the Origin characteristic and for the Positioning characteristic. Ethnic minority book met the inclusivity benchmark for ethnicity in the Origin characteristic. Neither statutory service nor individual donations met any of the characteristics or subcategories. Similar to inclusivity option 1, the scores for location in the NEON Collection and ethnicity in the ethnic minority book markedly exceeded the benchmarks.

Table 6. Inclusivity option 2: arbitrary thresholds.

Characteristics and corresponding INCRESE[a] items	NEON[b] Collection (n=687); threshold: n=69	Statutory service (n=78); threshold: n=8	Ethnic minority book (n=19); threshold: n=5	Individual donations (n=29); threshold: n=5
Sex and gender (INCRESE=Gender)	4	0	0	0
Origin				
Ethnicity	*69* [c]	0	*9*	1
Location	*245*	0	0	1
Disability				
Visual	3	0	0	0
Hearing	2	0	0	0
Mobility	18	1	0	0
Cognitive	26	3	0	1
Self-care	17	2	1	0
Positioning	*97*	5	4	2

[a]INCRESE: Inventory of Characteristics of Recovery Stories.

[b]NEON: Narrative Experiences Online.

[c]Numbers in italics indicate proportions above the threshold.

Strengths and Limitations of Each Approach

Simpson Diversity Index

A strength of calculating the diversity scores using the SDI is its practicality; to calculate the SDI, only the frequencies for each characteristic are required [63]. Less practical measures exist. For example, the Shannon Diversity Index is another established biodiversity index, which requires more data such as the rate of each species present out of the total organism population [64]. Likewise, Social Choice Methods were proposed in computer science; however, these methods require more data than required for the SDI (eg, social structures of power and influence) [65].

Three major weaknesses need to be noted. One is that there are no interpretation scores to indicate a level of the diversity in the SDI (eg, high, medium, and low). Another limitation is reliance on INCRESE data, which do not map onto all protected characteristics identified. In particular, there are no INCRESE items for Language and Military status. Moreover, even where an INCRESE item and a characteristic have the same label, the meaning may be different. For example, in this analysis, we chose an INCRESE item for location as part of the Origin characteristic. However, the response choices of this item are placed at a global level (eg, Europe, Asia, Africa), whereas the international treaties and national policies often regard the location of where people are from or live at a local level (eg, housing discrimination in the United Kingdom [66] and "buraku" [roughly defined as a defiled area] in Japan [67]). Both items and response choices in INCRESE can be extended to enable calculation of the diversity from the INCRESE data set. Lastly, while appealing as a simple and comprehensible metric, the SDI may misleadingly simplify the complex issue of diversity. For example, the diversity scores should not be treated as a target, which can be deprioritized once hit [68]. The diversity scores should rather be used as part of helping recipients from different backgrounds and with different identities feel connected with both a narrative and a narrator, maximizing the positive effects of RRNs [38].

Inclusivity Option 1

A strength of inclusivity option 1 is that it is a logical approach, comparing the proportion between a collection and its comparable population. This method allows a direct comparison with different contexts as far as the demographic data are available, and tailors the assignment of minoritized status to categories matching the comparison population.

Weaknesses include the time required, modest collection sizes, and comparator choice. First, inclusivity option 1 requires more time than option 2, because of the difficulties with finding the comparable data. There would be a health service resource allocation implication of choosing option 1 rather than option 2. Busy practitioners may not have time to identify comparable demographic data. For example, for self-care, finding the demographic proportion of people with self-care difficulties required a great amount of time. We used the demographic proportion of people suffering from self-neglect. Self-care difficulties and self-neglect may be similar; however, self-neglect can indicate a wider set of behaviors than self-care, such as hoarding and unwillingness to receive support [69,70]. Second, RRN collections often do not have many narratives (eg, the largest collection in the NEON Collection includes 78 narratives). Only a few narratives can meet the benchmarks in small collections. For example, in the ethnic minority book (n=19), only one narrative is sufficient to meet the benchmarks for the subcategories of gender, and visual, cognitive, and self-care difficulties, despite each having different demographic proportions. Lastly, a decision needs to be made on what an appropriate comparator is for this method. We used the UK general population; however, if a statutory mental health service

uses this method, the entire cohort of their service users may be more meaningful as a comparative tool. Identifying a meaningful comparator and retrieving comparison information may be complex [71].

Inclusivity Option 2

Strengths of inclusivity option 2, based on the arbitrary threshold, include practicality and representation. This approach allows a reasonable minimum number to be present, addressing the need for individuals from minority communities to "see themselves" [72] in the narrative collection.

However, weaknesses include the difficulty in justifying the benchmark numbers. Relatedly, the approach produces a binary outcome: the collection is either inclusive in a particular characteristic or it is not. This does not differentiate between a collection that just meets the benchmark versus one that markedly exceeds it.

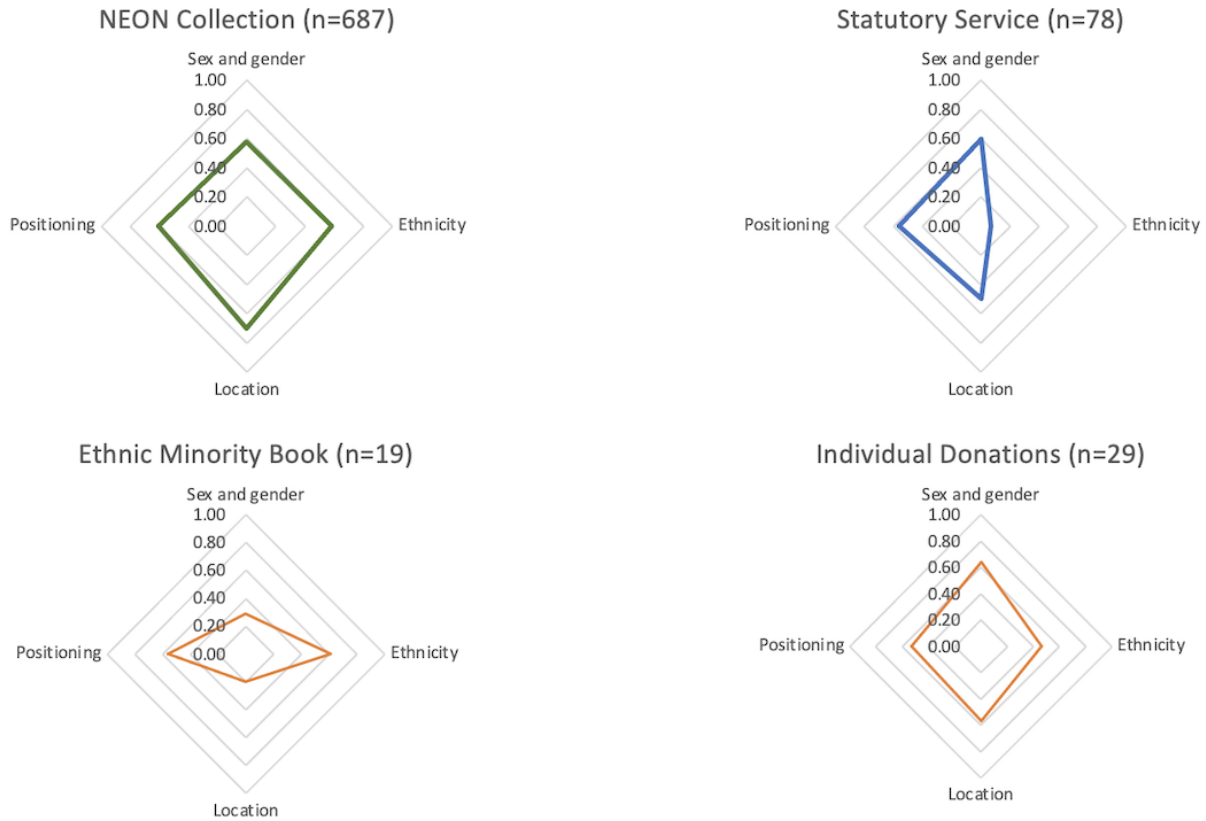
Overall, the diversity and inclusivity of the NEON Collection are higher than those of the three subgroups. One explanatory attribute is its size, as the three subgroups are part of the NEON Collection. Because the size is large, the NEON Collection has an inherent advantage with respect to diversity [73]. The more narratives a collection has, the more likely the collection will have different types of narratives. The size advantage also relates to the high inclusivity of the NEON Collection. The NEON Collection marked low yet above-benchmark scores in the characteristics where the benchmarks were low, whereas the other three collections did not (eg, scored a 0 indicating no relevant narrative identified).

Our case study has three implications. First, the diversity and inclusivity can be measured, although each metric has limits on its meaningfulness. For example, the diversity metric has no interpretation scores to indicate a level of the diversity. Inclusivity option 1 highlights a mismatch between INCRESE items and characteristics, and raises the question about defining an appropriate comparator. Inclusivity option 2 requires justification for the benchmark. Second, larger collections in general are more diverse and inclusive. Third, collections that have a specific focus (eg, on ethnicity) can be differentiated using these metrics.

Finally, how the output is presented needs to be discussed. A challenge with all measurement approaches is how to present the results [74]. One approach is a radar chart, used in the Four Layers of Diversity Model [75]. For example, the Sex and gender, Origin (ethnicity and location), and Positioning characteristics can be presented as shown in Figure 1. The items in the Disability characteristic were excluded as all five items included a 0, which is already visible in the table format.

One advantage of the radar chart is that the uniqueness of each collection can be visually highlighted [76]. Moreover, compared to a table format, chart formats such as a radar chart are often more reader-friendly and inclusive (eg, for people with dyslexia) [77,78]. A disadvantage is that not many characteristics can be included to maintain a reader-friendly presentation [76].

Figure 1. Radar charts to present the diversity scores for the Sex and gender, Origin (Ethnicity and Location), and Positioning characteristics in four collections. The radar chart presentation can visually highlight the uniqueness of each collection, and can be more reader-friendly and inclusive, but cannot present many characteristics. NEON: Narrative Experiences Online.



Conclusion

Several knowledge gaps exist. First, there is a mismatch between the INCRESE items and the identified characteristics. Candidate new items for INCRESE are Language and Military veteran. Moreover, there are currently only three characterization choices for Sex and gender in INCRESE ("male," "female," and "other"). More diverse and inclusive choices are needed (eg, "transgender," "nonbinary/nonconforming"). The content of INCRESE might be reviewed to enhance its use in diversity and inclusivity metrics. Second, the assessment of inclusivity requires development of a reliable and context-sensitive approach to identifying minority groups in each characteristic. Third, the optimal approach to inform clinical practice needs to be identified. In addition to concerns about comprehensibility of these candidate metrics, there may be specific clinical priorities; for example, some clinicians may not want to recommend a collection that includes many narratives about poor service experiences to their service users. Lastly, cultural adaptation of these metrics needs to be considered. For example, the Military veteran characteristic may hold more cultural importance in the United States than in many other countries [79]. In cultures such as Japan and South Korea, where age plays an important role [80], the Age characteristic may be more relevant. The next stages of research will include refinement of each metric with attention paid to minimizing the burden of calculation and developing interpretation guidance, the involvement of key stakeholders (ie, people living with mental health issues and mental health clinicians) in arbitrating between the candidate approaches, and real-world evaluation of the impact of more diverse and inclusive RRN collections.

Acknowledgments

This paper represents the outcome of independent research funded by the National Institute for Health and Care Research (NIHR) under its Programme Grants for Applied Research Programme (Programme Grants for Applied Research, Personal Experience as a Recovery Resource in Psychosis: NEON Programme, RP-PG-0615-20016). MS acknowledges the support of the NIHR Nottingham Biomedical Research Centre. FN acknowledges funding from an NIHR Advanced Fellowship (NIHR302218) and a University of Nottingham Anne McLaren Research Fellowship. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Authors' Contributions

Conceptualization: YK, SB, and MS; Methodology: YK, FN, and MS; Investigation: all authors; Resources: SRE and MS; Data curation: YK, SRE, FN, JLB, and CF; Writing—original draft preparation: YK and MS; Writing—review and editing: all authors; Project administration: YK and MS.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Income level and Western, Educated, Individualized, Rich and Democratic (WEIRD)/non-WEIRD status of 20 countries.
[DOCX File , 18 KB - mental_v10i1e44601_app1.docx]

Multimedia Appendix 2
Example calculations of the Simpson Diversity Index.
[DOCX File , 19 KB - mental_v10i1e44601_app2.docx]

References

1. Philippe TJ, Sikder N, Jackson A, Koblanski ME, Liow E, Pilarinos A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022 May 12;9(5):e35159 [FREE Full text] [doi: 10.2196/35159] [Medline: 35551058]
2. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
3. Weir K. The age of digital interventions. American Psychological Association Monitor on Psychology. 2021 Oct 01. URL: https://www.apa.org/monitor/2021/10/news-digital-interventions [accessed 2022-08-24]
4. Intellect launches digital mental health services in Japan. Intellect. 2022. URL: https://intellect.co/read/intellect-launches-digital-mental-health-services-in-japan/ [accessed 2022-08-24]
5. Ethics in Mental Health Digital Innovations for Young People in Africa: Digital Campaign. University of Oxford Department of Psychiatry News. URL: https://www.psych.ox.ac.uk/news/ethics-in-mental-health-digital-innovations-for-young-people-in-africa-digital-campaign [accessed 2022-08-24]

6. Adepoju P. Africa turns to telemedicine to close mental health gap. Lancet Digit Health 2020 Nov;2(11):e571-e572 [FREE Full text] [doi: 10.1016/S2589-7500(20)30252-1] [Medline: 33103096]
7. Balcombe L, De Leo D. Digital mental health challenges and the horizon ahead for solutions. JMIR Ment Health 2021 Mar 29;8(3):e26811 [FREE Full text] [doi: 10.2196/26811] [Medline: 33779570]
8. Kotera Y, Gilbert P, Asano K, Ishimura I, Sheffield D. Self‐criticism and self‐reassurance as mediators between mental health attitudes and symptoms: Attitudes toward mental health problems in Japanese workers. Asian J Soc Psychol 2018 Dec 13;22(2):183-192. [doi: 10.1111/ajsp.12355]
9. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
10. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
11. Riadi I, Kervin L, Dhillon S, Teo K, Churchill R, Card KG, et al. Digital interventions for depression and anxiety in older adults: a systematic review of randomised controlled trials. Lancet Healthy Longev 2022 Aug;3(8):e558-e571 [FREE Full text] [doi: 10.1016/S2666-7568(22)00121-0] [Medline: 36102765]
12. Khan K, Hall CL, Davies EB, Hollis C, Glazebrook C. The effectiveness of web-based interventions delivered to children and young people with neurodevelopmental disorders: systematic review and meta-analysis. J Med Internet Res 2019 Nov 01;21(11):e13478 [FREE Full text] [doi: 10.2196/13478] [Medline: 31682573]
13. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
14. Bergin AD, Vallejos EP, Davies EB, Daley D, Ford T, Harold G, et al. Preventive digital mental health interventions for children and young people: a review of the design and reporting of research. NPJ Digit Med 2020 Oct 15;3(1):133. [doi: 10.1038/s41746-020-00339-7] [Medline: 33083568]
15. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
16. Llewellyn-Beardsley J, Rennick-Egglestone S, Callard F, Crawford P, Farkas M, Hui A, et al. Characteristics of mental health recovery narratives: systematic review and narrative synthesis. PLoS One 2019 Mar 28;14(3):e0214678 [FREE Full text] [doi: 10.1371/journal.pone.0214678] [Medline: 30921432]
17. Borderline personality disorder (BPD) and recovery | Sarah's Mental Health Story | Mind. YouTube. 2015. URL: https://www.youtube.com/watch?v=srPBnXx3gd0 [accessed 2022-08-24]
18. Rennick-Egglestone S, Morgan K, Llewellyn-Beardsley J, Ramsay A, McGranahan R, Gillard S, et al. Mental health recovery narratives and their impact on recipients: systematic review and narrative synthesis. Can J Psychiatry 2019 Oct;64(10):669-679 [FREE Full text] [doi: 10.1177/0706743719846108] [Medline: 31046432]
19. McGranahan R, Rennick-Egglestone S, Ramsay A, Llewellyn-Beardsley J, Bradstreet S, Callard F, et al. Curation of mental health recovery narrative collections: systematic review and qualitative synthesis. JMIR Ment Health 2019 Oct 04;6(10):e14233 [FREE Full text] [doi: 10.2196/14233] [Medline: 31588912]
20. Scottish Recovery Network. URL: https://www.scottishrecovery.net/ [accessed 2023-08-23]
21. Smith-Merry J, Freeman R, Sturdy S. Implementing recovery: an analysis of the key technologies in Scotland. Int J Ment Health Syst 2011 May 15;5(1):11 [FREE Full text] [doi: 10.1186/1752-4458-5-11] [Medline: 21569633]
22. Repper J, Breeze J. User and carer involvement in the training and education of health professionals: a review of the literature. Int J Nurs Stud 2007 Mar;44(3):511-519. [doi: 10.1016/j.ijnurstu.2006.05.013] [Medline: 16842793]
23. Pinfold V, Thornicroft G, Huxley P, Farmer P. Active ingredients in anti-stigma programmes in mental health. Int Rev Psychiatry 2005 Apr;17(2):123-131. [doi: 10.1080/09540260500073638] [Medline: 16194782]
24. Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. Lancet 2016 Mar 12;387(10023):1123-1132. [doi: 10.1016/S0140-6736(15)00298-6] [Medline: 26410341]
25. Sheens E, Rhodes P, Dawson L. Encountering anorexia: challenging stigma with recovery stories. Adv Eat Disord 2016 Aug 11;4(3):315-322. [doi: 10.1080/21662630.2016.1217495]
26. Shaw L, Homewood J. The effect of eating disorder memoirs in individuals with self-identified eating pathologies. J Nerv Ment Dis 2015 Aug;203(8):591-595. [doi: 10.1097/NMD.0000000000000347] [Medline: 26230647]
27. Rennick-Egglestone S, Ramsay A, McGranahan R, Llewellyn-Beardsley J, Hui A, Pollock K, et al. The impact of mental health recovery narratives on recipients experiencing mental health problems: qualitative analysis and change model. PLoS One 2019;14(12):e0226201 [FREE Full text] [doi: 10.1371/journal.pone.0226201] [Medline: 31834902]
28. Ng F, Charles A, Pollock K, Rennick-Egglestone S, Cuijpers P, Gillard S, et al. The mechanisms and processes of connection: developing a causal chain model capturing impacts of receiving recorded mental health recovery narratives. BMC Psychiatry 2019 Dec 21;19(1):413 [FREE Full text] [doi: 10.1186/s12888-019-2405-z] [Medline: 31864315]

29. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. Br J Psychiatry 2011 Dec;199(6):445-452. [doi: 10.1192/bjp.bp.110.083733] [Medline: 22130746]
30. Oc87 Recovery Diaries. 2022. URL: https://oc87recoverydiaries.org/ [accessed 2023-11-02]
31. Cordle H, Fradgley J, Carson J, Holloway F, Richards P. Psychosistories of hope and recovery. London: Quay Books; 2011.
32. Bartel H. Men writing eating disorders: autobiographical writing and illness experience in English and German narratives. Bingley: Emerald Publishing Limited; 2021.
33. Yeo C, Rennick-Egglestone S, Armstrong V, Borg M, Charles A, Duke LH, et al. The influence of curator goals on collections of lived experience narratives: a qualitative study. J Recovery Ment Health 2021;4(2):16-28 [FREE Full text] [Medline: 34988285]
34. Roe J, Brown S, Yeo C, Rennick-Egglestone S, Repper J, Ng F, et al. Opportunities, enablers, and barriers to the use of recorded recovery narratives in clinical settings. Front Psychiatry 2020 Oct 30;11:589731 [FREE Full text] [doi: 10.3389/fpsyt.2020.589731] [Medline: 33192738]
35. Woods A, Hart A, Spandler H. The recovery narrative: politics and possibilities of a genre. Cult Med Psychiatry 2022 Jun 21;46(2):221-247 [FREE Full text] [doi: 10.1007/s11013-019-09623-y] [Medline: 30895516]
36. Kaiser BN, Varma S, Carpenter-Song E, Sareff R, Rai S, Kohrt BA. Eliciting recovery narratives in global mental health: Benefits and potential harms in service user participation. Psychiatr Rehabil J 2020 Jun;43(2):111-120 [FREE Full text] [doi: 10.1037/prj0000384] [Medline: 31355653]
37. Rennick-Egglestone S, Elliott R, Smuk M, Robinson C, Bailey S, Smith R, et al. Impact of receiving recorded mental health recovery narratives on quality of life in people experiencing psychosis, people experiencing other mental health problems and for informal carers: Narrative Experiences Online (NEON) study protocol for three randomised controlled trials. Trials 2020 Jul 20;21(1):661 [FREE Full text] [doi: 10.1186/s13063-020-04428-6] [Medline: 32690105]
38. Yeo C, Rennick-Egglestone S, Armstrong V, Borg M, Franklin D, Klevan T, et al. Uses and misuses of recorded mental health lived experience narratives in healthcare and community settings: systematic review. Schizophr Bull 2022 Jan 21;48(1):134-144 [FREE Full text] [doi: 10.1093/schbul/sbab097] [Medline: 34423840]
39. Costa L, Voronka J, Landry D, Reid J, Mcfarlane B, Reville D. "Recovering our stories": a small act of resistance. Stud Soc Just 2012;6(1):85-101 [FREE Full text] [doi: 10.26522/ssj.v6i1.1070]
40. Ng F, Newby C, Robinson C, Llewellyn-Beardsley J, Yeo C, Roe J, et al. How do recorded mental health recovery narratives create connection and improve hopefulness? J Ment Health 2022 Apr 05;31(2):273-280. [doi: 10.1080/09638237.2021.2022627] [Medline: 34983300]
41. Leaving no one behind: The imperative of inclusive development. United Nations Department of Economic and Social Affairs. 2016. URL: https://www.un.org/esa/socdev/rwss/2016/full-report.pdf [accessed 2022-08-24]
42. Roberson QM. Disentangling the meanings of diversity and inclusion in organizations. Group Organ Manag 2016 Jul 26;31(2):212-236. [doi: 10.1177/1059601104273064]
43. Magurran A, McGill B. Biological diversity: frontiers in measurement and assessment. Oxford: Oxford University Press; 2011.
44. Thompson S. Defining and measuring 'inclusion' within an organisation. UK Government Institute of Development Studies. 2017. URL: https://assets.publishing.service.gov.uk/media/5bace6fb40f0b62dd732c49f/123_-_Defining_and_measuring_inclusion_within_an_organisati.pdf [accessed 2022-08-24]
45. Ng F, Newby C, Robinson C, Llewellyn-Beardsley J, Yeo C, Roe J, et al. How do recorded mental health recovery narratives create connection and improve hopefulness? J Ment Health 2022 Apr 05;31(2):273-280. [doi: 10.1080/09638237.2021.2022627] [Medline: 34983300]
46. Henrich J. The weirdest people in the world: How the West became psychologically peculiar and particularly prosperous. Dublin: Penguin Random House; 2020.
47. What is the difference between sex and gender? Office for National Statistics. 2019. URL: https://www.ons.gov.uk/economy/environmentalaccounts/articles/whatisthedifferencebetweensexandgender/2019-02-21 [accessed 2022-12-21]
48. Heidari S, Babor TF, De Castro P, Tort S, Curno M. Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. Res Integr Peer Rev 2016;1:2 [FREE Full text] [doi: 10.1186/s41073-016-0007-6] [Medline: 29451543]
49. Yadav S, Lenka U. Diversity management: a systematic review. Equal Div Incl 2020 May 13;39(8):901-929. [doi: 10.1108/edi-07-2019-0197]
50. Manoharan A, Singal M. A systematic literature review of research on diversity and diversity management in the hospitality literature. Int J Hosp Manag 2017:77-91. [doi: 10.1016/j.ijhm.2017.07.002]
51. Jimenez MF, Laverty TM, Bombaci SP, Wilkins K, Bennett DE, Pejchar L. Underrepresented faculty play a disproportionate role in advancing diversity and inclusion. Nat Ecol Evol 2019 Jul 3;3(7):1030-1033. [doi: 10.1038/s41559-019-0911-5] [Medline: 31160738]
52. Ethnicity facts and figures. UK Government. 2011. URL: https://www.ethnicity-facts-figures.service.gov.uk/ [accessed 2022-08-01]

53. Handbook for external examiners. University of Chichester. 2021. URL: https://www.chi.ac.uk/app/uploads/2021/12/External-Examiners-Handbook-21-22.pdf [accessed 2022-08-24]
54. External examining handbook. University of Exeter. URL: https://as.exeter.ac.uk/academic-policy-standards/tqa-manual/externalexamininghandbook/ [accessed 2022-08-11]
55. Slade M, Rennick-Egglestone S, Llewellyn-Beardsley J, Yeo C, Roe J, Bailey S, et al. Recorded mental health recovery narratives as a resource for people affected by mental health problems: Development of the Narrative Experiences Online (NEON) intervention. JMIR Form Res 2021 May 27;5(5):e24417 [FREE Full text] [doi: 10.2196/24417] [Medline: 34042595]
56. Sexual orientation, UK: 2020. Office for National Statistics. 2019. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/sexuality/bulletins/sexualidentityuk/2020 [accessed 2022-08-24]
57. Number of individuals affected with vision loss in the United Kingdom from 1990 to 2020. Statista. 2020. URL: https://www.statista.com/statistics/1200836/vision-impairments-in-the-uk/ [accessed 2022-12-23]
58. Hearing Link Services. 2022. URL: https://www.hearinglink.org/your-hearing/about-hearing/facts-about-deafness-hearing-loss/#:~:text=Hearing%20loss%20statistics%20in%20the,common%20disability%20in%20the%20UK [accessed 2022-12-23]
59. Kirk-Wade E. UK disability statistics: prevalence and life experiences. House of Commons Library. 2022 Jul 29. URL: https://researchbriefings.files.parliament.uk/documents/CBP-9602/CBP-9602.pdf [accessed 2022-12-23]
60. How common is learning disability? Mencap. 2021. URL: https://www.mencap.org.uk/learning-disability-explained/research-and-statistics/how-common-learning-disability [accessed 2022-12-23]
61. Lauder W, Roxburgh M. Self-neglect consultation rates and comorbidities in primary care. Int J Nurs Pract 2012 Sep 25;18(5):454-461. [doi: 10.1111/j.1440-172x.2012.02065.x]
62. British in Europe. 2017. URL: https://www.britishineurope.org/page/1076136-facts-figures [accessed 2022-12-23]
63. He F, Hu X. Hubbell's fundamental biodiversity parameter and the Simpson diversity index. Ecol Letters 2005 Apr;8(4):386-390 [FREE Full text] [doi: 10.1111/j.1461-0248.2005.00729.x]
64. Morris EK, Caruso T, Buscot F, Fischer M, Hancock C, Maier TS, et al. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. Ecol Evol 2014 Sep;4(18):3514-3524 [FREE Full text] [doi: 10.1002/ece3.1155] [Medline: 25478144]
65. Mitchell M, Baker D, Moorosi N, Denton E, Hutchinson B, Hanna A. Diversity inclusion metrics in subset selection. 2020 Presented at: 2020 AAAI/ACM Conference on AI; February 7-9, 2020; New York. [doi: 10.1145/3375627.3375832]
66. Gulliver K. Racial discrimination in UK housing has a long history and deep roots. London School of Economics. 2017. URL: https://blogs.lse.ac.uk/politicsandpolicy/racial-discrimination-in-housing/ [accessed 2022-12-23]
67. Kobayakawa A. Japan’s modernization and discrimination: what are Buraku and Burakumin? Crit Sociol 2020 Apr 27;47(1):111-132. [doi: 10.1177/0896920520915493]
68. Trawalter S, Driskell S, Davidson M. What is good isn't always fair: on the unintended effects of framing diversity as good. Analy Soc Issues Public Policy 2015 Dec 14;16(1):69-99. [doi: 10.1111/asap.12103]
69. Lamkin J, Nguyen PT, Coverdale JH, Gordon MR. Towards a definition of "self-neglect" in psychiatric patients: descriptions of a case series. Psychiatr Q 2017 Sep 28;88(3):553-560. [doi: 10.1007/s11126-016-9467-6] [Medline: 27682463]
70. Kotera Y. De-stigmatising self-care: impact of self-care webinar during COVID-19. Int J Spa Wellness 2021 Mar 02;4(2-3):213-217. [doi: 10.1080/24721735.2021.1892324]
71. Rennick-Egglestone S, Mawson S. Homes of stroke survivors are a challenging environment for rehabilitation technologies. JMIR Rehabil Assist Technol 2021 Jun 17;8(2):e12029 [FREE Full text] [doi: 10.2196/12029] [Medline: 34137728]
72. Chowdhury R. Self-representation of marginalized groups: a new way of thinking through W. E. B. Du Bois. Bus Ethics Q 2021 Apr 30;31(4):524-548. [doi: 10.1017/beq.2021.5]
73. Lawson C, Fisher A. It's in the sample: the effects of sample size and sample diversity on the breadth of inductive generalization. J Exp Child Psychol 2011 Dec;110(4):499-519. [doi: 10.1016/j.jecp.2011.07.001] [Medline: 21816410]
74. Mishra P, Pandey CM, Singh U, Gupta A. Scales of measurement and presentation of statistical data. Ann Card Anaesth 2018;21(4):419-422 [FREE Full text] [doi: 10.4103/aca.ACA_131_18] [Medline: 30333338]
75. Gardenswartz L, Rowe A. Diverse teams at work : capitalizing on the power of diversity. Alexandria, VA: Society for Human Resource Management; 2008.
76. Lee A, Yang F, Chen C, Wang C, Sun C. Mining perceptual maps from consumer reviews. Decis Support Syst 2016 Feb;82:12-25. [doi: 10.1016/j.dss.2015.11.002]
77. Franzen L, Stark Z, Johnson AP. Individuals with dyslexia use a different visual sampling strategy to read text. Sci Rep 2021 Mar 19;11(1):6449. [doi: 10.1038/s41598-021-84945-9] [Medline: 33742007]
78. Kotera Y, Cockerill V, Green P, Hutchinson L, Shaw P, Bowskill N. Towards another kind of borderlessness: online students with disabilities. Distance Educ 2019 Apr 09;40(2):170-186. [doi: 10.1080/01587919.2019.1600369]
79. Knoester C, Davis EA. Patriotism, competition, nationalism, and respect for the military in US sports: public recognition of American institutionalized sports nationalism. Int Rev Sociol Sport 2021 Nov 05;57(7):1021-1043. [doi: 10.1177/10126902211048769]
80. Froese F, Sekiguchi T, Maharjan M. Human resource management in Japan and South Korea. In: Cooke FL, Kim S, editors. Routledge Handbook of Human Resource Management in Asia. London: Taylor & Francis; 2018:275-294.

Abbreviations

DMHI: digital mental health intervention
INCRESE: Inventory of Characteristics of Recovery Stories
NEON: Narrative Experiences Online
RRN: recorded recovery narrative
SDI: Simpson Diversity Index
SI: Simpson Index
WEIRD: Western, Educated, Individualized, Rich and Democratic

Edited by T Leung, T McCall; submitted 25.11.22; peer-reviewed by N Josifovski, A Lakdawala; comments to author 21.12.22; revised version received 24.12.22; accepted 01.01.23; published 17.04.23.

Please cite as:
Kotera Y, Rennick-Egglestone S, Ng F, Llewellyn-Beardsley J, Ali Y, Newby C, Fox C, Slade E, Bradstreet S, Harrison J, Franklin D, Todowede O, Slade M
Assessing Diversity and Inclusivity is the Next Frontier in Mental Health Recovery Narrative Research and Practice
JMIR Ment Health 2023;10:e44601
URL: https://mental.jmir.org/2023/1/e44601
doi:10.2196/44601
PMID:37067882

Viewpoint

Reintroducing the Effortless Assessment Research System (EARS)

Monika N Lind[1], MSc; Lauren E Kahn[1], PhD; Ryann Crowley[1], MSc; Wyatt Reed[1], BSc; Geordie Wicks[1], BSc; Nicholas B Allen[1], PhD

Center for Digital Mental Health, University of Oregon, Eugene, OR, United States

Corresponding Author:
Nicholas B Allen, PhD
Center for Digital Mental Health
University of Oregon
Straub Hall
Eugene, OR, 97403
United States
Phone: 1 541 346 4075
Email: nallen3@uoregon.edu

Abstract

This paper reintroduces the Effortless Assessment Research System (EARS), 4 years and 10,000 participants after its initial launch. EARS is a mobile sensing tool that affords researchers the opportunity to collect naturalistic, behavioral data via participants' naturalistic smartphone use. The first section of the paper highlights improvements made to EARS via a tour of EARS's capabilities—the most important of which is the expansion of EARS to the iOS operating system. Other improvements include better keyboard integration for the collection of typed text; full control of survey design and administration for research teams; and the addition of a researcher-facing EARS dashboard, which facilitates survey design, the enrollment of participants, and the tracking of participants. The second section of the paper goes behind the scenes to describe 3 challenges faced by the EARS developers—remote participant enrollment and tracking, keeping EARS running in the background, and continuous attention and effort toward data protection—and how those challenges shaped the design of the app.

(JMIR Ment Health 2023;10:e38920) doi:10.2196/38920

KEYWORDS

mobile sensing; passive sensing; personal sensing; digital phenotyping; ecological momentary assessment; digital mental health

Introduction

In 2018, we introduced the Effortless Assessment of Risk States (EARS) tool, a mobile sensing software that collects behavioral and interpersonal data via naturalistic smartphone use [1]. The original motivation for the development of the tool was to provide a new scalable approach for continuous measurement of behavior to facilitate the prediction of, and timely response to, mental health crises such as increases in suicidal thoughts and behaviors [1,2]. We designed the EARS tool to capture multiple indices of behavior by continuously collecting data from the various sensors in an individual's personal smartphone during their normal use the device. This makes the solution highly scalable because there is no need to provide the user with special wearables or to ask them to change their current behavior by integrating a new device into their lifestyle. This substantially reduces the compliance burden associated with the measurement. We selected the indices measured based on findings that demonstrate their links to mental health states such as depression and suicidality [1]. These indices included physical activity, geolocation, sleep, phone use and duration, music choice, facial expressions, acoustic vocal quality, and natural language use [1]. Although the continuous collection of behavioral data from smartphones has been extensively used in some commercial applications, the EARS tool was specifically designed to only be used in research studies that are regulated by properly constituted Institutional Review Boards or Human Research Ethics Committees that require, among other things, fully informed consent.

Much has changed since the launch of EARS, including its name. The acronym and the "bunny ears" icon remain (see Figure 1), but EARS now stands for the Effortless Assessment *Research System*. The name change not only reflects its broader application to a wide variety of use cases in behavioral research, in addition to our initial focus on the prediction of psychiatric risk, but also represents the extensive improvements made to

EARS over the last 4 years. The first section of this paper will highlight those improvements via a tour of EARS's capabilities across iOS and Android versions. The second section of this paper will go behind the scenes to describe 3 challenges faced by the EARS developers and how those challenges shaped the design of the app.

Figure 1. The Effortless Assessment Research System (EARS) app icon.

What EARS Does

Overview

The single most important improvement to EARS is the addition of a comparable iOS version to match the Android version. EARS on iOS is now programmed in Swift and is available on the Apple App Store (current version 2.1.2 at the time of this paper; search "EARS Mobile Sensing" in the Apple App Store). The iOS and Android versions share a set of core data streams (see Table 1). With minimal burden on the research participant, EARS can collect—with fully informed consent—data from a smartphone's accelerometer, GPS, activity and motion (ie, walking, running, cycling, and driving), music and media listening, selfies, and text typed into the keyboard. For research teams who wish to collect the timing of typed text but not the content, both versions can collect the timing of keystrokes without collecting content. In addition, EARS allows researchers to administer custom ecological momentary assessment (EMA) surveys. In the Android version only, EARS can also collect app usage data and ambient light, although recent developments in iOS version 16 have opened up the possibility of collecting app usage in iOS using the Screentime application programming interface. This capability is current being built into the iOS version of EARS. In the iOS version only, EARS can also collect battery level, charge status, and call status. Researchers can decide which data streams to collect for their study, and we protect the confidentiality of these data with industry-standard encryption.

Table 1. Data streams collected by Effortless Assessment Research System (EARS) on Android and iOS and example behavioral features to which they can contribute.

Data stream	Android	iOS	Relevant behavioral features
Accelerometer	✓	✓	Physical activity and sleep
GPS	✓	✓	Patterns of mobility and sematic location
Activity and motion	✓	✓	Physical activity
Music and media	✓	✓	Musical and emotional qualities of music listened to
Selfies	✓	✓	Facial expression
Typed text: content and timing	✓	✓	Language content analysis (eg, sentiment) and keyboard kinematics
Typed text: timing only	✓	✓	Keyboard kinematics
EMA[a]	✓	✓	Self-report of experiences, activities, and contexts
App usage	✓	Coming soon	Extent and timing of phone usage, broken down into specific app categories
Ambient light	✓		Sleep patterns
Battery level		✓	Regularity of daily behavioral patterns
Charge status		✓	Regularity of daily behavioral patterns
Call status		✓	Frequency of social contacts

[a]EMA: ecological momentary assessment.

Keystroke Data

The collection of text typed into the smartphone keyboard sets EARS apart from many other mobile sensing tools. All typed text, with the exception of anything typed into a secure field such as passwords and credit card numbers, are captured and stamped with the date, time, and application in which the keyboard is active. In other words, if a research participant sends a text message to their spouse while they have EARS installed, researchers can, with appropriate informed consent, know what they typed, when they typed it, and what application they used.

(Obviously, these data are very privacy sensitive.) However, to protect the confidentiality of the recipient of the message (who has presumably not consented to having their data collected), there is no record of who received the message or their reply. On the Android version, EARS captures these data regardless of which keyboard a research participant uses. On the iOS version, EARS's text capture depends on the participant using a custom EARS keyboard. The EARS keyboard for iOS has undergone a recent overhaul focused on improving its accuracy, autocorrection, and predictive text functions, as well as adding convenient features such as haptic feedback and trackpad mode (ie, long pressing the space bar for larger scrolling).

EMA Surveys

Another recent, major update to EARS transferred control of EMA administration to the researcher. Previously, EARS staff had to program each survey via customized development that required many clarifications and iterations. Now, researchers use a convenient EARS dashboard to assemble their survey(s) from a set of item types (slider, true or false, single choice, multiple choice, the time of day, text entry, and informational text), define survey question content and logic (eg, branching), and customize the EMA schedule. EMA delivery schedules can be predetermined by researchers, but currently, event-based triggering on EMA surveys (eg, by specific patterns of mobile sensing data) is not available. Researchers also choose whether to enable several EMA-related functions, including EMA streak-based gamification (to increase compliance), scheduling bursts, and a risk alert function. The risk alert function, which automatically scores a completed survey and detects instances of predefined response patterns indicating high risk (eg, endorsing suicidal intent), is suitable for research with high-risk populations. When a predefined high-risk response pattern is detected, the system alerts study staff or clinicians by automatically sending a text message or email containing the information required to conduct a safety check with the participant.

The Researcher Dashboard

The researcher dashboard itself represents a major update to EARS and serves 3 key functions. First, the dashboard facilitates the enrollment of participants and the installation of EARS on participant devices. Second, the dashboard enables data quality monitoring with daily reports on the specific data streams uploading from each participant. Third, as mentioned above, the dashboard allows researchers to design and administer EMAs.

Deprecated Features

The development and maintenance of EARS requires flexibility to respond to changing operating system parameters and researcher needs. As such, updates to EARS have also resulted in the deprecation of several features. EARS no longer measures SMS text messaging frequency, nor does it capture in-call acoustic voice properties. These depreciations are in response to technical challenges and legal complexities. For example, the passive collection of voice data proved to be complex not only at the technical level (eg, it was difficult to reliably trigger the collection of voice data when the research participants were speaking into their phone) but also legally (eg, ensuring that the collection of the participant's voice did not collect any voice data from a nonconsenting third party during phone calls). Because of our interest in emphasizing passive methods of assessment wherever possible, the video diary function (which required an active response from the participant) has also been phased out in favor of the selfie data stream. Sustaining EARS has also changed our approach to licensing and sharing open-source code. The requirements to maintain the software and support the implementation and maintenance of studies are very time-consuming. Although there are a number of free-to-the-user mobile sensing tools available, we have found that behavioral researchers are often looking for a solution that supports the implementation, administration, and analysis of these studies. As such, EARS is now licensed to a spin-out company, Ksana Health Inc, that can provide researchers with end-to-end support to conduct mobile sensing studies. This end-to-end support exceeded the capabilities of our university-based research center.

Data Delivery to Researchers

At the conclusion of their study, researchers have the option to take delivery of EARS data in a raw form or as extracted features. In this context, "features" refer to metrics derived from the raw EARS data produced by participants. EARS data features tend to be more tractable for many research teams than raw EARS data. For example, our data scientists can take raw accelerometer data—of which a typical participant produces 4,000,000 records per day—and derive interpretable sleep features, including bedtime, waketime, and sleep duration. In addition, our data scientists derive location and travel features from raw GPS data, including but not limited to time spent at home, time spent traveling, and the number of travel events per day. Following the sentiment analysis approach of Byrne and colleagues [3], which showed a relationship between EARS-typed text and stress, our data scientists offer typed text features that measure positive sentiment and negative sentiment words. We also offer features that measure first-person pronoun use and absolutist language, which are related to mental health, especially indices of depression and suicide risk [4,5].

To date, over 10,000 people across Europe, the United Kingdom, North America, and Australia have participated in research studies using EARS. Study lengths vary from 1 week to 1 year of mobile sensing data collection. EARS has gathered mobile sensing data from participants on 88% of study days. Due to challenges discussed below, the iOS version lags behind the Android version with respect to data completion, with EARS on iOS collecting at least some mobile sensing data on 85% of study days versus EARS on Android's 98%. Although EARS performs well on both platforms, the relative strength of EARS on Android supports its suitability for research conducted in low- and middle-income countries where Android dominates the mobile phone market.

The updates to EARS address a series of core challenges faced by EARS developers and researchers conducting studies using mobile sensing.

Challenge #1: Participant Tracking

In all, 3 major challenges shaped the development of EARS. The first challenge, familiar to every researcher who has collected longitudinal data, was participant tracking. Early problems with participant tracking included unpredictable participant behaviors such as deleting or losing the EARS app or purchasing a new phone. In the pilot version of EARS, we had no way of detecting these behaviors in real time, which meant that we often did not discover missing data until the end of data collection. Remote installation of EARS was also impossible, so even in the case of a conscientious participant who notified us of their new phone purchase, the reinstallation of EARS required a laboratory visit or an email exchange with EARS staff. (EARS can easily be enabled on a new phone by reinstalling the app, but research staff need to be able to track the installation codes used for the installations on the old and new phones so that the participants data can be concatenated into 1 file for the final data analyses). In addition, to avoid causing headaches for research coordinators, participant tracking required EARS to ensure that the correct version of EARS is installed by participants, link mobile sensing data with established participant identifiers, and prevent unknown people from installing EARS and uploading mystery data. EARS developers have addressed these requirements by building a researcher-facing dashboard.

Among other capabilities, the dashboard allows researchers to set up their study, enroll participants, and monitor mobile sensing data uploads. Study setup includes choosing which sensors to collect and customizing the content and schedule of EMA questions, which is then built into the bespoke version of EARS for that study. The enrollment of participants entails the creation on the dashboard of a custom, single-use installation code, in both 16-digit, hexadecimal form and QR code form. This installation code ensures that the custom version of EARS—tailored to the choices made by researchers during setup—is installed only on the devices of consenting participants. The dashboard also interfaces with Amazon Web Services (AWS) and Qualtrics to enable EARS installation as part of remote informed consent.

The monitoring of mobile sensing data uploads is facilitated by 3 figures on the dashboard: an overview (see Figure 2), a device list (see Figure 3), and a device dot plot (see Figure 4). The overview conveys the number of participants uploading without issues, uploading with issues (ie, missing expected sensors), and not uploading at all. The device list categorizes participant devices by status (eg, uploading without issues) and displays their last upload date, research site (important for multisite studies), and operating system. Finally, the device dot plot allows researchers to determine at a glance which sensors are uploading on which days for which participants. This information is crucial because many uploading issues are caused by participant behavior, which can be remedied by researcher reminders. For example, in iOS, it is possible for EARS participants to switch away from the EARS keyboard. If a participant does so, the dashboard will flag their device as not uploading text, which will signal the researcher to intervene.

Figure 2. The overview conveys the number of participants uploading without issues, uploading with issues, not uploading at all, and deactivated. "Deactivated" denotes Effortless Assessment Research System (EARS) installations that have been deactivated by the researcher, usually when the participant finishes the study.

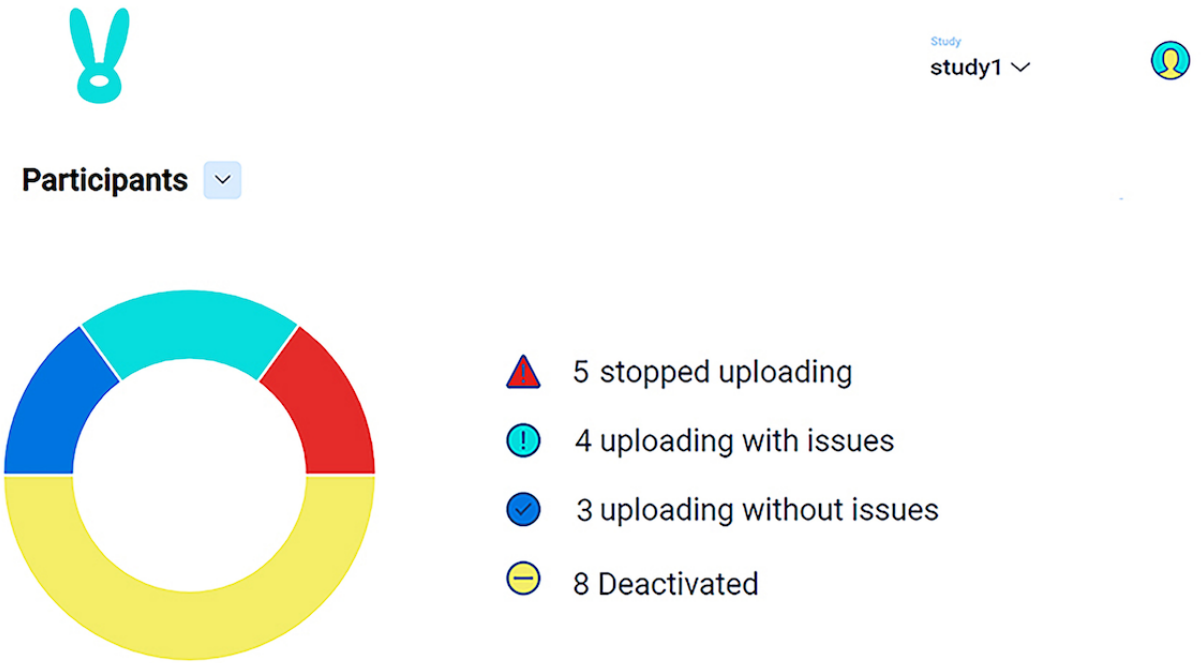
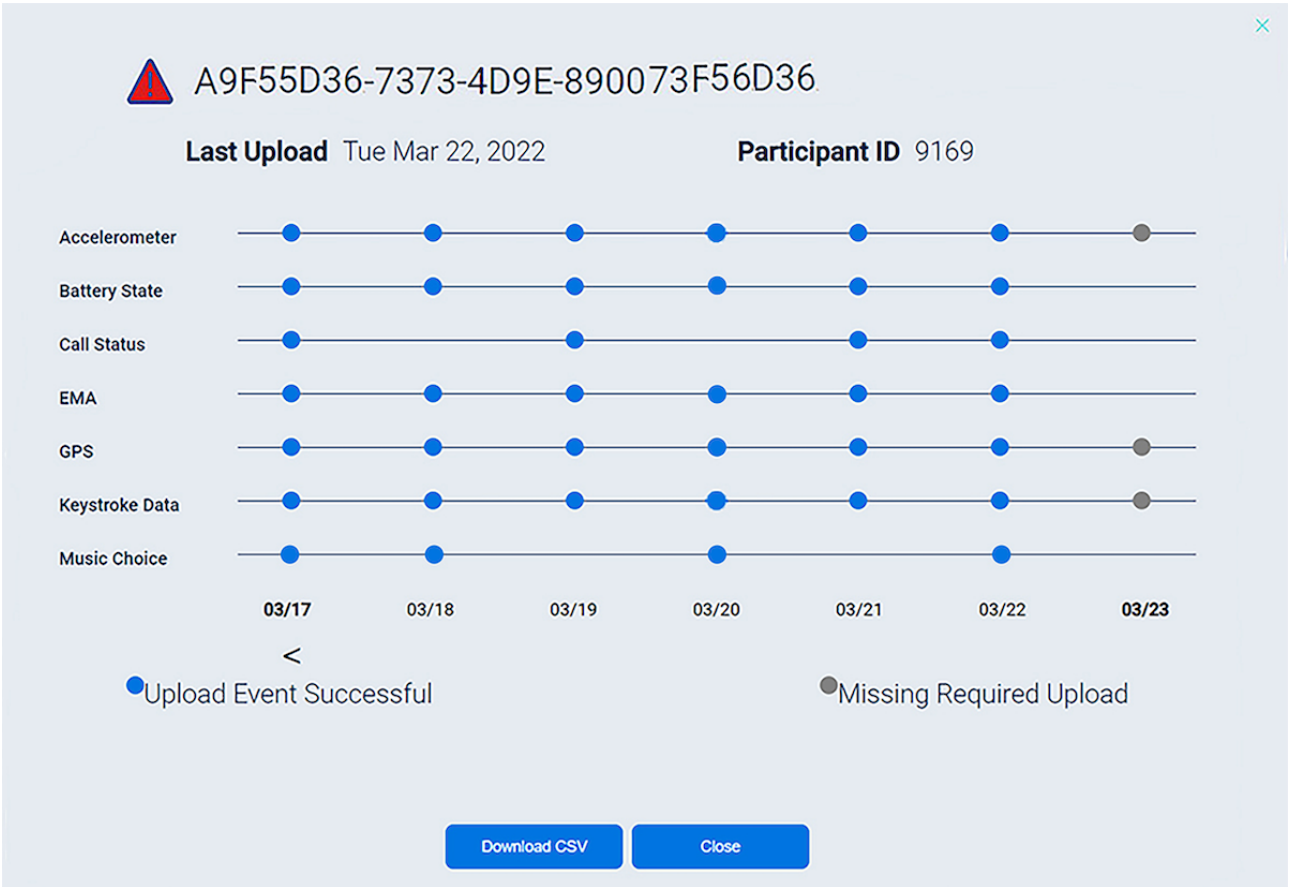


Figure 3. The device list elaborates on the participants categorized in the overview as uploading with issues and not uploading at all. The device list displays their participant ID, last upload date, research site, and operating system.

Figure 4. The device dot plot allows researchers to determine at a glance which sensors are uploading on which days for individual participants. EMA: ecological momentary assessment.



Challenge #2: Keeping EARS in Memory

The second challenge that shaped the development of EARS was keeping EARS running on participants' smartphones during data collection. Battery optimization and privacy, among other concerns, have driven both the Android and iOS operating systems to close mobile apps automatically under certain conditions. These conditions include apps that run in the background without active user interaction and apps that draw on sensors such as GPS above a certain threshold, both of which are related to mobile sensing functionality. The severity of this issue on Android depends on the Android manufacturer (see the "Don't kill my app!" website for more information [6]). A few Android handset manufacturers (eg, OnePlus and Huawei) account for the majority of the problems with keeping EARS in memory on Android. So far, the limited popularity of those manufacturers among our participants has contributed to the excellent performance of EARS on Android. iOS takes a more restrictive approach than Android.

After a relatively smooth Android-only launch of EARS in 2018, developing for iOS forced our programmers to solve 2 main problems. First, early testing on iOS revealed spotty collection of GPS and accelerometer data with evidence of the app getting repeatedly force-closed. It appeared that iOS took a dim view of our regular sampling of the GPS and accelerometer sensors. We adjusted by moving the GPS sensor to a threshold-based approach, that is, only collecting GPS data when the phone moved over 100 meters. We now apply this distance-based threshold in both iOS and Android versions. For the accelerometer, we changed collection to a retroactive application programming interface rather than collecting in real time. Not only did these changes prevent EARS from getting automatically closed by iOS, but they also reduced the impact of EARS on battery life.

The second main problem was that iOS optimizes user experience by automatically closing unused apps over time. Although the time varies among device models and user usage habits, on average, this results in most apps being closed after approximately 24 hours without active use. This means that EARS research participants must interact with the EARS app once per day to ensure ongoing sensor data collection. We addressed this problem by requiring that all research studies send 1 prompt to open the app to participants per day. The least onerous but still adequate version of this prompt requires just 1 tap: participants simply acknowledge a notification from EARS, which brings EARS to the foreground, although most studies take the opportunity to collect a short EMA survey at this time. An upside of this requirement is that it maintains participants' awareness of EARS running on their phones, thereby strengthening ongoing informed consent.

Common to Android and iOS, a third aspect of the challenge of keeping EARS in memory was training researchers to train participants. Continuous data collection by EARS depends on participant behavior, and researchers are the primary point of contact for participants. As such, it is critical that researchers receive comprehensive training from the EARS team. For example, in a scenario in which an iOS user switches away from the EARS keyboard, the researcher detects the switch from the dashboard and can prompt the participant to switch back to the EARS keyboard to ensure ongoing data collection. Ultimately, the goal is to achieve as close to continuous mobile sensing data collection as possible, and the participant's ability to facilitate this goal depends on the researcher's fluency with the dashboard and the participant's fluency with enabling EARS capabilities—both of which are supported by the extensive training, resources, and support provided by the EARS team.

Challenge #3: Data Protection

A third challenge that continues to shape the development of EARS is the protection of participant data. Many EARS researchers study adolescents, which increases the importance of rising to this challenge. Our recent data protection efforts have focused on 2 issues: encryption protocols and compliance with the with best practices and statutory requirements.

Our encryption protocols changed after careful consideration of likely threats to data protection. Per advice from data security professionals, we decided that the storage of the data in multiple locations presented a significant threat to data security. To address this increased risk, we made changes that minimized the storage locations and transfers. In short, we reduced our encryption layers from 2 to 1 and took advantage of the cloud-based data processing capabilities afforded by our cloud service, AWS. In our current approach, the data are single-encrypted in transit from the phone to the cloud. Upon arrival on the cloud, the data are decrypted then re-encrypted with AWS's standard server-side encryption. Data processing and feature extraction occur on the cloud without ever having to manually decrypt data, instead taking advantage of AWS's built-in decryption (and subsequent re-encryption) when temporarily bringing data into memory for processing. Researchers take delivery of their raw data and features in a single-encrypted transfer from the cloud. Some researchers opt to receive only extracted features to avoid ever having raw mobile sensing data stored on their local systems. Our long-term goal remains to provide researchers with the option to constrain data processing and feature extraction to the phone itself, thereby making the phone the exclusive home of the raw data. In the meantime, the cloud servers where EARS data live are also managed by a cybersecurity analyst employed by Ksana Health.

Ksana Health has now completed a third-party assessment for security and privacy controls (System and Organization Controls [SOC] 2 Type 2). A SOC 2 Type 2 report is an internal controls report capturing how a company safeguards data and how well those controls are operating. Companies that use cloud service providers use SOC 2 reports to assess and address the risks associated with third-party technology services. EARS received unqualified approval attesting to the strength of data security controls and compliance with best practices and statutory requirements, including the following:

- American Institute of Certified Public Accountants (AICPA), Trust Services Criteria
- Healthcare Insurance Portability and Accountability Act (HIPAA) Security Rule
- HIPAA Privacy Rule

- General Data Protection Regulation (GDPR) Controls Mapping

To highlight the details of one of these requirements, the European Union's adoption of the GDPR in 2016 and its enforcement starting in 2018 required us to ensure that EARS meets those standards. To the extent that the information we collect is health data or another special category of personal data subject to GDPR, we ask users for their explicit consent to process the data. We obtain this consent separately when they enroll in a study. Additionally, users can use the account settings to withdraw consent at any time, including by stopping the use of a feature, removing our access to a third-party service, unpairing the device, or deleting the data or the account. Users can also uninstall the app at any time, which halts data collection. We support advancements in data and privacy protections, and we expect regulations around the world to follow the European Union's lead. In fact, California already has with the California Consumer Privacy Act. As such, we view evolving data protection regulations as a challenge that will continue to improve EARS and mobile sensing research in general.

Future Directions

We hope that the increasing the adoption of EARS by mobile sensing researchers will enable us to develop EARS not just in response to challenges but in the pursuit of innovation. Mobile sensing opens up the possibility of collecting behavioral data in a highly scalable way that is continuous, ecological, and objective without creating significant participant burden. These data can be used to provide answers to questions that were not previously addressable with self-report or laboratory-based methods. However, to achieve this goal, we must have tools that address researchers' needs while respecting participants' rights to privacy and data security. EARS represents one attempt to continuously improve these tools to achieve these goals.

Acknowledgments

The authors wish to thank Michelle Byrne for her foundational contributions to the original Effortless Assessment of Risk States (EARS) app, Audra McClure-Begley for providing Effortless Assessment Research System (EARS) user compliance data, and Jennifer Magaña for providing images of the EARS dashboard. All authors (MNL, LEK, RC, WR, GW, and NBA) contributed to conception and design, drafted or revised the article, and approved the submitted version.

Conflicts of Interest

The authors (MNL, LEK, RC, WR, GW, and NBA) all hold equity interests in Ksana Health Inc, the company that has the sole commercial license for certain versions of the Effortless Assessment Research System (EARS) mobile phone app and some related EARS tools. The authors LEK, RC, WR, and NBA also receive salary from Ksana Health Inc.

References

1. Lind MN, Byrne ML, Wicks G, Smidt AM, Allen NB. The Effortless Assessment of Risk States (EARS) tool: an interpersonal approach to mobile sensing. JMIR Ment Health 2018 Aug 28;5(3):e10334 [FREE Full text] [doi: 10.2196/10334] [Medline: 30154072]
2. Allen NB, Nelson BW, Brent D, Auerbach RP. Short-term prediction of suicidal thoughts and behaviors in adolescents: can recent developments in technology and computational science provide a breakthrough? J Affect Disord 2019 May 01;250:163-169 [FREE Full text] [doi: 10.1016/j.jad.2019.03.044] [Medline: 30856493]
3. Byrne ML, Lind MN, Horn SR, Mills KL, Nelson BW, Barnes ML, et al. Using mobile sensing data to assess stress: associations with perceived and lifetime stress, mental health, sleep, and inflammation. Digit Health 2021 Aug 27;7:20552076211037227 [FREE Full text] [doi: 10.1177/20552076211037227] [Medline: 34777852]
4. Al-Mosaiwi M, Johnstone T. In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. Clin Psychol Sci 2018 Jul 05;6(4):529-542 [FREE Full text] [doi: 10.1177/2167702617747074] [Medline: 30886766]
5. Edwards T, Holtzman NS. A meta-analysis of correlations between depression and first person singular pronoun use. J Res Personal 2017 Jun;68:63-68 [FREE Full text] [doi: 10.1016/j.jrp.2017.02.005]
6. Don't kill my app!. Urbandroid Team. URL: https://dontkillmyapp.com/ [accessed 2023-03-19]

Abbreviations

AICPA: American Institute of Certified Public Accountants
AWS: Amazon Web Services
EARS: Effortless Assessment Research System (formerly Effortless Assessment of Risk States)
EMA: ecological momentary assessment
GDPR: General Data Protection Regulation
HIPAA: Healthcare Insurance Portability and Accountability Act
SOC: System and Organization Controls

Edited by J Torous; submitted 21.04.22; peer-reviewed by M Jensen, V Pejovic, P Chow; comments to author 02.06.22; revised version received 11.12.22; accepted 13.12.22; published 26.04.23.

Please cite as:
Lind MN, Kahn LE, Crowley R, Reed W, Wicks G, Allen NB
Reintroducing the Effortless Assessment Research System (EARS)
JMIR Ment Health 2023;10:e38920
URL: https://mental.jmir.org/2023/1/e38920
doi:10.2196/38920
PMID:37099361

Viewpoint

Health Economic Evaluation of Cognitive Control Training for Depression: Key Considerations

Constance Nève de Mévergnies[1,2], MSc; Nick Verhaeghe[2,3], PhD; Ernst H W Koster[1], PhD; Chris Baeken[4,5,6], MD, PhD; Yannick Vander Zwalmen[1], MSc; Kristof Hoorelbeke[1], PhD

[1]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium
[2]Department of Public Health and Primary Care, Ghent University, Ghent, Belgium
[3]Department of Public Health, Vrije Universiteit Brussel, Brussels, Belgium
[4]Department of Head and Skin, Ghent University, Ghent, Belgium
[5]Department of Psychiatry, University Hospital Brussel (UZBrussel), Brussels, Belgium
[6]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Kristof Hoorelbeke, PhD
Department of Experimental Clinical and Health Psychology
Ghent University
Henri Dunantlaan 2
Ghent, 9000
Belgium
Phone: 32 9 264 64 74
Email: kristof.hoorelbeke@Ugent.be

Abstract

Depression is a serious and burdensome psychiatric illness that contributes heavily to health expenditures. These costs are partly related to the observation that depression is often not limited to a single episode but can recur or follow a chronic pathway. In terms of risk factors, it is acknowledged that cognitive impairments play a crucial role in vulnerability to depression. Within this context, cognitive control training (CCT) has shown its effectiveness in reducing the risk for recurrence of depression. CCT is low cost intensive and can be provided as a web-based intervention, which makes it easy to disseminate. Despite increasing interest in the field, studies examining the cost-effectiveness of CCT in the context of depression are largely missing. Health economic evaluation (HEE) allows to inform decision makers with evidence-based insights about how to spend limited available (financial) resources in the most efficient way. HEE studies constitute a crucial step in the implementation of a new intervention in clinical practice. Approaching preventive measures for depression such as CCT from an HEE perspective is informative to health policy, fostering optimal use of health expenditures. The aim of this paper was to inform and guide researchers during the phase of designing HEE studies in the context of CCT for depression. A clear view of CCT's cost-effectiveness is paramount for its clinical implementation.

(JMIR Ment Health 2023;10:e44679) doi:10.2196/44679

KEYWORDS

health economic evaluation; cost utility; cognitive control training; CCT; depression recurrence; health policies

Introduction

Depression is a leading cause of disability worldwide, with devastating effects on individual well-being and functioning. Globally, >5% of the population is affected by depression [1]. In terms of 12-month and lifetime prevalence of major depressive disorder (MDD), a recent American study reported 10.4% and 20.6% prevalence rates, respectively [2].

Available and effective psychological treatments for depression are expensive and typically require extensive expertise, resulting in only a small portion of individuals with depression being able to profit from those. Despite significant investments in pharmacological and psychological interventions in the past decades, low-income and middle-income countries still have limited accessibility to treatment, resulting in 75% of people with mental health disorders worldwide not receiving adequate care [1].

The severity of depression, combined with the low accessibility of treatments, leads to important personal and societal costs [3,4]. In 2006, Kessler et al [5] estimated, among a sample of US workers, that a single episode of MDD was associated with an average of >5 weeks of lost productivity per worker, resulting in an annual capital loss of >US $36 billion to employers. As such, there is an unmet need for widely accessible and cost-effective treatments for depression. Recently, the use of various technologies to deliver treatments has shown potential, especially in its ability to reach a large proportion of individuals who would otherwise not have access to psychological interventions [3,6]. One such intervention that shows substantial promise is cognitive control training (CCT) [7,8].

This paper discusses the current state of the art regarding the role of cognitive control impairments as a risk factor for depression and, consequently, the relevance of CCT in the prevention of depression. Building on recent meta-analyses suggesting a need for high-quality and adequately powered health economic evaluation (HEE) studies [9], we argue that adopting a health economic perspective is paramount to comprehensively evaluate the dissemination potential of CCT for depression in an evidence-based way. For this purpose, we have provided theoretical and practical guidelines about how to design HEE studies that allow to address this specific question.

Cognitive Control Deficits as a Depressive Risk Factor

Identifying risk factors associated with recurrent depression has been the focus of extensive studies in the past decades, in which cognitive processes have shown to play an important role. Specifically, cognitive control—referring to the ability to flexibly adapt thoughts and behavior to achieve one's goals [10]—has been identified as an important cognitive risk factor in the context of depression [11]. That is, cognitive control deficits have been observed in individuals with MDD [12] and often persist following remission [13], placing them at risk for the recurrence of depression [14].

Individuals with remitted depression show, among other things, working memory impairments, diminished processing speed and attention, and executive functioning impairments. These deficits have major consequences on an individual's daily functioning and provide an important source of disability [15]. One of the main causes of costs associated with depression is loss of productivity owing to multiple or prolonged periods of sick leave or unemployment, where cognitive complaints interfere with occupational activities [16].

Previous studies suggest a central role for rumination in linking cognitive control deficits to risk for recurrent depression. Rumination can be defined as a recurrent process of perseverative negative thinking about feelings, problems, or disturbing experiences [17,18]. Elevated rumination, a key risk factor for depression, is associated with poor rates of recovery following pharmacological treatment for depression and slow response to psychotherapy [19,20]. In this context, it has been suggested that cognitive control is essential to be able to disengage from negative self-referential thoughts upon confrontation with stressors [21] or to inhibit habitual ruminative responses [22], where failure to do so may result in prolonged exposure to negative thoughts and affect. Consistent with this hypothesis, previous studies suggest that cognitive control deficits predict future increases in depressive symptomatology in patients with remitted depression, a relationship that appears to be mediated by rumination [23].

Although current pharmacological or psychological treatments have proven to be effective in decreasing depressive symptomatology, existing interventions typically do not target cognitive impairments directly. Moreover, there is a problem of recurrence of depression after initial treatment, where perturbations at the level of cognitive functioning show a positive association with the number of previous depressive episodes [24]. As such, it has been suggested that cognitive control deficits reflect increased vulnerability to recurrent depression [14]. Together, these findings suggest the need for innovative interventions aimed at reducing these specific risk factors.

CCT for Depression

A particularly promising approach is CCT—a computerized psychological intervention during which operations and mental exercises aimed at improving cognitive control are performed [8]. Within this context, cognitive training has shown to be a useful intervention to reduce vulnerability to depression [25-27].

Among the multiple CCT tasks used within the context of depression, the effects of the adaptive paced auditory serial addition task (aPASAT [8]; for a review, please see the paper by Koster et al [7]) have been the most intensively studied. During this computer task, participants are presented with digits in a sequential fashion. They are asked to provide the sum of the 2 last heard digits. The aPASAT tailors the exercise based on individual performance by modifying the intertrial interval following every 4 consecutive correct (−100 milliseconds) or incorrect (+100 milliseconds) responses. Within the context of this task, individuals are continuously required to update information in working memory while preventing interference from previous responses.

Several studies suggest the beneficial effects of aPASAT training on depressive symptomatology. For instance, Siegle et al [8] investigated the impact of CCT on the neurobiological and cognitive mechanisms underlying depression in a sample of patients with clinical depression (N=31). In the first pilot study, the authors combined treatment as usual (TAU) with 6 sessions of aPASAT training and an attention training session during a period of 2 weeks. They observed beneficial effects of CCT on rumination and level of depressive symptomatology, in comparison with the TAU control group. In an extended sample (N=43), Siegle et al [28] replicated the previous positive effects of CCT on rumination among patients with MDD and observed a decreased need for clinical care at 1-year follow-up, providing the first evidence for the long-term benefits of CCT. The clinical potential of cognitive training has been confirmed in other studies reporting the beneficial effects of CCT in patients with MDD (eg, the studies by Brunoni et al [29],

Iacoviello et al [30], Morimoto et al [31], and Sommer and Plewnia [32], but see the studies by Ferrari [33] and Moshier and Otto [34]).

Importantly, the beneficial effects of aPASAT training on depressive symptomatology were also observed in individuals at risk for (recurrence of) depression. For instance, in a sample of 68 individuals with remitted depression, Hoorelbeke and Koster [35] observed beneficial effects of 10 sessions of aPASAT training on rumination and depressive symptomatology compared with an active control condition. Similarly, in a recent randomized controlled trial (N=92 individuals with remitted depression) in which CCT was combined with an intensive experience sampling procedure, Hoorelbeke et al [36] observed beneficial effects of CCT on the deployment of rumination in daily life. Moreover, compared to an active control condition, CCT reduced the risk for recurrence of depression in individuals with remitted depression, as shown by lower recurrence rates in a period of one year following training [37].

CCT has shown interesting preventive effects in the context of depression. Consistent with this finding, a recent meta-analysis suggested small to medium effects of aPASAT training on depressive symptomatology and rumination, both immediately after training and at follow-up [38]. Interestingly, the authors explored the clinical relevance of their results by converting the pooled effect sizes into number needed to treat (NNT; ie, the number of patients who would need to be treated by an intervention to observe improvement in 1 patient). The results suggest that a similar number of patients should be treated with aPASAT training (NNT=6.15) to approximate the effects of monotherapeutic pharmacological (NNT=5.95) or psychotherapeutic interventions (NNT=7.13 [39]). Despite these encouraging findings, several challenges remain, such as identifying the precise mechanisms of transfer and the optimal parameters of administration (eg, training dosage and use of therapist-supported vs unguided delivery), which have mostly been left unexplored so far [40] (but see the study by Launder et al [25]).

Taken together, CCT is a highly promising preventive intervention for recurrent depression. Given that recent findings suggest that CCT may reduce the risk for recurrence of depression [37], CCT is also likely to influence related factors, including work absenteeism and costs associated with depression. By targeting underlying vulnerability factors that are currently not targeted by traditional interventions for depression, CCT has multiple characteristics that make it an interesting choice from a clinical perspective: (1) it could be relatively low cost intensive for patients; (2) it is highly accessible as it can be provided as a web-based intervention; (3) it is easily disseminated [7]; (4) it is a relatively inexpensive intervention, especially when provided on a large scale; and (5) compared with pharmacological treatments, limited side effects have been reported for CCT [26].

Introducing a Health Economic Perspective

In recent decades, the importance of investing in health has been widely emphasized, resulting in a strong emphasis on the efficient use of budgets (eg, encouraging investments in preventive efforts) [41]. Therefore, policy makers have been strongly encouraged to opt for the most efficient interventions to ensure population health [42]. On the basis of effectiveness studies, digital health interventions seem to hold great promise as preventive interventions for depression [6,43-45]. In particular, CCT is a potentially scalable preventive intervention for recurrent depression that would suit web-based dissemination [38,40]. However, it is clear that a sole focus on the efficacy and effectiveness of CCT is insufficient for policy making. This would require HEE studies, examining the value for money of CCT compared with alternative strategies, aiming to inform different stakeholders (eg, decision makers) with evidence-based insights [42].

Health economics and HEE can be defined as “the study of how scarce resources are allocated among alternative uses for the care of sickness and the promotion, maintenance and improvement of health, including the study of how healthcare and health-related services, their costs and benefits, and health itself are distributed among individuals and groups in society” [46]. Multimedia Appendix 1 [42,46-51] provides an overview of the key health economic terms relevant to this paper.

Although multiple studies have explored the cost-effectiveness of other digital health interventions for depression [52-54] (for a recent meta-analysis, refer to the paper by Rohrbach et al [9]), the cost-effectiveness of CCT remains to be evaluated. If we aim to provide clear information to policy makers and governments, costs and health outcomes of CCT should be explored, compared with the available alternatives. As such, HEE forms a crucial step for the implementation of CCT.

To ensure that HEEs are methodologically correct, identifiable, interpretable, and useful for decision-making, several elements need to be considered, accounted for, and reported. The aim of this paper was to introduce HEE in the context of CCT and to discuss key considerations when planning an HEE. A focus will be placed on key methodological concepts when preparing an HEE of CCT for depression. This may guide researchers during the design phase of future studies. Multimedia Appendix 2 [55-68] provides an overview of these concepts and their translation to the context of CCT.

Evaluating the Cost-Effectiveness of CCT

Overview

HEE focuses on the comparative analysis of ≥2 alternative courses (intervention vs comparator) of action in terms of both their costs and effects. In the context of CCT, CCT can be considered as the intervention, whereas the comparator may be a waiting list condition, TAU, or a condition in which participants complete an alternative training procedure, with or without TAU. Overall, 4 main types of full HEE exist: cost-minimization analysis, cost-benefit analysis,

cost-effectiveness analysis, and cost-utility analysis (CUA; for the definitions of these constructs, refer to Multimedia Appendix 1 [48]). As one of the most common HEEs used, CUA allows for health-related quality of life (HRQoL) adjustments to a given set of treatment outcomes, while simultaneously providing a generic outcome measure for comparison of costs and effects between different intervention strategies. Therefore, this paper specifically focused on this type of economic evaluation.

HEE—Health Effects and Costs

As previously mentioned, in an HEE, both costs and effects are considered. Main *cost categories* include direct and indirect costs. Direct costs are those directly associated with the disease or condition (in this case, depression) that is considered. They can be divided into direct medical costs (eg, psychologist visit and hospitalization) and direct nonmedical costs (eg, transportation costs). Indirect costs include those associated with productivity losses owing to the disease or condition that is considered (Multimedia Appendix 1 [49]).

The types of costs that are considered in an HEE depend on the *perspective of the analysis*. Every time an economic question is asked, it is crucial to carefully consider the analytic viewpoint. Commonly used perspectives include (1) patient perspective (only considering the costs for the patient), (2) health insurance perspective (considering the costs for the health insurer; ie, direct medical costs), (3) payer perspective (ie, the patient and the health insurer), and (4) societal perspective (considering direct costs and indirect costs). Importantly, decision makers must be informed about the viewpoint that has been taken in the HEE.

The *effects of the intervention* may be measured in natural units (eg, avoided recurrent episode of depression, avoided hospitalization, avoided complications, and avoided psychotherapy sessions) or in utilities. In a CUA, the effects are typically expressed in terms of quality-adjusted life years (QALYs; Multimedia Appendix 1). QALYs are a standardized measure reflecting the extent to which interventions improve HRQoL [69]. QALYs are calculated by multiplying a utility for a given condition (HRQoL weight) by the time an individual experiences the condition. These utilities are often derived from generic questionnaires such as the Short Form Health Survey (SF-36) or EuroQol–5 Dimension. The SF-36 is a 36-item patient-reported survey allowing to measure health status [55,56]. Similarly, the EuroQol–5 Dimension is a widely used measure of HRQoL, allowing to estimate utilities to calculate QALYs [57].

QALYs are a recognized metric for evaluating treatments, allowing the comparison of intervention alternatives for different disorders. However, at the same time, the suitability of QALYs to assess changes in mental health has been questioned [69,70]. That is, measurement tools used to estimate QALYs typically contain a limited focus on (specific aspects of) mental health [70]. Therefore, measures such as the SF-36 have been criticized in terms of the extent to which they allow to capture the heterogeneous nature of depression. As depression is a complex construct to capture [66], the key challenges to obtaining QALYs include describing this state and valuing it (eg, in terms of the morbidity and quality of life associated with living in this health state) [71]. In this context, authors have highlighted the need for new conceptualizations of the QALYs, with the goal of basing it on a valid and comprehensive model of quality of life specific to mental illness. This would require an instrument that is more sensitive to detect expected changes in this context [72,73].

To overcome these challenges and increase efficiency in measurement, it has been suggested to combine QALYs with other outcomes. As such, we would advise the reader to consider different kinds of measurements, allowing a multimethod evaluation of cost-effectiveness (refer to Multimedia Appendix 2 for an overview of relevant outcome measures in the context of CCT and how these can be operationalized). In this context, patient-reported outcomes (PRO) and PRO measures (PROMs) are increasingly used. PROs contain information from patients about their quality of life, their own health, and the functional status related to their treatment or the health care they receive. PROMs are the instruments or tools that are used to measure PROs [74]. Considering both QALYs and PROMs in HEE could facilitate decision-making, improve the quality of health care, and stimulate improvements in services, essentially by enabling the comparison of providers' performances [74,75].

HEE—Outcome

The outcome of a CUA is usually expressed as the incremental cost-utility ratio (ICUR; Multimedia Appendix 1). This is calculated by dividing the difference in costs (incremental costs) between the intervention and the comparator by the difference in their effects (incremental effects). The ICUR results in the following outcome: € per QALY. An ICUR as such does not allow to draw conclusions about whether an intervention can be considered as cost-effective.

To determine the cost-effectiveness of an intervention, a reference value above which the intervention is considered as not cost-effective can be used. Reference values can be based on a fixed amount threshold or calculated based on gross domestic product (GDP) per capita. This is the case in most low-income and middle-income countries. The World Health Organization promotes a threshold defined as 1 to 3 times the national annual GDP per capita as a guide for determining the cost-effectiveness of health interventions. We refer interested readers to the papers by Kazibwe et al [76] and Woods et al [77] for a more detailed discussion about this topic.

The reference values used typically differ among countries. For example, in France, no explicit threshold is applied [78]. This is similar to Belgium [48], but a threshold of 1 time the GDP per capita, representing a value of €35,000 (US $38,080) to €40,000 (US $43,520) per QALY, is often used in HEEs [79-81]. In the Netherlands, the threshold is up to €80,000 (US $87,039) per QALY [82], whereas in Spain, this is typically €30,000 (US $32,640) per QALY [83]. The National Institute for Health and Care Excellence considers a threshold ranging between £20,000 (US $25,488) and £30,000 (US $38,232) for England and Wales [84]. In the United States, a threshold between US $50,000 and US $100,000 is still often used by researchers, insurers, and public and private policy makers, but there is a debate about whether the threshold should be adjusted to inflation [85,86]. In India, researchers often use the World

Health Organization per capita GDP as a cost-effectiveness threshold [87].

HEE—Sources When Thinking About an HEE

Guidelines may be useful for researchers when planning an HEE. Well-known reporting guidelines include the Consolidated Health Economic Evaluation Reporting Standards statement that was recently updated, providing resources reflecting the most recent developments in HEE methods [88]. This checklist includes 28 items and is primarily intended to guide researchers when reporting economic evaluations for peer-reviewed journals and peer reviewers and editors assessing these manuscripts for publication. Taking such guidelines into account in the phase of planning an HEE study is likely to add to the quality of the design.

Another checklist that has been widely used in this context is the Consensus Health Economic Criteria checklist. It has been designed for assessment of the methodological quality of economic evaluations in systematic reviews [89]. These contain items that need to be included as good practice when reporting results of an HEE. Many countries have also developed specific national guidelines. For instance, Sharma et al [90] identified 31 national HEE guidelines, published between 1997 and August 2020.

In addition to the guidelines mentioned previously, published HEEs within the field of CCT and depression prevention also provide valuable sources of information. We advise the reader to search appropriate literature for what is already known regarding the cost-effectiveness of a given intervention, preferentially by referring to peer-reviewed and evidence-based sources. To the best of our knowledge, there are currently no studies examining the cost-effectiveness of CCT for depression. Although not specific for CCT, a recent meta-analysis suggests that digital health interventions are slightly more cost-effective than usual care [9]. Consistent with previous studies [45,91,92], guided digital health interventions outperformed unguided interventions [9]. A factor that may have contributed to this is that most unguided digital health interventions typically do not compensate for the lack of therapist support by including additional content or technological features (eg, to support adherence) [93]. In this context, it should be noted that studies exploring the effects of CCT on indicators of vulnerability to depression strongly differ in the extent to which they have been delivered as a therapist-assisted or unguided web-based intervention. This, in addition to other factors related to the delivery mode of CCT [40], remains as a parameter that requires further investigation and should be taken into account when considering HEE of CCT.

Overall, the current paucity of research on HEE of CCT for depression suggests the need for future CCT studies to include a focus on HEE. To advance such research in the context of interventions for depression, standardization of studies in terms of measurement and analytic approaches is paramount [9]. For this purpose, Multimedia Appendix 2 provides an overview of methodological questions to consider when setting up such studies. In addition, for an overview of the optimal procedures to perform HEE (ie, data analysis), including step-by-step guidelines about how to interpret the results of such HEE, we refer to Briggs et al [67].

Several challenges related to performing HEE should also be considered. For instance, HEE outcomes will always be context specific and country specific (eg, an intervention that is cost-effective in Belgium might not be cost-effective in Portugal). In addition, findings may also differ depending on the perspective taken. For instance, a recent meta-analysis found the cost-effectiveness of digital health interventions to be moderated by the economic perspective used [9]. Studies using a societal perspective yielded a relatively lower incremental net monetary benefit (Multimedia Appendix 1) than studies adopting a health care perspective, suggesting that the cost-effectiveness of internet interventions compared with control conditions cannot be assumed when maintaining a societal perspective [9].

Depending on the economic perspective taken, additional resources may be available. For instance, Powell and Torous [94] recently proposed a patient-centered method for estimating the economic value of clinical improvement following the use of digital health interventions. In particular, the authors suggest that an estimate of the economic value of clinical response to digital health interventions can be obtained by multiplying the country-specific willingness-to-pay (WTP) threshold per QALY, impact of the mental health condition in terms of QALYs, engagement level of digital health intervention users, effect size of the intervention (eg, in terms of percentage of improvement in symptoms), and duration of the impact of the intervention.

Extending the nation-specific example provided by Powell and Torous [94] to the context of CCT for depression, the economic value of aPASAT training for depression is estimated at US $555.61 per individual treated in the United States. That is, following Powell and Torous [94], for the United States, we rely on WTP threshold values of US $175,000 per QALY and 0.159 QALYs lost per year of depression. We combine these estimates with data obtained from a recent study during which aPASAT training was used as a web-based digital health intervention targeting repetitive negative thinking [95]. A total of 382 participants completed the baseline assessment of this study, of which 152 reported a level of depressive symptoms of at least moderate severity (≥14) based on the Depression Anxiety and Stress Scales [96]. For this example, we consider these scores as potentially indicative for the presence of a depressive episode. Overall, 97 (64%) of the participants with elevated depressive symptoms completed the web-based intervention. Of the latter group, posttraining and follow-up data were available for 90 individuals, suggesting a median reduction in level of depressive symptoms of 33% and 39% immediately following the intervention and at 1 month follow-up, respectively. Using these parameters, we estimate the economic value of aPASAT training for depression as follows: US $175,000 per QALY × 0.159 QALYs lost per year of depression × 64% completing the intervention × 39% reduction in symptoms × 0.08 years of improvement (at least until 1 month follow-up)=US $555.61. This value could be compared with values obtained for other interventions and costs related to the use of the intervention (eg, subscription fees to obtain access to the digital health intervention) or be used to

investigate the impact of country effects and context effects (eg, WTP threshold used).

Although informative to the patient, limitations of such methods include the imprecise nature of the estimation, owing to the input parameters used. For instance, in the example provided above, adherence and response rates were based on 1 study that relied on a convenience sample, in which evaluation of effects was limited to 1 month follow-up [95]. It is possible that the context in which CCT was provided offers an overestimation of adherence rates and, related to this, economical gains. Vice versa, previous research suggests that effects of CCT likely extend beyond the duration included in this example [35]. Other important limitations of this approach include the exclusive focus on the patient perspective and the lack of consideration of indirect benefits of the intervention (eg, increased productivity in a professional context). Such information can only be obtained from HEE studies designed to address the question of cost-effectiveness.

In this context, it is recommended for clinical trials in this area to systematically adopt measures that allow HEE based on input from various stakeholders. As CCT can be relatively inexpensive on a sufficiently large scale, we recommend researchers to cooperate with all the relevant stakeholders including patients, clinicians, health insurance companies, and policy makers. In addition, health economic researchers need to educate stakeholders about the relevance of cost-effectiveness analysis in policy and practice, how such analyses are performed, what uncertainties can be involved, and how the results should be interpreted. In addition to accurate communication with stakeholders, a clear view about CCT's cost-effectiveness is paramount to its clinical implementation.

HEE—Dealing With Uncertainty

It is worth noting that HEEs are frequently characterized by a certain degree of uncertainty or methodological considerations related to input parameters (eg, utilities and costs). This uncertainty can be addressed by performing 1-way sensitivity analyses and probabilistic sensitivity analyses (for more information, refer to Drummond et al [42]).

As noted previously, HEE serves evidence-based decision-making processes in health care. As such, we recommend that researchers include an HEE in their studies as a crucial step toward implementation. More precisely, information about the cost-effectiveness of CCT for depression will provide an extra piece of evidence that can be considered as being highly relevant for different stakeholders (eg, researchers, clinicians, and policy makers). However, the accuracy of such evidence is highly dependent on the quality of HEE methodological approaches used. This requires thorough reflection about methodological choices when planning future studies, which may be guided by the tools provided previously.

Conclusions

Previous studies suggest the potential of CCT to be used as a scalable and effective preventive intervention for recurrent depression. Although previous studies have mainly focused on the effects of CCT in terms of cognitive and emotional transfer, so far, little attention has been given to the cost-benefit ratio in this context. As such, there is a strong need for investigation of the cost-effectiveness of CCT. This paper provided guidelines about how future studies can address this gap in the literature as a means to influence health care policy, paving the way toward clinical implementation of CCT.

Acknowledgments

The authors would like to thank Ruben Willems, PhD, for providing feedback about the manuscript. This study was supported by an Applied Biomedical (Toegepast Biomedisch onderzoek met een primair Maatschappelijke finaliteit [TBM]) grant from the Research Foundation–Flanders (Fonds Wetenschappelijk Onderzoek Vlaanderen [FWO]), awarded to the Prevention of Depression (PrevenD) project 2.0 (T000720N). KH is funded by FWO (12J9722N).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key health economic constructs and definitions of basic concepts.
[PDF File (Adobe PDF File), 147 KB - mental_v10i1e44679_app1.pdf]

Multimedia Appendix 2
Key considerations when designing health economic studies in the context of cognitive control training for depression.
[PDF File (Adobe PDF File), 185 KB - mental_v10i1e44679_app2.pdf]

References

1. Depressive disorder (depression). World Health Organization. 2023 Mar. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2023-05-08]

2. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. JAMA Psychiatry 2018 Apr 01;75(4):336-346 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.4602] [Medline: 29450462]
3. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
4. Kessler RC. The costs of depression. Psychiatr Clin North Am 2012 Mar;35(1):1-14 [FREE Full text] [doi: 10.1016/j.psc.2011.11.005] [Medline: 22370487]
5. Kessler RC, Akiskal HS, Ames M, Birnbaum H, Greenberg P, Hirschfeld RM, et al. Prevalence and effects of mood disorders on work performance in a nationally representative sample of U.S. workers. Am J Psychiatry 2006 Sep;163(9):1561-1568 [FREE Full text] [doi: 10.1176/ajp.2006.163.9.1561] [Medline: 16946181]
6. Kim J, Aryee LM, Bang H, Prajogo S, Choi YK, Hoch JS, et al. Effectiveness of digital mental health tools to reduce depressive and anxiety symptoms in low- and middle-income countries: systematic review and meta-analysis. JMIR Ment Health 2023 Mar 20;10:e43066 [FREE Full text] [doi: 10.2196/43066] [Medline: 36939820]
7. Koster EH, Hoorelbeke K, Onraedt T, Owens M, Derakshan N. Cognitive control interventions for depression: a systematic review of findings from training studies. Clin Psychol Rev 2017 Apr;53:79-92. [doi: 10.1016/j.cpr.2017.02.002] [Medline: 28273486]
8. Siegle GJ, Ghinassi F, Thase ME. Neurobehavioral therapies in the 21st century: summary of an emerging field and an extended example of cognitive control training for depression. Cognit Ther Res 2007 Mar 23;31(2):235-262 [FREE Full text] [doi: 10.1007/s10608-006-9118-6]
9. Rohrbach PJ, Dingemans AE, Evers C, Van Furth EF, Spinhoven P, Aardoom JJ, et al. Cost-effectiveness of internet interventions compared with treatment as usual for people with mental disorders: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2023 Jan 05;25:e38204 [FREE Full text] [doi: 10.2196/38204] [Medline: 36602854]
10. Cohen JD. Cognitive control: core constructs and current considerations. In: Egner T, editor. The Wiley Handbook of Cognitive Control. Chichester, UK: John Wiley & Sons; 2017:1-28.
11. Joormann J, Stanton CH. Examining emotion regulation in depression: a review and future directions. Behav Res Ther 2016 Nov;86:35-49. [doi: 10.1016/j.brat.2016.07.007] [Medline: 27492851]
12. Levens SM, Gotlib IH. Updating positive and negative stimuli in working memory in depression. J Exp Psychol Gen 2010 Nov;139(4):654-664 [FREE Full text] [doi: 10.1037/a0020283] [Medline: 21038984]
13. Levens SM, Gotlib IH. Updating emotional content in recovered depressed individuals: evaluating deficits in emotion processing following a depressive episode. J Behav Ther Exp Psychiatry 2015 Sep;48:156-163 [FREE Full text] [doi: 10.1016/j.jbtep.2015.03.009] [Medline: 25889375]
14. De Raedt R, Koster EH. Understanding vulnerability for depression from a cognitive neuroscience perspective: a reappraisal of attentional factors and a new conceptual framework. Cogn Affect Behav Neurosci 2010 Mar;10(1):50-70. [doi: 10.3758/CABN.10.1.50] [Medline: 20233955]
15. Knight MJ, Air T, Baune BT. The role of cognitive impairment in psychosocial functioning in remitted depression. J Affect Disord 2018 Aug 01;235:129-134. [doi: 10.1016/j.jad.2018.04.051] [Medline: 29655074]
16. Clark M, DiBenedetti D, Perez V. Cognitive dysfunction and work productivity in major depressive disorder. Expert Rev Pharmacoecon Outcomes Res 2016 Aug;16(4):455-463. [doi: 10.1080/14737167.2016.1195688] [Medline: 27268275]
17. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. Perspect Psychol Sci 2008 Sep;3(5):400-424. [doi: 10.1111/j.1745-6924.2008.00088.x] [Medline: 26158958]
18. Watkins ER. Constructive and unconstructive repetitive thought. Psychol Bull 2008 Mar;134(2):163-206 [FREE Full text] [doi: 10.1037/0033-2909.134.2.163] [Medline: 18298268]
19. Watkins E. Psychological treatment of depressive rumination. Curr Opin Psychol 2015 Aug;4:32-36 [FREE Full text] [doi: 10.1016/j.copsyc.2015.01.020]
20. Jones NP, Siegle GJ, Thase ME. Effects of rumination and initial severity on remission to cognitive therapy for depression. Cognit Ther Res 2008 Aug 01;32(4):591-604 [FREE Full text] [doi: 10.1007/s10608-008-9191-0] [Medline: 24353355]
21. Koster EH, De Lissnyder E, Derakshan N, De Raedt R. Understanding depressive rumination from a cognitive science perspective: the impaired disengagement hypothesis. Clin Psychol Rev 2011 Feb;31(1):138-145. [doi: 10.1016/j.cpr.2010.08.005] [Medline: 20817334]
22. Watkins ER, Roberts H. Reflecting on rumination: consequences, causes, mechanisms and treatment of rumination. Behav Res Ther 2020 Apr;127:103573. [doi: 10.1016/j.brat.2020.103573] [Medline: 32087393]
23. Demeyer I, De Lissnyder E, Koster EH, De Raedt R. Rumination mediates the relationship between impaired cognitive control for emotional information and depressive symptoms: a prospective study in remitted depressed adults. Behav Res Ther 2012 May;50(5):292-297 [FREE Full text] [doi: 10.1016/j.brat.2012.02.012] [Medline: 22449892]
24. Vanderhasselt MA, De Raedt R. Impairments in cognitive control persist during remission from depression and are related to the number of past episodes: an event related potentials study. Biol Psychol 2009 Jul;81(3):169-176. [doi: 10.1016/j.biopsycho.2009.03.009] [Medline: 19482234]

25. Launder N, Minkov R, Davey C, Finke C, Malmberg GH, Lampit A. Computerized cognitive training in people with depression: a systematic review and meta-analysis of randomized clinical trials. medRxiv. Preprint posted online March 26, 2021 [FREE Full text] [doi: 10.1101/2021.03.23.21254003]
26. Motter JN, Pimontel MA, Rindskopf D, Devanand DP, Doraiswamy PM, Sneed JR. Computerized cognitive training and functional recovery in major depressive disorder: a meta-analysis. J Affect Disord 2016 Jan 01;189:184-191. [doi: 10.1016/j.jad.2015.09.022] [Medline: 26437233]
27. Legemaat AM, Semkovska M, Brouwer M, Geurtsen GJ, Burger H, Denys D, et al. Effectiveness of cognitive remediation in depression: a meta-analysis. Psychol Med 2021 Apr 14;52(16):1-16 [FREE Full text] [doi: 10.1017/S0033291721001100] [Medline: 33849674]
28. Siegle GJ, Price RB, Jones NP, Ghinassi F, Painter T, Thase ME. You gotta work at it: pupillary indices of task focus are prognostic for response to a neurocognitive intervention for rumination in depression. Clin Psychol Sci 2014 Jul;2(4):455-471 [FREE Full text] [doi: 10.1177/2167702614536160]
29. Brunoni AR, Boggio PS, De Raedt R, Benseñor IM, Lotufo PA, Namur V, et al. Cognitive control therapy and transcranial direct current stimulation for depression: a randomized, double-blinded, controlled trial. J Affect Disord 2014 Jun;162:43-49 [FREE Full text] [doi: 10.1016/j.jad.2014.03.026] [Medline: 24767004]
30. Iacoviello BM, Wu G, Alvarez E, Huryk K, Collins KA, Murrough JW, et al. Cognitive-emotional training as an intervention for major depressive disorder. Depress Anxiety 2014 Aug;31(8):699-706. [doi: 10.1002/da.22266] [Medline: 24753225]
31. Morimoto SS, Altizer RA, Gunning FM, Hu W, Liu J, Cote SE, et al. Targeting cognitive control deficits with neuroplasticity-based computerized cognitive remediation in patients with geriatric major depression: a randomized, double-blind, controlled trial. Am J Geriatr Psychiatry 2020 Sep;28(9):971-980. [doi: 10.1016/j.jagp.2020.05.023] [Medline: 32591170]
32. Sommer A, Plewnia C. Depression treatment by tDCS-enhanced cognitive control training: a test of two stimulation intensities. Brain Stimul 2021 Jul;14(4):987-989 [FREE Full text] [doi: 10.1016/j.brs.2021.06.004] [Medline: 34147374]
33. Ferrari GR, Vanderhasselt M, Rinck M, Demeyer I, De Raedt R, Beisel S, et al. A cognitive control training as add-on treatment to usual care for depressed inpatients. Cogn Ther Res 2021 Feb 27;45(5):929-943 [FREE Full text] [doi: 10.1007/s10608-020-10197-y]
34. Moshier SJ, Otto MW. Behavioral activation treatment for major depression: a randomized trial of the efficacy of augmentation with cognitive control training. J Affect Disord 2017 Mar 01;210:265-268. [doi: 10.1016/j.jad.2017.01.003] [Medline: 28068613]
35. Hoorelbeke K, Koster EH. Internet-delivered cognitive control training as a preventive intervention for remitted depressed patients: evidence from a double-blind randomized controlled trial study. J Consult Clin Psychol 2017 Feb;85(2):135-146. [doi: 10.1037/ccp0000128] [Medline: 27362792]
36. Hoorelbeke K, Van den Bergh N, De Raedt R, Wichers M, Albers CJ, Koster EH. Regaining control of your emotions? Investigating the mechanisms underlying effects of cognitive control training for remitted depressed patients. Emotion 2023 Feb;23(1):194-213. [doi: 10.1037/emo0001067] [Medline: 35175068]
37. Hoorelbeke K, Van den Bergh N, De Raedt R, Wichers M, Koster EH. Preventing recurrence of depression: long-term effects of a randomized controlled trial on cognitive control training for remitted depressed patients. Clin Psychol Sci 2021 Jul;9(4):615-633 [FREE Full text] [doi: 10.1177/2167702620979775]
38. Vander Zwalmen Y, Liebaert E, Hoorelbeke K, de Mévergnies CN, Baeken C, Verhaeghe N, et al. Treatment response following adaptive PASAT training for depression vulnerability: a systematic review and meta-analysis. Neuropsychol Rev (Forthcoming) 2023 Feb 28. [doi: 10.1007/s11065-023-09581-8] [Medline: 36853526]
39. Cuijpers P, Turner EH, Mohr DC, Hofmann SG, Andersson G, Berking M, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. Psychol Med 2014 Mar;44(4):685-695. [doi: 10.1017/S0033291713000457] [Medline: 23552610]
40. Vander Zwalmen Y, Hoorelbeke K, Liebaert E, Nève de Mévergnies C, Koster EH. Cognitive remediation for depression vulnerability: Current challenges and new directions. Front Psychol 2022 Jul 22;13:903446 [FREE Full text] [doi: 10.3389/fpsyg.2022.903446] [Medline: 35936259]
41. Littlejohns P, Weale A, Chalkidou K, Faden R, Teerawattananon Y. Social values and health policy: a new international research programme. J Health Organ Manag 2012;26(3):285-292. [doi: 10.1108/14777261211238945] [Medline: 22852452]
42. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 4th edition. New York, NY: Oxford University Press; 2015.
43. Sander L, Rausch L, Baumeister H. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016 Aug 17;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
44. Välimäki M, Anttila K, Anttila M, Lahti M. Web-based interventions supporting adolescents and young people with depressive symptoms: systematic review and meta-analysis. JMIR Mhealth Uhealth 2017 Dec 08;5(12):e180 [FREE Full text] [doi: 10.2196/mhealth.8624] [Medline: 29222079]

45. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
46. Glossary of frequently encountered terms in health economics. National Information Center on Health Services Research and Health Care Technology. URL: https://wayback.archive-it.org/org-350/20210820174512/https://www.nlm.nih.gov/nichsr/edu/healthecon/glossary.html [accessed 2023-05-08]
47. A glossary of health economic terms. York Health Economics Consortium. 2016. URL: https://yhec.co.uk/resources/glossary/ [accessed 2023-03-21]
48. Cleemput I, Neyt M, Van de Sande S, Thiry N. Belgian guidelines for economic evaluations and budget impact analyses: second edition. Belgian Health Care Knowledge Centre. URL: https://kce.fgov.be/sites/default/files/2021-11/KCE_183_economic_evaluations_second_edition_Report_update.pdf
49. Rice DP. Cost of illness studies: what is good about them? Inj Prev 2000 Sep;6(3):177-179 [FREE Full text] [doi: 10.1136/ip.6.3.177] [Medline: 11003181]
50. Health-related quality of life (HRQOL). Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/hrqol/index.htm [accessed 2023-04-10]
51. Kim DD, Wilkinson CL, Pope EF, Chambers JD, Cohen JT, Neumann PJ. The influence of time horizon on results of cost-effectiveness analyses. Expert Rev Pharmacoecon Outcomes Res 2017 Dec;17(6):615-623. [doi: 10.1080/14737167.2017.1331432] [Medline: 28504026]
52. Buntrock C, Berking M, Smit F, Lehr D, Nobis S, Riper H, et al. Preventing depression in adults with subthreshold depression: health-economic evaluation alongside a pragmatic randomized controlled trial of a web-based intervention. J Med Internet Res 2017 Jan 04;19(1):e5 [FREE Full text] [doi: 10.2196/jmir.6587] [Medline: 28052841]
53. Kolovos S, Kenter RM, Bosmans JE, Beekman AT, Cuijpers P, Kok RN, et al. Economic evaluation of internet-based problem-solving guided self-help treatment in comparison with enhanced usual care for depressed outpatients waiting for face-to-face treatment: a randomized controlled trial. J Affect Disord 2016 Aug;200:284-292. [doi: 10.1016/j.jad.2016.04.025] [Medline: 27155071]
54. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010 Dec 19;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
55. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
56. Ware JE. SF-36 health survey. In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment. 2nd edition. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 1999:1227-1246.
57. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol group. Ann Med 2001 Jul;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]
58. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics 1993 Nov;4(5):353-365. [doi: 10.2165/00019053-199304050-00006] [Medline: 10146874]
59. Manual iMTA Medical Cost Questionnaire (iMCQ). institute for Medical Technology Assessment. 2018. URL: https://www.imta.nl/questionnaires/ [accessed 2023-07-10]
60. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-57. [Medline: 9881538]
61. Zimmerman M, Martinez JH, Attiullah N, Friedman M, Toba C, Boerescu DA, et al. A new type of scale for determining remission from depression: the remission from depression questionnaire. J Psychiatr Res 2013 Jan;47(1):78-82. [doi: 10.1016/j.jpsychires.2012.09.006] [Medline: 23102820]
62. Beck AT, Steer RA, Brown G. Beck depression inventory–II. APA PsycTests. 1996. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft00742-000 [accessed 2023-05-08]
63. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
64. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): validation of a content-independent measure of repetitive negative thinking. J Behav Ther Exp Psychiatry 2011 Jun;42(2):225-232 [FREE Full text] [doi: 10.1016/j.jbtep.2010.12.003] [Medline: 21315886]
65. Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta Earthquake. J Pers Soc Psychol 1991 Jul;61(1):115-121. [doi: 10.1037//0022-3514.61.1.115] [Medline: 1890582]
66. Fried EI, Flake JK, Robinaugh DJ. Revisiting the theoretical and methodological foundations of depression measurement. Nat Rev Psychol 2022 Apr 14;1(6):358-368 [FREE Full text] [doi: 10.1038/s44159-022-00050-2]
67. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press; 2006.

68. Lokkerbol J, Wijnen B, Ruhe HG, Spijker J, Morad A, Schoevers R, et al. Design of a health-economic Markov model to assess cost-effectiveness and budget impact of the prevention and treatment of depressive disorder. Expert Rev Pharmacoecon Outcomes Res 2021 Oct;21(5):1031-1042 [FREE Full text] [doi: 10.1080/14737167.2021.1844566] [Medline: 33119427]
69. Prieto L, Sacristán JA. Problems and solutions in calculating quality-adjusted life years (QALYs). Health Qual Life Outcomes 2003 Dec 19;1:80 [FREE Full text] [doi: 10.1186/1477-7525-1-80] [Medline: 14687421]
70. Knapp M, Mangalore R. "The trouble with QALYs...". Epidemiol Psichiatr Soc 2007 Oct;16(4):289-293. [doi: 10.1017/s1121189x00002451] [Medline: 18333423]
71. Luyten J, Naci H, Knapp M. Economic evaluation of mental health interventions: an introduction to cost-utility analysis. Evid Based Ment Health 2016 May;19(2):49-53. [doi: 10.1136/eb-2016-102354] [Medline: 27075444]
72. Brazier J. Measuring and valuing mental health for use in economic evaluation. J Health Serv Res Policy 2008 Oct;13 Suppl 3:70-75. [doi: 10.1258/jhsrp.2008.008015] [Medline: 18806195]
73. Brazier J. Is the EQ-5D fit for purpose in mental health? Br J Psychiatry 2010 Nov;197(5):348-349. [doi: 10.1192/bjp.bp.110.082453] [Medline: 21037210]
74. Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013 Jan 28;346:f167 [FREE Full text] [doi: 10.1136/bmj.f167] [Medline: 23358487]
75. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ 2015 Feb 10;350:g7818. [doi: 10.1136/bmj.g7818] [Medline: 25670183]
76. Kazibwe J, Gheorghe A, Wilson D, Ruiz F, Chalkidou K, Chi YL. The use of cost-effectiveness thresholds for evaluating health interventions in low- and middle-income countries from 2015 to 2020: a review. Value Health 2022 Mar;25(3):385-389 [FREE Full text] [doi: 10.1016/j.jval.2021.08.014] [Medline: 35227450]
77. Woods B, Revill P, Sculpher M, Claxton K. Country-level cost-effectiveness thresholds: initial estimates and the need for further research. Value Health 2016 Dec;19(8):929-935 [FREE Full text] [doi: 10.1016/j.jval.2016.02.017] [Medline: 27987642]
78. Massetti M, Aballéa S, Videau Y, Rémuzat C, Roïz J, Toumi M. A comparison of HAS and NICE guidelines for the economic evaluation of health technologies in the context of their respective national health care systems and cultural environments. J Mark Access Health Policy 2015 Mar 12;3:24966 [FREE Full text] [doi: 10.3402/jmahp.v3.24966] [Medline: 27123190]
79. Vandepitte S, Putman K, Van Den Noortgate N, Verhaeghe N, Annemans L. Cost-effectiveness of an in-home respite care program to support informal caregivers of persons with dementia: a model-based analysis. Int J Geriatr Psychiatry 2020 Jun;35(6):601-609. [doi: 10.1002/gps.5276] [Medline: 32011773]
80. Werbrouck A, Schmidt M, Putman K, Annemans L, Benhalima K, Simoens S, et al. Cost-utility analysis of lifestyle interventions to prevent type 2 diabetes in women with prior gestational diabetes. Eur J Public Health 2020 Jun 01;30(3):473-478. [doi: 10.1093/eurpub/ckz196] [Medline: 31665267]
81. Pil L, Hoorens I, Vossaert K, Kruse V, Tromme I, Speybroeck N, et al. Cost-effectiveness and budget effect analysis of a population-based skin cancer screening. JAMA Dermatol 2017 Feb 01;153(2):147-153. [doi: 10.1001/jamadermatol.2016.4518] [Medline: 28002834]
82. Schurer M, Matthijsse SM, Vossen CY, van Keep M, Horscroft J, Chapman AM, et al. Varying willingness to pay based on severity of illness: impact on health technology assessment outcomes of inpatient and outpatient drug therapies in The Netherlands. Value Health 2022 Jan;25(1):91-103 [FREE Full text] [doi: 10.1016/j.jval.2021.08.003] [Medline: 35031104]
83. Piera-Jiménez J, Etzelmueller A, Kolovos S, Folkvord F, Lupiáñez-Villanueva F. Guided internet-based cognitive behavioral therapy for depression: implementation cost-effectiveness study. J Med Internet Res 2021 May 11;23(5):e27410 [FREE Full text] [doi: 10.2196/27410] [Medline: 33973857]
84. Updated guide to the methods of technology appraisal. National Institute for Health and Clinical Excellence. 2013. URL: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 [accessed 2023-05-08]
85. Menzel P. How should willingness-to-pay values of quality-adjusted life-years be updated and according to whom? AMA J Ethics 2021 Aug 01;23(8):E601-E606 [FREE Full text] [doi: 10.1001/amajethics.2021.601] [Medline: 34459726]
86. Ubel PA, Hirth RA, Chernew ME, Fendrick AM. What is the price of life and why doesn't it increase at the rate of inflation? Arch Intern Med 2003 Jul 28;163(14):1637-1641. [doi: 10.1001/archinte.163.14.1637] [Medline: 12885677]
87. Prinja S, Sundararaman T, Muraleedharan V. Cost-effectiveness threshold and health opportunity cost achieving universal health coverage. Economic & Political Weekly EPW. 2020 Jan. URL: https://hss.iitm.ac.in/ctap/wp-content/uploads/2022/05/CM_LV_1_110120_Shankar_Prinja.pdf [accessed 2023-05-08]
88. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, CHEERS 2022 ISPOR Good Research Practices Task Force. Consolidated health economic evaluation reporting standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Int J Technol Assess Health Care 2022 Jan 11;38(1):e13. [doi: 10.1017/S0266462321001732] [Medline: 35007499]
89. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. Int J Technol Assess Health Care 2005;21(2):240-245 [FREE Full text] [Medline: 15921065]

90. Sharma D, Aggarwal AK, Downey LE, Prinja S. National healthcare economic evaluation guidelines: a cross-country comparison. Pharmacoecon Open 2021 Sep;5(3):349-364 [FREE Full text] [doi: 10.1007/s41669-020-00250-7] [Medline: 33423205]
91. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
92. Kählke F, Buntrock C, Smit F, Ebert DD. Systematic review of economic evaluations for internet- and mobile-based interventions for mental health problems. NPJ Digit Med 2022 Nov 23;5(1):175 [FREE Full text] [doi: 10.1038/s41746-022-00702-w] [Medline: 36424463]
93. Burger F, Neerincx MA, Brinkman WP. Technological state of the art of electronic mental health interventions for major depressive disorder: systematic literature review. J Med Internet Res 2020 Jan 20;22(1):e12599 [FREE Full text] [doi: 10.2196/12599] [Medline: 31958063]
94. Powell A, Torous J. A patient-centered framework for measuring the economic value of the clinical benefits of digital health apps: theoretical modeling. JMIR Ment Health 2020 Oct 30;7(10):e18812 [FREE Full text] [doi: 10.2196/18812] [Medline: 33124995]
95. Hoorelbeke K, Vervaeke J, Siegle GJ, Baeken C, Koster EH. Individual differences associated with treatment adherence and transfer effects following gamified web-based cognitive control training for repetitive negative thinking. Internet Interv 2022 Feb 12;27:100507 [FREE Full text] [doi: 10.1016/j.invent.2022.100507] [Medline: 35242588]
96. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the beck depression and anxiety inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]

Abbreviations

aPASAT: adaptive paced auditory serial addition task
CCT: cognitive control training
CUA: cost-utility analysis
GDP: gross domestic product
HEE: health economic evaluation
HRQoL: health-related quality of life
ICUR: incremental cost-utility ratio
MDD: major depressive disorder
NNT: number needed to treat
PRO: patient-reported outcome
PROM: patient-reported outcome measure
QALY: quality-adjusted life year
SF-36: Short Form Health Survey
TAU: treatment as usual
WTP: willingness-to-pay

Edited by J Torous; submitted 29.11.22; peer-reviewed by A Peckham, J Torous; comments to author 04.03.23; revised version received 22.05.23; accepted 10.06.23; published 18.08.23.

Please cite as:
Nève de Mévergnies C, Verhaeghe N, Koster EHW, Baeken C, Vander Zwalmen Y, Hoorelbeke K
Health Economic Evaluation of Cognitive Control Training for Depression: Key Considerations
JMIR Ment Health 2023;10:e44679
URL: https://mental.jmir.org/2023/1/e44679
doi:10.2196/44679
PMID:37594847

Viewpoint

The Potential Influence of AI on Population Mental Health

Catherine K Ettman[1], PhD; Sandro Galea[2], MD, MPH, DrPH

[1]Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
[2]Office of the Dean, Boston University School of Public Health, Boston, MA, United States

Corresponding Author:
Catherine K Ettman, PhD
Department of Health Policy and Management
Johns Hopkins Bloomberg School of Public Health
624 N Broadway Street
Baltimore, MD, 21205
United States
Phone: 1 410 516 8000
Fax: 1 410 955 4775
Email: cettman1@jhu.edu

Abstract

The integration of artificial intelligence (AI) into everyday life has galvanized a global conversation on the possibilities and perils of AI on human health. In particular, there is a growing need to anticipate and address the potential impact of widely accessible, enhanced, and conversational AI on mental health. We propose 3 considerations to frame how AI may influence population mental health: through the advancement of mental health care; by altering social and economic contexts; and through the policies that shape the adoption, use, and potential abuse of AI-enhanced tools.

(JMIR Ment Health 2023;10:e49936) doi:10.2196/49936

KEYWORDS

mental health; artificial intelligence; AI; policy; policies; population health; population; ChatGPT; generative; tools; digital mental health

Introduction

The widespread incorporation of artificial intelligence (AI) in daily use has sparked a global dialogue about the potential benefits and risks of AI on human well-being. Specifically, there is an increasing urgency to anticipate and address the potential impact of widely accessible, enhanced, and conversational AI on mental health. We propose 3 points to consider when determining how AI may influence population mental health: through the advancement of mental health care; by altering social and economic contexts; and through the policies that shape the adoption, use, and potential abuse of AI-enhanced tools (Figure 1).

Figure 1. Influence of artificial intelligence on population mental health.

Prevention, Screening, and Treatment of Mental Health Disorders

With over 970 million people living with a mental disorder worldwide [1], as well as a shortage of accessible care for many people, leveraging tools such as artificial intelligence (AI) could influence mental health through prevention and treatment. AI-enabled tools can prevent more severe mental illness from developing by identifying higher-risk populations that lead to quicker intervention. AI can detect, assess, and predict stress [2]. For example, AI can process natural language from electronic health records to detect early cognitive impairment [3] or child maltreatment [4], which can have effects on mental health across the course of one's life.

In addition to preventing mental health challenges through more effective and rapid screening, AI has the potential to improve access to mental health care [5]. One could imagine a world where AI serves as the "front line" for mental health, providing a clearinghouse of resources and available services for individuals seeking help. In addition, targeted interventions delivered digitally can help reduce the population burden of mental illness, particularly in hard-to-reach populations and contexts, for example, through stepped care approaches that aim to help populations with the highest risk following natural disasters.

While AI has promise in terms of early identification of risk and in triaging and treating large volumes of patients, significant flaws exist in using AI for this purpose [6], including bias that may lead to inaccurate assessment and perpetuation of stereotypes. AI efforts to improve risk prediction thus far have been met with mixed results, such as suicide risk prediction by AI being no better than simpler models [7]. The recent improvements in AI technology, however, suggest that as AI improves, it can rapidly become more useful to identify risk for personalized interventions [5]. While some efforts are attempting to leverage AI to deliver mental health care, such as in the form of responsive chatbots, there remains a gulf between vision and implementation—as well as understanding the long-term consequences of replacing human compassion, judgment, and experience with AI-generated responses.

Social and Economic Contexts That Shape Mental Health

More foundationally, AI may shift or exacerbate differences in the distribution of assets, which serve as a buffer against mental health challenges. Mental health is sensitive to economic and social contexts. First, it is possible that AI may transform or modify existing economic contexts, such as distributions of wealth and employment, which both protect mental health. Unemployment is associated with adverse mental health outcomes long after initial job loss occurs [8]. Potential loss of jobs that may follow AI replacement of specific tasks and industries could lead to psychological sequelae, particularly among workers more vulnerable to job loss, borne disproportionately by populations with fewer assets [7]. In this way, AI could widen existing economic gaps between groups and exacerbate mental health inequities [9,10], thereby fulfilling cumulative inequality theory [11]. Alternatively, AI may benefit mental health through the creation of new entrepreneurial opportunities and access to capital previously unavailable.

Second, the use of AI and generative AI in particular, with human-like responses, may shift how people interact with each other. Meaningful social connections and social support serve as protective mechanisms against diminished health, and AI may shift how people interact with each other. AI may lead to greater polarization and extremism as users consume curated information and may lead to further breakdown of social networks [12] and ties that bond and protect mental health.

Policy, Regulation, and Guardrails

The policy environment we live in, along with the values that drive our policies, will inform how AI can influence mental health. AI may create opportunities to rapidly synthesize seemingly unlimited information about individuals; if used maliciously, these tools can cause harm to the health of populations. Three considerations, therefore, will be important in this area as we consider how AI may influence population mental health.

First, policies, standards, and regulations should consider how to safeguard sensitive patient information and individuals' privacy. Given rapidly evolving technology, services, and functions, regulation has not yet kept up with the potential use and misuse of targeted data. Particularly in the case of sharing sensitive mental health data, it will be important to ensure that patients are protected from exposure to malefactors who can exploit their mental health status. While the Health Insurance Portability and Accountability Act (HIPAA) protects digital patient health information in certain settings, it does not extend to new health ecosystems such as the medical internet of things [13] and mobile health (mHealth) applications that collect copious data about individuals and their environments. As the landscape of mental health care and well-being evolves, policies to protect privacy will need to evolve. While there may be benefits to highly accurate data, such as faster arrival of support following suicide and crisis lifeline calls [14], costs include lack of patient privacy and potential abuse by bad actors.

Second, alignment on values and implementation of policy to reduce the influence of bias in AI will be critical to ensure that existing gaps are not exacerbated and that groups are not targeted, mistreated, or maligned intentionally or unintentionally. A growing awareness of the importance of algorithmic fairness has prompted discussion on the appropriate use of AI and machine learning; in the absence of thoughtful intervention, existing algorithms could perpetuate bias and heighten health disparities across groups [15]. Given a history of stigma around mental health in particular, alignment by stakeholders across sectors on the values and sensitivities of using AI broadly will be needed to prevent the exacerbation of stigma and mental health disparities.

Third, guardrails around AI-generated responses can prevent harm. Suicide attempts are more successful when the means used are more lethal; it is possible that users could leverage AI to learn more quickly about self-harm or harming others. Ensuring that AI has built-in guardrails to prevent the proliferation of lethal means and to instead leverage resources to create a pathway to treatment may help to prevent unfavorable outcomes of AI-human engagement.

Conclusion

While AI may pose potential risks and benefits to human mental health, the mechanism by which they occur is through the real world. Mental health and physical health are experienced in real life. Perhaps the best way to prepare for the oncoming changes that new tools will bring will be to ensure that even as we develop new digital tools, we continue to invest in the basic infrastructure, assets, and social connections that we know protect mental health—and make human life worth living.

Conflicts of Interest

None declared.

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psych 2022 Feb;9(2):137-150 [FREE Full text] [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]
2. Mentis AA, Lee D, Roussos P. Applications of artificial intelligence-machine learning for detection of stress: a critical overview. Mol Psychiatry 2023 Apr 05:1-13. [doi: 10.1038/s41380-023-02047-6] [Medline: 37020048]
3. Penfold RB, Carrell DS, Cronkite DJ, Pabiniak C, Dodd T, Glass AM, et al. Development of a machine learning model to predict mild cognitive impairment using natural language processing in the absence of screening. BMC Med Inform Decis Mak 2022 May 12;22(1):129 [FREE Full text] [doi: 10.1186/s12911-022-01864-z] [Medline: 35549702]
4. Negriff S, Lynch FL, Cronkite DJ, Pardee RE, Penfold RB. Using natural language processing to identify child maltreatment in health systems. Child Abuse Negl 2023 Apr;138:106090. [doi: 10.1016/j.chiabu.2023.106090] [Medline: 36758373]
5. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim H, et al. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep 2019 Nov 07;21(11):116 [FREE Full text] [doi: 10.1007/s11920-019-1094-0] [Medline: 31701320]
6. Tornero-Costa R, Martinez-Millana A, Azzopardi-Muscat N, Lazeri L, Traver V, Novillo-Ortiz D. Methodological and quality flaws in the use of artificial intelligence in mental health research: systematic review. JMIR Ment Health 2023 Feb 02;10:e42045 [FREE Full text] [doi: 10.2196/42045] [Medline: 36729567]
7. Shortreed SM, Walker RL, Johnson E, Wellman R, Cruz M, Ziebell R, et al. Complex modeling with detailed temporal predictors does not improve health records-based suicide risk prediction. NPJ Digit Med 2023 Mar 23;6(1):47 [FREE Full text] [doi: 10.1038/s41746-023-00772-4] [Medline: 36959268]
8. Gallo WT, Bradley EH, Dubin JA, Jones RN, Falba TA, Teng H, et al. The persistence of depressive symptoms in older workers who experience involuntary job loss: results from the health and retirement survey. J Gerontol B Psychol Sci Soc Sci 2006 Jul 01;61(4):S221-S228 [FREE Full text] [doi: 10.1093/geronb/61.4.s221] [Medline: 16855043]

9. Ettman CK, Adam GP, Clark MA, Wilson IB, Vivier PM, Galea S. Wealth and depression: a scoping review. Brain Behav 2022 Mar 08;12(3):e2486 [FREE Full text] [doi: 10.1002/brb3.2486] [Medline: 35134277]
10. Ettman CK, Fan AY, Philips AP, Adam GP, Ringlein G, Clark MA, et al. Financial strain and depression in the U.S.: a scoping review. Transl Psychiatry 2023 May 13;13(1):168 [FREE Full text] [doi: 10.1038/s41398-023-02460-z] [Medline: 37179345]
11. Ferraro KF, Shippee TP. Aging and cumulative inequality: how does inequality get under the skin? Gerontologist 2009 Jun 17;49(3):333-343 [FREE Full text] [doi: 10.1093/geront/gnp034] [Medline: 19377044]
12. Santos FP, Lelkes Y, Levin SA. Link recommendation algorithms and dynamics of polarization in online social networks. Proc Natl Acad Sci U S A 2021 Dec 14;118(50):e2102141118 [FREE Full text] [doi: 10.1073/pnas.2102141118] [Medline: 34876508]
13. Theodos K, Sittig S. Health information privacy laws in the digital age: HIPAA doesn't apply. Perspect Health Inf Manag 2021;18(Winter):1l [FREE Full text] [Medline: 33633522]
14. Purtle J, Chance Ortego J, Bandara S, Goldstein A, Pantalone J, Goldman ML. Implementation of the 988 Suicide & Crisis Lifeline: estimating state-level increases in call demand costs and financing. J Ment Health Policy Econ 2023 Jun 01;26(2):85-95. [Medline: 37357873]
15. Agarwal R, Bjarnadottir M, Rhue L, Dugas M, Crowley K, Clark J, et al. Addressing algorithmic bias and the perpetuation of health inequities: an AI bias aware framework. Health Policy Technol 2023 Mar;12(1):100702. [doi: 10.1016/j.hlpt.2022.100702]

Abbreviations

AI: artificial intelligence
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health

Edited by J Torous; submitted 14.06.23; peer-reviewed by M del Pozo Banos, A Bachu; comments to author 20.08.23; revised version received 09.10.23; accepted 29.10.23; published 16.11.23.

Please cite as:
Ettman CK, Galea S
The Potential Influence of AI on Population Mental Health
JMIR Ment Health 2023;10:e49936
URL: https://mental.jmir.org/2023/1/e49936
doi:10.2196/49936
PMID:37971803

Original Paper

Understanding the Adoption and Use of Digital Mental Health Apps Among College Students: Secondary Analysis of a National Survey

Justine Bautista[1], MA; Stephen M Schueller[1,2,3], PhD

[1]School of Social Ecology, University of California, Irvine, Irvine, CA, United States

[2]Department of Psychological Science, University of California, Irvine, Irvine, CA, United States

[3]Department of Informatics, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Justine Bautista, MA
School of Social Ecology
University of California, Irvine
Social Ecology I, 102
Irvine, CA, 92617
United States
Phone: 1 (949) 824 6861
Email: Jrbautis@uci.edu

Abstract

Background: Increasing rates of mental health diagnoses in college students signal the need for new opportunities to support the mental health of this population. With many mental health apps being efficacious, they may be a promising resource for college campuses to provide support to their students. However, it is important to understand why (or why not) students might want to use apps and their desired features.

Objective: Information on students' interest in mental health apps may inform which apps are to be provided and how campuses can support their use. This study aimed to understand the interest and hesitation in app use and the relationship between mental health needs, as defined by depression, anxiety, and positive mental health, and app use.

Methods: The web-based *Healthy Minds Study* collected information on mental health needs, perceptions, and service use across colleges and universities. We used a sample of 989 participants who completed the survey between 2018 and 2020 and an elective module on digital mental health. We analyzed the elective module responses using a mixed methods approach, including both descriptive and inferential statistics, along with thematic coding for open text responses.

Results: The Results from this study revealed that anxiety (b=−0.07; $P<.001$), but not depression (b=0.03; $P=.12$) and positive mental health (b=−0.02; $P=.17$), was a significant predictor of app adoption. Prominent qualitative findings indicated that the most desired app features included tips and advice, access to resources and information, and on-demand support that involves interaction throughout the day. The participants also suggested an overall desire for human interaction to be integrated into an app. As predicted, hesitancy was encountered, and the qualitative results suggested that there was a lack of interest in the adoption of mental health app and preference.

Conclusions: The findings from this study underscore that simply providing digital mental health apps as tools may be insufficient to support their use in college campuses. Although many students were open to using a mental health app, hesitation and uncertainty were common in the participant responses. Working with colleges and universities to increase digital literacy and provide resources that allow students to gauge when app use is appropriate may be helpful when implementing mental health apps as resources in college campuses.

(JMIR Ment Health 2023;10:e43942) doi:10.2196/43942

KEYWORDS

mental health; mental health apps; college students; digital health; app; anxiety

Introduction

Background

From 2007 to 2017, the rate of mental health disorders increased from 22% to 36% among college students [1]. The most prevalent mental disorders in this group are depression and anxiety [2]. Compared with other health services offered on college campuses, mental health services have the highest number of visits per patient, indicating recurring service use and high demand for these services [3]. Increasing the number of traditional mental health services alone is unlikely to meet this demand. Many college campuses are facing budgetary challenges, making additional investments unlikely. Even when funding is made available, campuses face difficulties in filling open positions because of a shortage of qualified providers, especially those with experience in treating college student mental health disorders [4]. When services do exist, students may not use them because of their preferences or a need for care outside regular operating hours of services [5]. Most US counties suffer from a shortage of mental health providers, leaving college counseling centers with an unmet need and difficulty in recruitment [6]. With a shortage of mental health providers in general, colleges and universities must compete for a limited pool of candidates that are able to support college students and have the necessary qualifications [7], In light of this, campuses may need to explore new opportunities to support college student mental health. One opportunity includes digital mental health resources that can address this demand, with limited additional burden on traditional service providers.

Mental Health App Engagement in College Students

Digital mental health is broadly defined as any form of technology used for mental health assessment, support, prevention, and treatment [8]. In particular, many campuses are offering digital mental health apps to their students either by purchasing or recommending them [9]. Digital mental health apps refer to software programs that provide tools and resources such as tracking, psychoeducation, and exercises to help people self-manage their conditions. Digital mental health apps are efficacious at addressing the mental health concerns of college students. Smartphone ownership is even higher in adults aged between 18 and 29 years (96%) than in adults (85%) generally [8]. Thus, college students may be an ideal population to target such resources. Because they can be accessed via mobile devices, mental health apps are widely available at any time, creating an especially accessible option for those who may have difficulty accessing mental health services. This is especially beneficial to college students, who may experience difficulty accessing campus resources when needed because of time constraints and staff shortages [5,7]. The convenience and flexibility of mental health apps may be especially appealing to college students. However, challenges in mental health app use include a lack of regulation of and research on many mental health apps available. With so many mental health apps being widely available, students may find it difficult to identify apps that contain evidence-based content, because research evidence for most mental health apps is scarce [9]. Furthermore, privacy and safety continue to be areas of concern, especially as privacy breaches continue to occur [9]. Nonetheless, several apps have been evaluated through prior research and deemed effective for alleviating mental health symptoms [10]. Thus, it is essential that college campuses provide safe, evidence-based, and effective tools for their students.

A systematic review of apps for college student mental health found that the majority of the apps evaluated were either effective or partially effective in alleviating symptoms of depression and anxiety or enhancing overall psychological well-being [10]. Despite their overall effectiveness, how best to get students to use mental health apps remains an open question. In light of this, more research is necessary to understand more specifically what college students may like or dislike in digital mental health apps. One study found that despite 53% of students downloading a digital mental health app at one point, only 19% were currently using a digital mental health app [11]. This signals a clear interest in using and downloading the apps, but a decrease in their actual use. In addition, most digital mental health interventions show low rates of adherence among college students, signaling a lack of sustained app use [12]. Nonetheless, those with clinical diagnoses of depression and anxiety seem to show increased interest in mental health apps [13]. In particular, web-based health resource use is especially high in those with mild to severe symptoms of depression [14]. The existing interest in those with anxiety and depression combined with their clear efficacy in alleviating symptoms leads us to believe that anxiety and depression may have an association with app use. However, low rates of adherence have also led us to explore whether this interest in apps translates to their actual use.

This study seeks to bridge the gap between interest in apps and app use, exploring why college students may choose to use an app or not. Specifically, we explored college students' existing attitudes toward mental health apps and the relationship between mental health needs, including depression, anxiety, and positive mental health, and app use. We hypothesized that we would see both positive and negative themes in attitudes with respect to mental health apps and that apps would be generally well received but would continue to encounter attitudes of uncertainty about their efficacy. We also hypothesized that depression and anxiety would have a significant association with mental health app adoption and the frequency of app use. Few studies have examined the association between positive mental health and app use; nevertheless, we hypothesized that positive mental health would not be a significant predictor, as no concrete link between the 2 variables has been identified in the past.

Methods

The Healthy Minds Study

This study used data from the *Healthy Minds Study*, an annual web-based survey distributed to colleges and universities across the United States. The *Healthy Minds Study* addresses various topics related to college student mental health, campus climate, help-seeking, service use, and overall student experience [15]. Incentives included a national sweepstakes, in which 2 students were selected for US $500 gift cards, while 10 were selected for US $100 gift cards [16]. However, institutions may also

provide their own incentives of their choice, with approval from the Healthy Minds Study team [16].

The survey is structured in 2 parts: standard modules and elective modules. Each participating college or university is required to administer the standard modules, consisting of basic demographic information (age, sex, race and ethnicity, socioeconomic status, etc), academic information, school experiences, mental health status, and mental health service use or help-seeking. Colleges and universities are also provided with a list of elective modules on topics such as substance use, sleep, eating and body image, sexual assault, overall health, and financial stress. Each college or university may select which of these elective modules are to be provided along with the standard modules. Thus, each student completing the survey receives the combination of standard modules and the university- or college-selected elected modules. Our analysis focuses on data collected from 2018 to 2020 after *Attitudes About Mobile Resources* elective module was introduced. However, the exact dates of data collection were unavailable in the data set provided.

Ethics Approval

Institutional Review Board (IRB) approval for the Healthy Minds Study was obtained through Advarra, an independent IRB service in North America. This study makes use of the publicly available and deidentified Healthy Minds Study data that are made available through this website [17]. This secondary data analysis was deemed exempt from an additional IRB review by the University of California, Irvine.

Participants

Participants consisted of college students who completed the *Healthy Minds Study* survey from 2018 to 2020. A broad range of colleges and universities participated in this study, with students pursuing an associate's degree, bachelor's degree, master's degree, JD, MD, PhD, or equivalent, as well as nondegree students being included in this sample. Overall, the data set consisted of 151,211 participants, of which 62,029 (41.02%) participants were from 78 colleges and universities during the 2018 to 2019 academic year, and 89,182 (58.98%) participants were from 75 colleges and universities during the 2019 to 2020 academic year. The 2018 to 2019 survey had a 16% participation rate while in the 2019 to 2020 academic year; the participation rate was 13% in the fall and 16% in the winter. Participation rate data were provided by The Healthy Minds Network. The 2018 to 2019 survey did not differentiate between fall and winter survey participation rates. Only a subset of participants completed our module of interest, "Attitudes About Mobile Resources." In the 2018 to 2019 academic year, 400 (0.64%) participants completed the module, whereas 589 (0.66%) participants completed the module in the 2019 to 2020 academic year, resulting in a total of 989 participants.

Overall, participants who completed the elective module were relatively similar to the broader sample. Both the samples had participants who were on an average aged 23 (SD 4.92) years; mostly female (overall: 101,000/151,211, 66.79%; elective module: 673/961, 70%); and mostly White (overall: 107,003/151,211, 70.8%; elective module: 788/961, 82%). Those who completed the elective module and the overall sample both had an average score of 44 (8.12) on the Diener and Diener Flourishing Scale but the sample that completed the elective module had slightly elevated percentages of students with depression (overall: 45,193/151,211, 29.89%; elective module: 310/961, 32.3%) and anxiety (overall: 39,126/151,211, 25.87%; elective module: 274/911, 28.5%). Furthermore, those who completed the elective module tended to pursue higher levels of education. Fewer students who completed the elective module were pursuing an associate's degree (overall: 14,150/151,211, 9.36%; elective module: 14/961, 1.6%), while a higher percentage of students were pursuing a bachelor's degree (overall: 100,548/151,211, 66.49%; elective module: 611/961, 69.1%); master's degree (overall: 18,325/151,211, 12.12%; elective module: 179/961, 20.2%); JD (overall: 980/151,211, 0.65%; elective module: 11/961, 1.2%); MD (overall: 1926/151,211, 1.27%; elective module: 17/961, 1.9%); or PhD (overall: 10,683/151,211, 7.06%; elective module: 130/961, 14.7%).

Measures

From the standard module, we used 3 measures to identify the mental health status of the participants. Depression was measured using the Patient Health Questionnaire-9 (PHQ-9), anxiety using the Generalized Anxiety Disorder Scale-7 (GAD-7), and positive mental health using the Diener and Diener Flourishing Scale.

The PHQ-9 is a 9-item measure of depressive symptoms [18]. PHQ-9 total scores range from 0 to 27, with a PHQ-9 score of >9 indicating clinically elevated levels of depressive symptoms [18]. The psychometric properties of the PHQ-9 evaluated in college students demonstrated its reliability and validity, even in diverse college student populations [19].

The GAD-7 contains 7 items measuring generalized anxiety [20]. GAD-7 total scores range from 0 to 21 [20]. A GAD-7 score of >9 indicates clinically elevated levels of anxiety symptoms [20]. Prior analyses of the GAD-7 suggested that it has strong reliability and validity when used as a screener for college students [21].

The Flourishing Scale is an 8-item scale measuring self-perceived success in various important areas of a participant's life [15]. Higher overall scores on the Flourishing Scale indicate greater psychological resources, with scores ranging from 8 to 56 [22]. The Flourishing Scale has demonstrated high reliability in measuring the well-being among university students [23].

The "Attitudes About Mobile Resources" elective module included the following questions:

1. "Would you be open to using an app for wellness or mental/emotional health?" (Yes; maybe; no) "Have you ever used a smartphone app to manage your wellness or mental/emotional health?" (No, never; yes)
2. "When did you use a smartphone app to manage your wellness or mental/emotional health?" (Before starting college; since starting college; I currently use an app)
3. "What are the reasons why you have not used a mental health app?" (I have concerns about privacy and security of data; there is lack of research support available; I'm

unsure about how useful the app will be; I have concerns about cost; apps seem difficult to use; I don't know if I could find a suitable app; I don't know which app to download; I don't have time to use apps; I'm not interested in using mental health apps; I don't think I need these kinds of apps; Other)

4. "Now imagine you are trying to decide which wellness or mental/emotional health app to use. How important would each of the following features be in your decision? (The app has research supporting its benefits; the app is well designed and easy to use; the app has information about data privacy and storage policies; the app has reviews from users; the app has reviews from experts in the field; the app developer; the cost of the app; the time commitment required by the app seems manageable to me; Something else)." (Not at all important; slightly important; moderately important; important; very important)
5. "How helpful, overall, do you think the smartphone app(s) was or has been for your mental or emotional health?" (Very helpful; helpful; somewhat helpful; not helpful)
6. "What would you hope to get out of an app for wellness or mental/emotional health?"
7. "Why would you not use a mental health app?"

Data Analysis

This study used a mixed methods approach, drawing from both qualitative and quantitative methods. Quantitative methods included both descriptive and inferential statistics to assess the variables that may affect app adoption. Qualitative methods included thematic analyses of open text responses to questions related to app use. The qualitative methods were intended to expand the findings and understand the quantitative responses.

Quantitative Methods

We aimed to understand the relationship between mental health (as measured by PHQ-9, GAD-7, and Flourishing scale) and mental health app use (adoption and frequency). To do so, we conducted a binary logistic regression with mental health status predicting adoption while controlling for demographic variables, including race and ethnicity, sex, age, and socioeconomic status. Independent binary logistic regressions were conducted for each mental health status variable (ie, depression, anxiety, and flourishing). Logistic regression was used to examine the relationship between mental health and the frequency of app use. To determine if clinical levels of mental health symptoms were related to app use, chi-squared difference tests were conducted with clinical thresholds identified by the PHQ-9 and GAD-7 (>9 for both measures). As we were simultaneously conducting statistical tests for separate mental health variables, Bonferroni correction was used to adjust the α level and address multiple comparisons. The adjusted α level was .017. Raw percentages were also calculated for the number of students who had ever used a mental health app and who would or would not be open to using a mental health app, the reasons for hesitation in app use, and the time points when students initiated app use.

Qualitative Methods

A thematic analysis [23] was conducted on the 2 open text responses in the elective module, and not all students answered each question. The response totals are reported below.

1. "What would you hope to get out of an app for wellness or mental/emotional health?" (541 responses)
2. "Why would you not use a mental health app?" (131 responses)

We used open coding to identify themes in attitudes toward mental health apps and their use. We created the codebook by first performing an initial analysis of the responses from all participants and then revising the codebook iteratively after an initial subset of the data was coded. After coding 10% of the data, the coders reviewed the responses and identified areas where more clarity was needed in the codebook. The 2 coders then coded the next 10% and continued to revise the codebook. Coding in segments also promoted increased reliability and consistency, allowing the 2 coders to review discrepancies and areas where the codebook could be refined. After coding 20% of the data, discussing, and revising the codebook, coders independently completed the remainder of the responses. In the final codebook, question 1 had a total of 38 codes, whereas question 2 had a total of 12 codes. We assessed the interrater reliability with percent agreement, which was 80.3% for question 1 and 91.2% for question 2. Given the high level of agreement and the goal to identify themes present in the data, rather than requiring consensus, a code was counted as present if either of the coders indicated that code for a particular response. Thus, some responses received 2 codes when the coders were not in agreement.

Results

Demographic Information

A total of 961 students completed the elective module. These 961 students were predominantly White (n=788, 82%) and female (n=673, 70%). The average age of the participants was 22.63 (SD 4.92) years. The mean level of the mental health variables was 7.98 (SD 5.94) for depression, 6.98 (SD 5.38) for anxiety, and 44.03 (SD 8.12) for positive mental health. The rates of clinically elevated depression and anxiety levels were 32.3% (310/961) and 28.5% (274/961), respectively.

Quantitative Results

Table 1 presents the percentages of app use and the reasons for hesitation across the sample. The results of the binary logistic regression for depression, anxiety, and positive mental health as predictors of app use found that after controlling for sex, age, socioeconomic status, and race and ethnicity, anxiety was the only significant predictor of app adoption (b=−0.07; P<.001). Neither depression (b=0.03; P=.12) nor positive mental health (b=−0.02; P=.17) were significant predictors of app adoption. We used McFadden pseudo-R^2 (ρ^2) to calculate goodness of fit (ρ^2=0.03), which indicated a small effect size.

We also examined whether depression, anxiety, and positive mental health would predict greater frequency of app use. The resulting model did not show a significant relationship between

the mental health variables and frequency of app use (R^2=0.05; $F_{13,230}$=0.9; P=.56). Investigation of the coefficients showed that the presence of depression (b=0.02; P=.63), anxiety (b=−0.003; P=.91), and positive mental health (b=0.01; P=.57) had no significant relationship with the frequency of app use.

We used a chi-squared test of independence to analyze whether those who met the clinical thresholds of depression and anxiety showed higher rates of app adoption. No relationship was found between anxiety (χ^2_1=2.8, N=961; P=.10) and app adoption and depression (χ^2_1=0.9, N=961; P=.34) and app adoption.

Table 1. App use and hesitation.

Response	Values, n (%)
Students who have used a mental health app	
Yes	245 (25.4)
No	719 (74.6)
Students who would be open to using a mental health app	
Yes	368 (38.02)
No	144 (14.88)
Maybe	456 (47.11)
Time points when students used a mental health app	
Before starting college	64 (26.1)
Since starting college	172 (70.2)
Currently using	85 (34.7)
Hesitation to use a mental health app	
I'm unsure about how useful the app will be	351 (48.8)
I don't know which app to download	290 (40.3)
I don't need these kinds of apps	262 (36.4)
I don't know if I could find a suitable app	193 (26.8)
I'm not interested in using mental health apps	191 (26.6)
I have concerns about privacy and security of data	139 (19.3)
I don't have time to use apps	101 (14.1)
I have concerns about cost	82 (11.4)
I don't have a suitable device or enough space to download apps	34 (4.7)
There is a lack of research support available	68 (9.5)
Apps seem difficult to use	15 (2.1)

Qualitative Results

The first question that we qualitatively analyzed asked students what they would like to see in a mental health app. We grouped the codes into categories as shown in Table 2. The first was "what they want" and the second was "how they want it." The "what they want" category consisted of codes that addressed specific app features. Upon calculating the frequency of the codes, we found that the 3 features that students were most interested in were related to tips and advice. The third most prominent code was also closely related to this, with many asking for "relaxation/calming tips," which refer to apps that may have features for calming or reducing stress in the user. Some examples of student responses within this theme are "Tips and tools for when feeling anxious..." and "Tips/strategies for dealing with stressors in my life."

"Access to resources and information" was also a prominent code. Example responses included, "information to read, resources listed" and "...find in-person resources." Another larger proportion of code seemed to suggest that students want on-demand support that involved interaction throughout the day. Students wanted "tracking and documenting/journaling" features along with "reminders and check-ins." Example responses include, "tracking factors that contribute to emotional wellbeing" and "reminders of how to not be anxious or deal with situations, or just reminders."

Table 2. Desired mental health app features.

Code	Values, n (%)
What they want	
Tips and advice (general)	90 (16.7)
Anxiety and mental wellness tips	68 (12.6)
Relaxation or calming tips	58 (10.8)
Not sure or I don't know	58 (10.8)
Access to resources and information	57 (10.6)
Tracking and documenting or journaling	52 (9.6)
Reminders and check-Ins	42 (7.8)
Guided meditation	41 (7.6)
Access to medical help and appointments	30 (5.6)
Encouragement and Motivation	20 (3.7)
Daily tasks or routines and goal-setting	18 (3.3)
Positive affirmations	16 (3)
Holistic well-being	15 (2.8)
Activities	13 (2.4)
An emotional outlet	13 (2.4)
Confidence and empowerment	12 (2.2)
Breathing exercises	11 (2)
Health and physical exercise tips	10 (1.9)
Productivity and time management	9 (1.7)
Clarity and perspective	8 (1.5)
Comfort	8 (1.5)
Stability	8 (1.5)
Depression help	4 (0.7)
Self-care tips	3 (0.6)
Sleeping tips	3 (0.6)
Help with suicidal thoughts	1 (0.2)
How they want it	
Communication and real-time support	68 (12.6)
Professional support	24 (4.5)
Peer support	23 (4.3)
Convenience in accessibility	23 (4.3)
Affordability	7 (1.3)
Anonymity	7 (1.3)
Similarity to other apps	7 (1.3)
Usability	6 (1.1)
Confidentiality	5 (0.9)
Unique to existing resources	4 (0.7)
No judgment	3 (0.6)
Hesitation in using mental health apps	
Not helpful, ineffective, or not useful	29 (22.3)
I don't need it	25 (19.2)

Code	Values, n (%)
I would not use it	25 (19.2)
Prefers face-to-face counseling or in-person interaction	16 (12.3)
Privacy and confidentiality concerns	14 (10.8)
Dislike for apps	11 (9.2)
Smartphone worsens mental health	12 (8.5)
Impersonal	8 (6.2)
Already using a mental health app	5 (3.8)
I don't own or have limited access to a smartphone	4 (3.1)
Not sure or I don't know	4 (3.1)
Religious and family support	3 (2.3)

In the "how they want it" category, students showed an overwhelming desire for "communication and real-time support" features. This code had 2 subcodes, "professional support" and "peer support," which were both prominent codes in our analysis. Example responses include, "An anonymous peer-peer chat group or access to a mental health professional" and "convenience of a mental health professional anywhere and anytime." This suggests that one theme of these responses is the overall desire for human interaction to be integrated into an app. Students wanted the ability to communicate in real time with both peers and professionals.

Our second question aimed to explore reasons why students may not use a mental health app, providing insight into hesitation in mental health app adoption. The most prominent responses were coded as "not helpful/ineffective/not useful." "I don't need it" and "I would not use it" were also prominent codes, which reflects a lack of interest. Example responses include, "i don't think talking to a computer could help," "worried about the ability of such an app," and "I do not think that it could provide realistic or helpful information." These 3 responses may provide insight into how resources in college and university settings may not be well received by some students. In addition, the responses indicated a preference for face-to-face counseling or in-person interaction, adhering to more traditional methods of mental health help-seeking.

Discussion

Principal Findings

Our findings suggest that simply providing digital mental health apps may not be sufficient to increase their use among college students. Rather, we identified considerable hesitation and uncertainty toward using mental health apps and the various factors related to app adoption. Despite numerous studies examining perceptions and outcomes of mental health app use, few studies give college students the space to openly discuss and identify desired individual app features, hesitation, and barriers to app use. Our findings provide information on how college students use mental health apps, their interest in particular app features or capabilities, and hesitation that may affect app adoption.

Contrary to our hypothesis, which suggested that anxiety and depression would both be significant predictors of app adoption, our quantitative findings suggested that anxiety was the only significant predictor of app adoption. This aligns with our qualitative findings in which one prominent feature that students wanted in mental health apps was tips and advice related to anxiety and relaxation. Students wanted apps that directly addressed how to reduce anxiety symptoms and promote relaxation. Prior literature suggests that meditation apps are among the most commonly downloaded mental health apps, and meditation is largely popular among those with anxiety [24,25]. With meditation apps being popular and appealing to those with anxiety, college campuses should consider them as a potentially valuable resource. Surprisingly, the qualitative analysis also found very few responses indicating that the students wanted to see features that directly addressed depression. This could be due in part to meditation apps, with meditation appealing to those with anxiety and being the most popular category of mental health apps and the most well known. Those with anxiety may feel that the most popular mental health apps are more tailored to their needs than those with depression. Regarding positive mental health, our findings also suggested that it did not affect app use. One explanation for these findings is that negative emotions, and the desire to reduce them, might drive the adoption of mental health apps among college students rather than positive emotions. Some conceptualizations of depression and anxiety characterize them by the presence of positive and negative emotions [26]. In such conceptualizations, depression is characterized by low positive affect and low negative affect, whereas anxiety is characterized by low positive affect and high negative affect. This would be consistent with the lack of significant findings for positive mental health, which would be thought of as consisting of high positive and low negative affect [27].

It should also be noted that only about 25% (245/961) of the respondents reported using a mental health app. This is in contrast to prior literature, which suggests that 53% of college students have used a mental health app [28]. Much of the prior work, however, has focused on 4-year universities, and our data set consisted of a wide variety of colleges and universities. Indeed, one study of community college students found that 21.2% reported using a mental health app [29]. Thus, adoption rates may vary across types of colleges and universities, and

more work should be conducted that is more representative of different types of higher educational settings.

Although anxiety predicted app adoption, it did not predict more frequent app use. One explanation would be that stress may inhibit sustained mental health app use [11]. As college students face stressors associated with academics, financial issues, and peer relationships, using a mental health app may not fit into their already busy lives. This may also be due to the uncertainty surrounding app use and perhaps a lack of digital literacy regarding the use of an app. Students may be unsure of how to use the app or identify how often to use it. This could explain why students with elevated anxiety are not likely to use a mental health app more frequently, although they are more likely to download a digital mental health app. Even if a student with anxiety were to download a mental health app, they may be unable to identify how often they should use it based on their symptoms. In contrast, it is also possible that anxiety did not predict frequent app use because those who used apps frequently experienced reduced anxiety. Given that our data were cross-sectional, we could not determine how adoption might lead to changes in clinical outcomes and, ultimately, impact long-term use.

With other methods, such as traditional in-person therapy or prescribed medications, patients receive direct instruction and are given a schedule for therapy sessions or medication dosage. When using a mental health app, students must independently decide what a "healthy" amount of app use looks like and what apps would be useful in relation to their symptoms. The second largest point of hesitation was uncertainty about which apps they should download. Increasing digital literacy and providing tools that allow students to gauge when app use is appropriate may be helpful when implementing digital mental health apps as resources in college campuses. Working with ecosystems that already exist in students' lives may be especially beneficial. For example, Kaiser Permanente created a set of mental health apps to provide to their members. To support delivery and integration, health care providers within the Kaiser Permanente network were trained to provide a web-based mental health tool as a resource [30]. Similar efforts could benefit college campuses by training counseling center staff or other providers, peers, or students to increase digital literacy and find effective evidence-based mental health apps. Digital navigators, or people designated to evaluate and recommend apps, set up technology, and collect data within the organization, can work with counseling center staff to implement mental health apps on campus [31]. In addition, factors that predict a student's decision to begin using an app may not predict their sustained use. It is possible that those with higher anxiety levels did not experience sufficient improvement to continue using the app. However, it is also important to note that these data were cross-sectional. Consequently, it is difficult to determine the full scope of discovery, adoption, and sustainment.

In students who had used a mental health app since starting college, only about half were still using one at the time the survey was completed. In line with findings from the study by Kern et al [32], students have a clear interest in mental health apps but do not sustain their use. Again, it is possible that as students experience better mental health, the need for a mental health app no longer exists, leading them to delete the app. However, it is also possible that uncertainty around app use may prevent students from using an app or continuing its use.

In addition, it is not surprising that an overarching theme was that students wanted apps to provide resources and increased access to mental health care. Access to resources and information along with access to medical help and appointments ranked highly on the list of the desired app features. The most common theme for the "how they want it" section suggested that students wanted apps that provided real-time support and connected them with both peers and professionals. Most mental health apps seem to have similar features, with mood tracking and journaling being the most common [33]. Apps that connect students to resources and increase access to medical care, professional support, and peer support may be worthwhile for this population. However, none of these features are mentioned in the list of the top 5 most common offerings of mental health apps [31]. In the study by Lagan et al [33], the analysis revealed that very few apps offered peer support and connection to a coach or therapist, whereas none offered access to resources and information or access to medical care. In addition, no themes emerged regarding cultural competency, although our sample consisted of White individuals predominantly. Integrating campus resources with mental health apps may be of particular interest to college students [34]. Fortunately, there are some products that demonstrate these features. For example, YOU at College is a "full continuum of campus care," and it provides multiple features. Their "YOU" app connects students to resources available on their campus along with other resources and content related to mental health. In addition, their "NOD" app is a cognitive behavior therapy–based app that provides social challenges, prompts, and reflection opportunities to combat loneliness and avoidance. The "NOD" app is built on evidence-based principles and has been shown to reduce loneliness in susceptible students [35]. Resources such as these may be useful for campuses to consider, especially as services, such as YOU at College require enterprise adoption, that is, a college or university must purchase it to make it available for its students.

In students who had not used a mental health app, a number of factors led to hesitancy. It was most common for students to indicate that they were unsure of how useful the app would be. The second major point of hesitation was the uncertainty around which app to download. With so many apps existing in the app store, it is understandable why it might be difficult for students to select one. The app store houses thousands of mental health apps, with very few actually having empirical support [36]. Therefore, it may be confusing for many students to make informed decisions about which app to download. Providing empirical evidence and helping students identify whether the app is useful for them may be especially helpful. Teaching students how to identify a suitable evidence-based mental health app in the app store may also help in clearing uncertainty about whether a mental health app could actually be useful, which may be done through the *One Mind PsyberGuide* [37]. Interestingly, very few students were concerned about the costs associated with the app and access to technology in both our quantitative and qualitative findings. This suggests that students

do feel they have sufficient resources to be able to access the app but are unsure of whether it would be useful or which they should select.

Qualitative findings around hesitancy suggested that many students simply felt that they did not need a mental health app or that it would not be effective. Although there are mental health apps that are effective, many mental health apps are not evidence based [38]. Therefore, students' perceptions of mental health apps may be accurate. This also aligns with prior findings on mental health service use in college campuses. College students are often skeptical of the effectiveness of existing services [39]. Thus, hesitancy toward mental health services among college students may be a more pervasive issue that exists even outside the digital realm. Our qualitative findings suggest that there seems to be a preference toward face-to-face counseling, which also exhibits the assumption that using a mental health app would mean abandoning in-person therapy. Rather than considering mental health apps as a replacement for traditional forms of therapy, it should be noted that they may be used as adjuncts to care. Preference for face-to-face counseling and in-person interaction was the fourth major point of hesitation in using mental health apps, with about 1 in 8 students mentioning this preference. It may be useful for campuses to market mental health apps as additional resources rather than as a replacement for traditional forms of therapy. Although campuses do not directly suggest that an app should be a replacement, it may be helpful to reiterate to students that the app should be used in addition to other resources to combat the misconceptions that many students have. Furthermore, students often experience long wait times and issues with scheduling appointments for college counseling services [5]. These apps could be useful for students while waiting for counseling services, as mental health apps can be accessed immediately.

Limitations

As previously mentioned in *Methods* section, demographic characteristics of the elective module sample and overall sample tended to be relatively similar. However, the elective module participants had a larger percentage of students pursuing higher levels of education. More students in the elective module were pursuing graduate studies. Furthermore, more students were pursuing a bachelor's degree in the elective module, while a smaller percentage were pursuing an associate's degree when compared with the overall sample. We noted earlier that app adoption may differ among types of college or university settings and differences in adoption rates might also exist among types of degrees. In addition, issues such as accessibility of resources could differ in the overall sample compared with the elective module, as community college students often experience severe psychological concerns with less institutional support for mental health resources [40]. With fewer students in the elective module pursuing an associate's degree, there remains uncertainty about whether students in a 2-year university may have had different attitudes toward mental health apps.

It should be noted that the pandemic may have affected this study and attitudes toward mental health apps. Some students might have provided data after the March 2020 shutdown, which had an impact on their mental health, campus resources, and broader aspects of student life. The data set did not provide the survey completion dates for each participant; therefore, we were unable to exclude those who completed the survey after March 2020 or compare results across those who completed before and after March 2020. However, it is difficult to predict the long-term impact of the pandemic, as it was shaping people's knowledge and interest in using digital resources to support their mental health [41]. Certainly, college and university campuses are reconsidering opportunities for remote and hybrid services for a variety of purposes, including mental health care.

Conclusions

Our findings reflect the need for college campuses to provide resources tailored to the needs of students. Campuses should not only provide mental health apps as a resource but also provide resources that clear uncertainty regarding available apps. It is essential to consider why people use or do not use mental health apps and what is needed, in addition to the apps to support their use. On the basis of the responses of the students, plenty of hesitation continues to inhibit app adoption, with many students experiencing uncertainty about using a mental health app. In students who chose to use a mental health app, the presence of anxiety seemed to predict app adoption. As students tend to begin using mental health apps in colleges, it has become imperative to provide the resources necessary to promote mental well-being on campus.

Acknowledgments

Data permissions were provided by The Healthy Minds Network. The authors would like to thank Kidra Setayeshi, who assisted with the qualitative analysis of this project.

Conflicts of Interest

SMS received consulting payments from Otsuka Pharmaceuticals for work unrelated to this manuscript and is on the Scientific Advisory Board for Headspace for which he receives compensation.

References

1. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007-2017). Psychiatr Serv 2019 Jan 01;70(1):60-63 [FREE Full text] [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]

2. Pedrelli P, Nyer M, Yeung A, Zulauf C, Wilens T. College students: mental health problems and treatment considerations. Acad Psychiatry 2015 Oct;39(5):503-511 [FREE Full text] [doi: 10.1007/s40596-014-0205-9] [Medline: 25142250]
3. Turner JC, Keller A. College health surveillance network: epidemiology and health care utilization of college students at US 4-year universities. J Am Coll Health 2015;63(8):530-538 [FREE Full text] [doi: 10.1080/07448481.2015.1055567] [Medline: 26086428]
4. Ponsford LR. Assessing mental health and violence on college campuses using the vulnerability model. J Am Assoc Nurse Pract 2016 Apr;28(4):212-217. [doi: 10.1002/2327-6924.12351] [Medline: 27098402]
5. Cohen KA, Graham AK, Lattie EG. Aligning students and counseling centers on student mental health needs and treatment resources. J Am Coll Health 2022 Apr;70(3):724-732. [doi: 10.1080/07448481.2020.1762611] [Medline: 32432973]
6. Butryn T, Bryant L, Marchionni C, Sholevar F. The shortage of psychiatrists and other mental health providers: causes, current state, and potential solutions. Int J Acad Med 2017;3(1):5-9 [FREE Full text] [doi: 10.4103/IJAM.IJAM_49_17]
7. Thomas KC, Ellis AR, Konrad TR, Holzer CE, Morrissey JP. County-level estimates of mental health professional shortage in the United States. Psychiatr Serv 2009 Oct;60(10):1323-1328. [doi: 10.1176/ps.2009.60.10.1323] [Medline: 19797371]
8. Wies B, Landers C, Ienca M. Digital mental health for young people: a scoping review of ethical promises and challenges. Front Digit Health 2021 Sep 06;3:697072 [FREE Full text] [doi: 10.3389/fdgth.2021.697072] [Medline: 34713173]
9. Lattie EG, Kornfield R, Ringland KE, Zhang R, Winquist N, Reddy M. Designing mental health technologies that support the social ecosystem of college students. Proc SIGCHI Conf Hum Factor Comput Syst 2020 Apr;2020:1-15 [FREE Full text] [doi: 10.1145/3313831.3376362] [Medline: 32656549]
10. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
11. Melcher J, Camacho E, Lagan S, Torous J. College student engagement with mental health apps: analysis of barriers to sustained use. J Am Coll Health 2022;70(6):1819-1825. [doi: 10.1080/07448481.2020.1825225] [Medline: 33048626]
12. Becker TD, Torous JB. Recent developments in digital mental health interventions for college and university students. Curr Treat Options Psychiatry 2019 Jun 14;6(3):210-220 [FREE Full text] [doi: 10.1007/s40501-019-00178-8]
13. Lipschitz JM, Connolly SL, Miller CJ, Hogan TP, Simon SR, Burdick KE. Patient interest in mental health mobile app interventions: demographic and symptom-level differences. J Affect Disord 2020 Feb 15;263:216-220. [doi: 10.1016/j.jad.2019.11.083] [Medline: 31818779]
14. Rideout V, Fox S, Peebles A, Robb MB. Coping with COVID-19: how young people use digital media to manage their mental health. Common Sense and Hopelab. San Francisco, CA, USA: Common Sense and Hopelab; 2021. URL: https://tinyurl.com/ym7tr6ax [accessed 2022-12-11]
15. Healthy Minds Study among Colleges and Universities, (2018-2020). Healthy Minds Network, University of Michigan, University of California Los Angeles, Boston University, and Wayne State University. 2020. URL: https://tinyurl.com/pxjxzaar [accessed 2022-02-18]
16. The Healthy Minds Network. URL: https://healthymindsnetwork.org/ [accessed 2023-03-14]
17. School participation overview guide | Academic Year 2020-2021. Healthy Minds Network, University of Michigan, University of California Los Angeles, Boston University, and Wayne State University. URL: https://tinyurl.com/2p8yee5t [accessed 2022-02-21]
18. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
19. Keum BT, Miller MJ, Inkelas KK. Testing the factor structure and measurement invariance of the PHQ-9 across racially diverse U.S. college students. Psychol Assess 2018 Aug;30(8):1096-1106. [doi: 10.1037/pas0000550] [Medline: 29565614]
20. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
21. Byrd-Bredbenner C, Eck K, Quick V. Psychometric properties of the Generalized Anxiety Disorder-7 and Generalized Anxiety Disorder-mini in United States university students. Front Psychol 2020 Sep 24;11:550533 [FREE Full text] [doi: 10.3389/fpsyg.2020.550533] [Medline: 33071867]
22. Diener E, Wirtz D, Biswas-Diener R, Tov W, Kim-Prieto C, Choi DW, et al. New measures of well-being. In: Diener E, editor. Assessing Well-Being: The Collected Works of Ed Diener. Dordrecht, The Netherlands: Springer; 2009:247-266.
23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
24. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019 Jun 17;2:54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
25. Histon T, Heisler WS, Ozer H, Castaldo P. X project: getting grounded—how project Chamai is poised to serve emotional wellness needs of kaiser Permanente members. Perm J 2018;22:18-071. [doi: 10.7812/TPP/18-071-01]
26. Watson D, Clark LA, Carey G. Positive and negative affectivity and their relation to anxiety and depressive disorders. J Abnorm Psychol 1988 Aug;97(3):346-353. [doi: 10.1037//0021-843x.97.3.346] [Medline: 3192830]

27. Diener E, Pressman SD, Hunter J, Delgadillo-Chase D. If, why, and when subjective well-being influences health, and future needed research. Appl Psychol Health Well Being 2017 Jul;9(2):133-167 [FREE Full text] [doi: 10.1111/aphw.12090] [Medline: 28707767]
28. Mobile fact sheet. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-12-08]
29. Borghouts J, Eikey EV, Mark G, De Leon C, Schueller SM, Schneider M, et al. Understanding mental health app use among community college students: web-based survey study. J Med Internet Res 2021 Sep 14;23(9):e27745 [FREE Full text] [doi: 10.2196/27745] [Medline: 34519668]
30. Mordecai D, Histon T, Neuwirth E, Heisler WS, Kraft A, Bang Y, et al. How Kaiser Permanente created a mental health and wellness digital ecosystem. NEJM Catal Innov Care Deliv 2021 Jan 01;2(1):1 [FREE Full text] [doi: 10.1056/cat.20.0295]
31. Wisniewski H, Torous J. Digital navigators to implement smartphone and digital tools in care. Acta Psychiatr Scand 2020 Apr;141(4):350-355 [FREE Full text] [doi: 10.1111/acps.13149] [Medline: 31930477]
32. Kern A, Hong V, Song J, Lipson SK, Eisenberg D. Mental health apps in a college setting: openness, usage, and attitudes. Mhealth 2018 Jun 30;4:20 [FREE Full text] [doi: 10.21037/mhealth.2018.06.01] [Medline: 30050916]
33. Lagan S, D'Mello R, Vaidyam A, Bilden R, Torous J. Assessing mental health apps marketplaces with objective metrics from 29,190 data points from 278 apps. Acta Psychiatr Scand 2021 Aug;144(2):201-210. [doi: 10.1111/acps.13306] [Medline: 33835483]
34. Lattie EG, Lipson SK, Eisenberg D. Technology and college student mental health: challenges and opportunities. Front Psychiatry 2019 Apr 15;10:246 [FREE Full text] [doi: 10.3389/fpsyt.2019.00246] [Medline: 31037061]
35. Bruehlman-Senecal E, Hook CJ, Pfeifer JH, FitzGerald C, Davis B, Delucchi KL, et al. Smartphone app to address loneliness among college students: pilot randomized controlled trial. JMIR Ment Health 2020 Oct 20;7(10):e21496 [FREE Full text] [doi: 10.2196/21496] [Medline: 33079071]
36. Wang L, Fagan C, Yu CL. Popular mental health apps (MH apps) as a complement to telepsychotherapy: guidelines for consideration. J Psychother Integr 2020 Jun;30(2):265-273 [FREE Full text] [doi: 10.1037/int0000204]
37. Neary M, Bunyi J, Palomares K, Mohr DC, Powell A, Ruzek J, et al. A process for reviewing mental health apps: using the one mind PsyberGuide credibility rating system. Digit Health 2021 Oct 29;7:20552076211053690 [FREE Full text] [doi: 10.1177/20552076211053690] [Medline: 34733541]
38. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019 Mar 22;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
39. Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011 May;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]
40. Son C, Hegde S, Smith A, Wang X, Sasangohar F. Effects of COVID-19 on college students' mental health in the united states: interview survey study. J Med Internet Res 2020 Sep 03;22(9):e21279 [FREE Full text] [doi: 10.2196/21279] [Medline: 32805704]
41. Ben-Zeev D. The digital mental health genie is out of the bottle. Psychiatr Serv 2020 Dec 01;71(12):1212-1213. [doi: 10.1176/appi.ps.202000306] [Medline: 32576123]

Abbreviations

GAD-7: Generalized Anxiety Disorder Scale-7
PHQ-9: Patient Health Questionnaire-9

Edited by J Torous; submitted 31.10.22; peer-reviewed by A Peipert, K Cohen, C Benjet, I Shubina; comments to author 02.12.22; revised version received 15.12.22; accepted 15.12.22; published 22.03.23.

Please cite as:
Bautista J, Schueller SM
Understanding the Adoption and Use of Digital Mental Health Apps Among College Students: Secondary Analysis of a National Survey
JMIR Ment Health 2023;10:e43942
URL: https://mental.jmir.org/2023/1/e43942
doi:10.2196/43942
PMID:36947115

Original Paper

Intervening on Social Comparisons on Social Media: Electronic Daily Diary Pilot Study

Fernanda C Andrade[1*], MA; Savannah Erwin[1*], MA; Kaitlyn Burnell[2], PhD; Jalisa Jackson[1], BA; Marley Storch[1], BS; Julia Nicholas[3], BS; Nancy Zucker[1], PhD

[1]Department of Psychology & Neuroscience, Duke University, Durham, NC, United States

[2]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[3]Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY, United States

[*]these authors contributed equally

Corresponding Author:
Fernanda C Andrade, MA
Department of Psychology & Neuroscience
Duke University
417 Chapel Drive
Campus Box 90086
Durham, NC, 27708
United States
Phone: 1 (919) 660 5640
Email: fernanda.andrade@duke.edu

Abstract

Background: Literature has underscored the dark aspects of social media use, including associations with depressive symptoms, feelings of social isolation, and diminished self-esteem. Social comparison, the process of evaluating oneself relative to another person, is thought to contribute to these negative experiences such that people with a stronger tendency to compare themselves with others are particularly susceptible to the detrimental effects of social media. Social media as a form of social connection and communication is nevertheless an inevitable—and arguably integral—part of life, particularly for young adults. Therefore, there is a need to investigate strategies that could alter the manner in which people interact with social media to minimize its detrimental effects and maximize the feelings of affiliation and connection.

Objective: This pilot study examined the feasibility, acceptability, and effectiveness of a brief web-based intervention designed to alter engagement with social media and promote psychological well-being by encouraging social savoring as an alternative to social comparison. Social savoring was operationalized as experiencing joyful emotions related to the happiness of someone else's experiences (ie, feeling happy for someone else).

Methods: Following an intensive longitudinal design, 55 college students (mean age 19.29, SD 0.93 years; n=43, 78% women and n=23, 42% White) completed baseline measures (individual differences, psychological well-being, connectedness, and social media use) and then 14 days of daily surveys on their social media activity and well-being. On day 8, the group that was randomized to receive the intervention watched a video instructing them on the skill of social savoring and was asked to practice this skill during days 8 to 14.

Results: Overall, participants reported positive perceptions of the intervention. Participants who watched the intervention video reported significantly higher performance self-esteem (P=.02) at posttest than those in the control condition, after controlling for baseline levels. Participants also reported significantly higher state self-esteem (P=.01) on days in which they engaged in more social savoring while using social media, and the use of social savoring increased significantly (P=.01) over time, suggesting that participants found it helpful. Participants in both conditions reported significantly lower levels of social comparison (control: P=.01; intervention: P=.002) and higher levels of connectedness (control: P<.001; intervention: P=.001) at posttest than at baseline.

Conclusions: Initial evidence from this pilot study suggests that a web-based social savoring intervention may help minimize the potentially harmful consequences of social media use, at least in some domains. Future work is needed to examine the effectiveness and acceptance of this intervention in different age groups and in clinical samples that are in part characterized by higher levels of comparison with others (eg, people with eating disorders).

(JMIR Ment Health 2023;10:e42024) doi:10.2196/42024

KEYWORDS

social media; social comparison; young adults; social savoring; intervention; self-esteem; depression

Introduction

Social Comparisons on Social Media

Today, 72% of the American public uses some type of social media, and most people interact with at least 1 social media site daily [1]. Social media platforms help people synchronously and asynchronously connect with others. Adolescents and young adults, in particular, use social media to build, maintain, and strengthen their social networks [2-5]. Beyond fortifying social networks, engagement with social media content has been associated with a tendency to engage in social comparisons [6]. Social comparison, the act of comparing oneself with others, stems, in part, from people's need to evaluate their own opinions and abilities by comparing themselves with other people [7]. Such comparisons can have positive intentions and outcomes, such as motivating self-improvement, enhancing learning, and fostering positive self-views [8].

Notwithstanding, exposure to social media content can increase the risk of poor mental health outcomes [9,10], particularly among those with higher tendencies to engage in social comparisons [11]. Comparisons with others who are perceived to be better off than oneself (upward comparisons) tend to have a more marked negative impact on well-being than comparisons with others who are perceived to be on an equal level (lateral comparisons) or who are perceived to be worse off than oneself (downward comparisons) [12]. Upward social comparisons have consistently been related to more negative self-judgments, lower self-esteem [9], and the presence of disordered eating behaviors [13,14]. The visual nature of many social media platforms creates a rich environment for upward social comparisons, particularly with respect to performance and physical appearance, which can have negative consequences for well-being [12,15,16]. For example, a meta-analysis of 156 studies found that social comparison is positively associated with body dissatisfaction, especially among women and younger people [17]. Furthermore, adolescents and young adults often portray ideal versions of themselves on social media using advanced filters, photo editing, and video editing. They may seek out experiences with specific considerations of how such content will be perceived when posted on social media, rather than how intrinsically motivating it is [18,19]. Young adults viewing this carefully curated content often perceive others on social media as having better lives [20] and thus are prone to the negative effects of these comparisons. Indeed, a systematic review and meta-analysis of 70 published studies showed that social comparison on social media is positively related to higher levels of depression and anxiety [21].

Intervening on Social Comparisons

A growing body of work has investigated the impact of interventions designed to reduce social comparisons made during social media use. A subset of these interventions targets a lack of realism in social media content by manipulating participants' exposure to different content. For instance, a study found that participants who primarily viewed selfies of people without makeup reported lower facial dissatisfaction than those who viewed selfies of people with makeup [22]. Likewise, participants who viewed parodied versions of celebrity images reported lower body dissatisfaction and greater positive mood than those who were exposed to thin idealized images of celebrities [23]. In another intervention, women who saw side-by-side images of idealized versus nonidealized targets reported lower body dissatisfaction than those who viewed the idealized images alone [24].

Although changing the social media content with which people engage can reduce the negative impact of social comparisons, such an approach may not be realistic. Moreover, not all upward social comparison is harmful. Unlike comparisons that foster contrast between oneself and others, comparisons that foster assimilation (ie, the belief that one can obtain the same status as someone else) are characterized by a selective focus on the similarities between oneself and others [25], which can motivate positive feelings about oneself [26].

Consistent with this evidence, other interventions have targeted how participants interact with potentially harmful social media content using, for example, self-compassion and mindfulness-based interventions. Self-compassion refers to kindness toward oneself, acceptance of one's humanness, and understanding that negative experiences are universal [27]. Self-compassion may help reduce the negative impact of social media by fostering more stable self-worth and less externally contingent self-esteem. People with higher self-compassion tend to feel less inadequate and less judged or evaluated than people with lower reported self-compassion when exposed to the coveted experiences of others [27,28]. Mindfulness-based interventions enhance users' attention to their present experiences and foster curiosity, rather than judgment, about one's states. Mindfulness is negatively related to fear of missing out (FoMO), which refers to the apprehension that others may be having more positive experiences than oneself [29,30]. Similar to upward social comparisons that decrease positive affect, FoMO is pervasive on social media and related to more depressive symptoms [29]. Social media users who endorse higher mindfulness also report lower depression than users with lower mindfulness [31], and mindfulness-based interventions have shown promise in reducing body dissatisfaction and negative mood [32,33].

Although promising, other studies have found no evidence in support of these interventions designed to impact how people interact with social media [28,34]. These contradictory findings could stem from methodological differences, including intervention duration or the lack of control over prior states. It is also possible that these strategies are effective only among those with a higher tendency to compare themselves with others, as these people tend to be more negatively impacted by comparisons on social media [10,11,23,24,34].

Savoring Interventions

A related strategy, yet less explored with reference to social media, is savoring. Savoring is an emotion regulation approach by which people focus their awareness on pleasant experiences and appreciate the pleasure of past, present, and future experiences [35,36]. At a trait level, savoring appears to moderate the relationship between positive or negative personal experiences and various outcomes, including happiness [37], life satisfaction [38], and depression [39,40]. At a state level, savoring can bolster and maintain levels of reward sensitivity [41] and improve people's capacity to recognize and enjoy positive experiences even in the face of negative events [39].

Savoring is a promising target for psychological interventions [42]. A 6-week intervention that involved weekly sessions where savoring skills were taught reduced negative affect and increased positive affect in a sample of people newly diagnosed with HIV [43]. Similarly, a 2-week intervention showed that participants who engaged in and logged daily savoring practices reported lower levels of depressive symptoms and negative affect than those who did not receive the intervention [44]. Another study used social media platforms as a means to practice savoring by asking participants to describe and post about joyous activities performed during the week on social media platforms [45]. This intervention helped decrease depressive symptoms and negative mood compared with inactive controls.

Although a growing body of literature suggests that savoring can increase positive affect and reduce negative affect and depression, this skill has received little attention as a means to manage the negative impact of social comparisons on social media. In this project, we built on the concept of savoring and adapted it to be an interpersonal, empathic process we term "social savoring": a focus on feeling happy or joyous for the positive experiences of another person. Social savoring attempts to enhance one's positive feelings in response to *others'* positive experiences, amplify attention to these positive feelings toward others [37], and thereby increase a sense of connection with others [35]. In essence, people focus on how good it feels to experience happiness and gratitude for another person when that person experiences a joyous moment. Empathy has been shown to be a powerful tool for enhancing the feelings of social connectedness and self-esteem [46]. When empathy is integrated with strategies that foster savoring, the resulting combination may prove to be a useful alternative to social comparisons.

This Study

The purpose of this pilot study was to test the effectiveness and acceptance of a brief computer-delivered social savoring intervention to improve the quality of young adults' interactions with social media, ultimately increasing social connectedness and self-esteem and reducing depressive symptoms, loneliness, and contingent self-worth. The present intervention introduced social savoring as a task in which participants were asked to observe (on social media) a joyous moment experienced by someone else, reflect on what it would feel like for that person to have that experience, and then allow themselves to feel joy for that other person. To better understand how the intervention impacted participants' typical experiences with social media, we assessed their social media interactions for 7 days before receiving the intervention and 6 days after receiving the intervention.

We hypothesized that, relative to baseline, participants in the intervention group would report fewer depressive symptoms, lower loneliness, less contingent self-worth, reduced social comparison orientation, less FoMO, greater social connectedness, and higher self-esteem at posttest. At the daily level, we also expected that, on days after the intervention, participants in the intervention group would report decreased loneliness and social comparison and increased social connectedness, state self-esteem, and positive affect in response to social comparison compared with control participants.

Methods

Recruitment

Participants

Participants were 55 college students (mean age 19.29, SD 0.93 years; n=43,78% women and n=23, 42% White) enrolled in a private Southeastern university in the United States. Participants were recruited from the departmental research participation pool during the 2021 spring semester. Interested students completed a screening questionnaire to indicate where they currently resided and completed the Iowa-Netherlands Social Comparison Orientation Scale [47]. Those who were aged ≥18 years, lived in the United States, and scored >22 (out of 55) on the Social Comparison Orientation Scale—indicating a higher tendency to engage in social comparisons—were eligible to participate in the study.

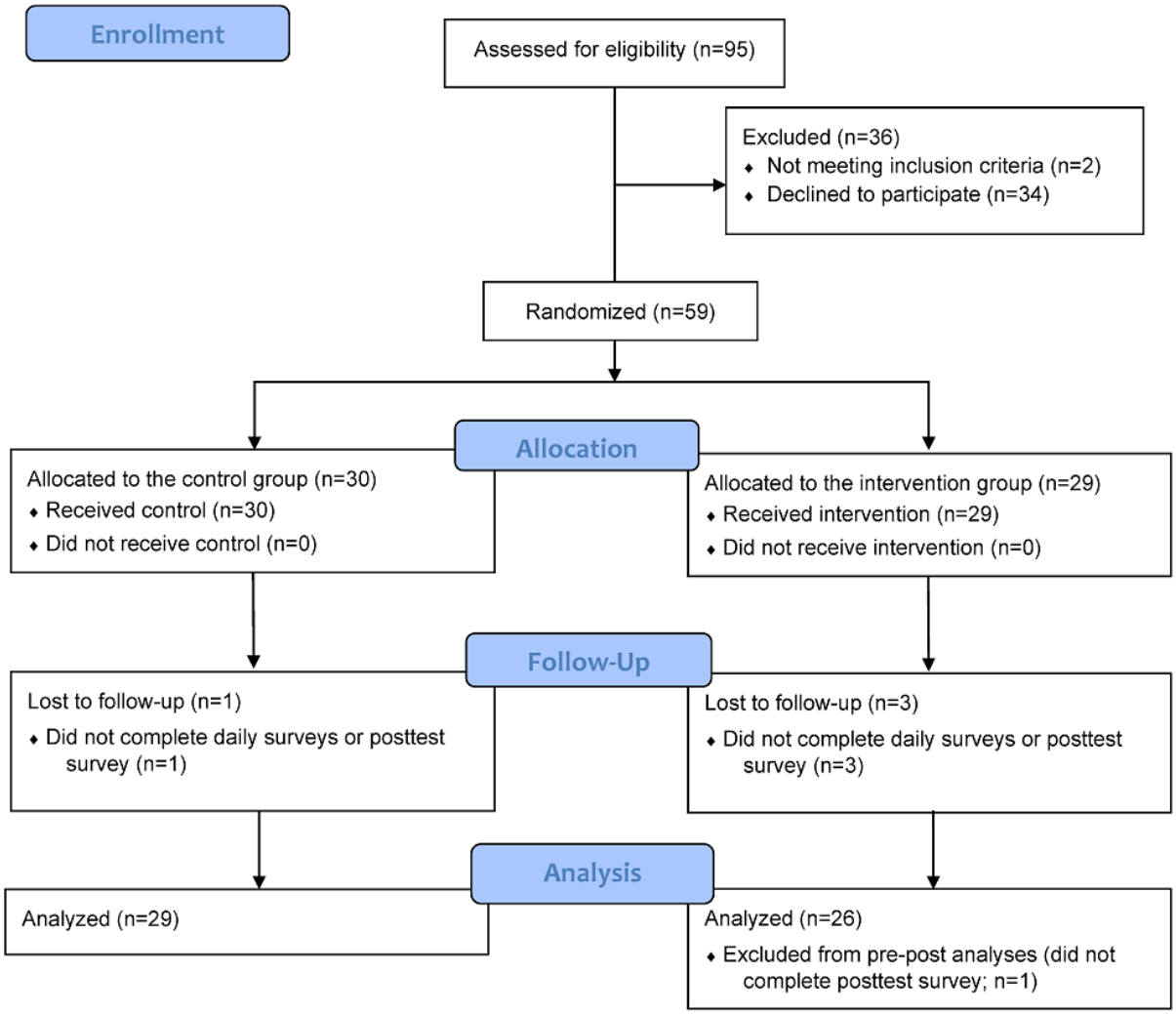
Power

A 2-week, in-person, savoring intervention [44] in college students found an effect size of Cohen d=0.41 for the difference in the average level of depression between control and intervention participants. A post hoc power analysis indicated that this study had 85% power to find an effect size of Cohen d≥0.41 with an α of .05 and sample of 55 participants.

Data Exclusion

As shown in Figure 1, of the 95 participants who completed the screening survey advertised on the web, 59 (62%) completed the study. A total of 7% (4/59) of participants completed only the baseline survey but did not complete any daily surveys or the posttest survey, and 2% (1/59) of participants from the intervention condition completed the daily surveys but did not complete the posttest survey. The latter participant was excluded from the analyses of pretest and posttest responses but was included in the analyses of daily interactions. Among the participants who completed at least 1 of the 13 daily surveys (55/59, 93%), compliance was high: participants completed an average of 12.45 (SD 1.09; range 8-13) surveys, resulting in a compliance rate of 96%.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of trial.



Ethics Approval

All procedures and materials were approved by the Duke University Campus Institutional Review Board (protocol 2021-0066) and are available on the web [48].

Procedure

This study followed an intensive longitudinal design to enhance statistical power and deepen our understanding of participants' typical interactions with social media. After providing consent to participate, participants completed the baseline survey (day 1) and were randomly assigned to a waitlist control or intervention condition. On the following day, all participants completed the first of 13 surveys (for days 2 to 14) delivered daily by text and email. After completing the seventh daily survey (day 8), participants who received the social media use intervention (26/55, 47%) viewed the intervention video and were instructed to practice the intervention skills daily (for approximately 5 minutes) and continue completing the daily surveys (days 9 to 14). Participants in the waitlist control condition (29/55, 53%) did not receive any new material but continued completing daily surveys through day 14. On the last day of the study (day 15), participants completed the posttest survey, which included items for assessing their experience with the intervention in addition to the same items as those in the baseline survey (day 1).

Materials

Overview

Participants completed a baseline survey, daily surveys, and a posttest survey, as described in the subsequent sections. McDonald ω was used to calculate the internal consistency of the baseline and posttest scales. Participants in the intervention condition watched a video describing the social savoring intervention.

Baseline Survey

Social Comparison Orientation

Social comparison was assessed using the 11-item Iowa-Netherlands Comparison Orientation Measure [47], which assesses engagement in comparisons with others (eg, "I always like to know what others in a similar situation would do"). Responses were provided on a 5-point scale (1=*disagree strongly*; 5=*agree strongly*) and were summed so that higher values indicate higher social comparison. The internal consistency was appropriate for the present sample (ω=0.76).

Self-esteem

Self-esteem was measured with 20 items from the appearance (6 items, eg, "I feel unattractive"; ω=0.89), social (7 items, eg, "I feel that others respect and admire me"; ω=0.84), and performance (7 items, eg, "I feel like I'm not doing well"; ω=0.63) subscales of the State Self-Esteem Scale [49], which assesses the extent to which one feels positively valued in the moment. Responses were reported on a 5-point scale (1=*not at all*; 5=*extremely*) and were averaged so that higher values indicate higher levels of appearance, social, and performance self-esteem.

Depression

Depression was assessed using the 21-item Beck Depression Inventory–second edition (BDI-II) [50], which asks participants to report on various symptoms of depression (eg, self-criticalness and sadness) by selecting the descriptions that best reflect their experiences over the last 2 weeks, including the day of the survey (eg, "I cry over every little thing"; ω=0.87). Response options range from 0 (indicating no presence of a symptom) to 3 (indicating the extreme presence of a symptom). BDI-II scores were summed so that higher values indicate higher levels of depressive symptoms (possible range 0-63).

Trait Loneliness

Trait loneliness was assessed using the 20-item UCLA Loneliness Scale [51], which assesses the feelings of loneliness and social isolation (eg, "I am no longer close to anyone"; ω=0.95). Participants answered the items on a 4-point scale (1=*never*; 4=*always*). Scores were averaged so that higher values indicate higher levels of loneliness.

Social Connectedness

Social connectedness was assessed using the 8-item Social Connectedness Scale [52], which assesses connectedness and closeness to others (eg, "I don't feel related to anyone"; ω=0.95). Responses were provided on a 6-point scale (1=*strongly disagree*; 6=*strongly agree*), and negatively worded items were rescored so that higher values indicate higher levels of social connectedness.

Contingencies of Self-worth

Contingencies of self-worth were assessed using the appearance and approval from others subscales of the Contingencies of Self-Worth Scale [53]. The measure assesses the extent to which one's self-esteem depends on the validation of one's appearance ("My sense of self-worth suffers whenever I think I don't look good"; ω=0.81) and approval from others (eg, "I can't respect myself if others don't respect me"; ω=0.83). Responses were provided on a 7-point scale (1=*strongly disagree*; 7=*strongly agree*), and negatively worded items were rescored so that higher values indicate higher levels of each subscale.

FoMO Scale

We measured FoMO using the 10-item FoMO Scale [29], which assesses the fear that others may be having rewarding experiences that one is "missing out" on (eg, "I fear my friends have more rewarding experiences than me"; ω=0.83). The response options ranged from 1 (*not at all true of me*) to 5 (*extremely true of me*). Scores were averaged so that higher values indicate higher levels of FoMO.

Daily Interactions

Loneliness

Loneliness was assessed using a 1-item measure of state loneliness [54] (ie, "I feel lonely"), with response options ranging from 1 (*not at all*) to 5 (*very much*).

State Self-esteem

State self-esteem was assessed using 1 item asking, "As a whole, how do you feel about yourself right now?" [49]. Responses were provided using a dial ranging from 0 (*terrible*) to 100 (*terrific*).

Daily Social Comparison

Daily social comparison was measured using 1 item asking, "To what extent did you compare yourself to the social media content you viewed today?" It was rated on a scale of 1 (*not at all*) to 5 (*a great deal*). Participants who responded with ≥2 were further asked, "Considering how you were feeling about yourself before engaging with social media, how did comparing yourself to the social media content make you feel?" Responses were provided on a 5-point scale ranging from 1 (*much worse*) to 5 (*much better*).

Social Savoring

Social savoring was assessed with 1 item asking how often participants engaged in social savoring while using social media that day. Response options ranged from 1 (*0%-20% of the time on social media*) to 5 (*80%-100% of the time on social media*). Only participants assigned to the social savoring condition saw this item, which was presented on days 9 to 14 (ie, after the delivery of the intervention).

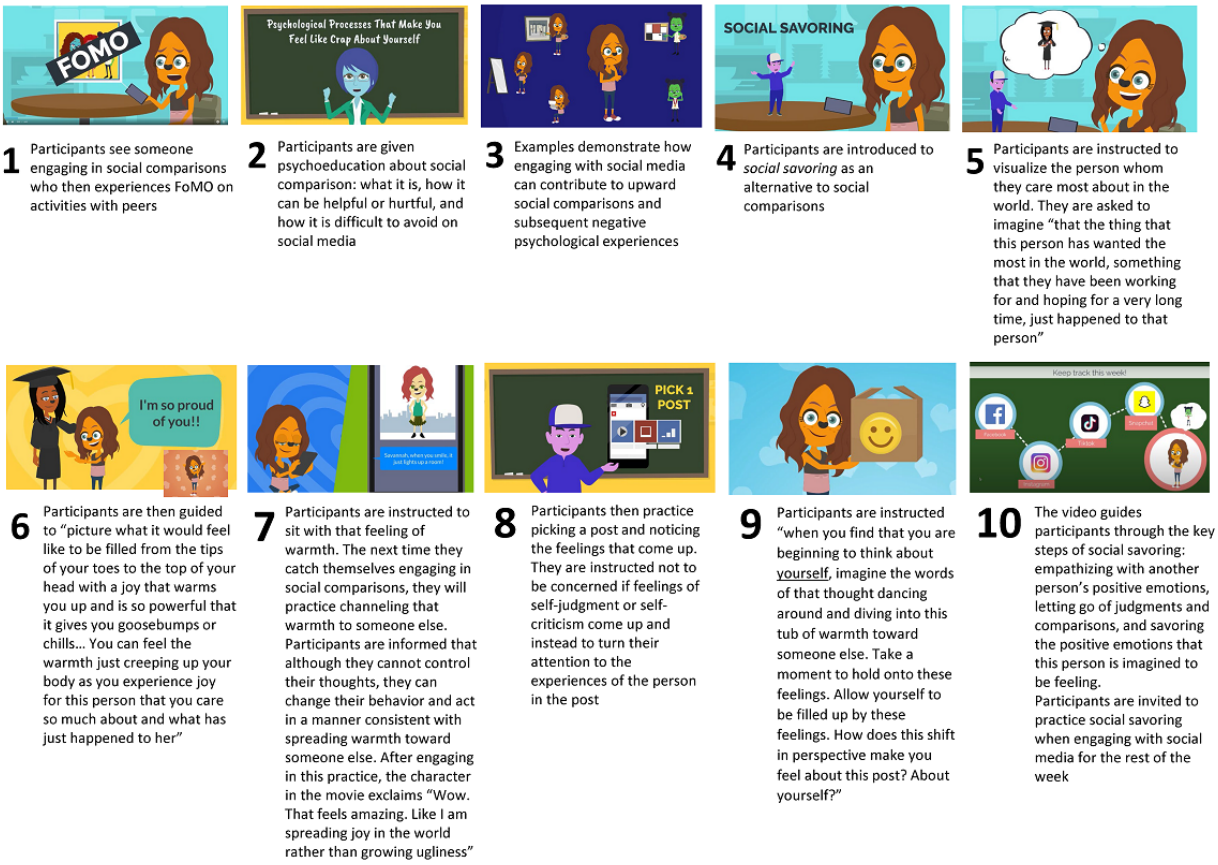
Social Media Use

Each day, participants reported the top 3 contents with which they interacted the most while using social media that day. Response options included fitness and sports, dieting and eating, food and cooking, political and societal issues, memes and comedy, beauty, nature and animals, art, and vacation.

Intervention

The intervention consisted of a 7-minute animated video presented on study day 8 to participants assigned to the intervention condition. The video introduced the concept of social comparison, explained how it manifests on social media and how social savoring can help minimize the impact of social comparisons on psychological well-being, and provided step-by-step instructions on how to practice social savoring (Figure 2). We intentionally adopted nonhuman characters in the video to reduce the possibility of physical appearance comparisons with the actors in the video. On subsequent study days (days 9 to 14), participants in the intervention condition were reminded of the basic steps of social savoring and asked to practice social savoring at least once before taking the daily survey.

Figure 2. Summary of the intervention steps. FoMO: fear of missing out.



Posttest Survey

On the final day of the study, participants completed the same battery of measures as those completed in the pretest phase: social comparison orientation (ω=0.79), appearance self-esteem (ω=83), social self-esteem (ω=0.87), performance self-esteem (ω=0.69), depression (ω=0.92), trait loneliness (ω=0.95), social connectedness (ω=0.95), appearance contingencies of self-worth (ω=0.77), approval from others contingencies of self-worth (ω=0.86), and FoMO (ω=0.91). Participants in the intervention condition also provided feedback on their experiences. Through open-ended questions, participants were asked whether social savoring was a new skill to them, what they thought about the cartoon-style intervention video, whether they agreed with the video's message, how they perceived the duration of the intervention, and whether they had any suggestions for improvement. Through multiple-choice questions, participants were asked whether they would recommend this intervention to a friend, whether they preferred a live-action skit for the intervention video, and how likely they were to use this skill in the future.

Statistical Analysis

This pilot study examined how social comparisons manifest on social media and whether social savoring can be a tool for helping buffer the negative impact of social comparisons. We performed multilevel analyses to examine whether the type of content participants daily engaged with was associated with the daily reports of social comparison and comparison affect. We used multilevel analyses to also examine whether the daily reports of social savoring were associated with the daily reports of social comparison, affect experienced from social comparison, loneliness, and state self-esteem.

To investigate participants' acceptance of the intervention, we descriptively examined their responses to the multiple-choice and open-ended follow-up items, which were completed by participants in the intervention condition. To test the effectiveness of the brief computer-delivered social savoring intervention, we performed a 2-way repeated-measures ANOVA and paired-samples 2-tailed *t* tests to test for between-condition differences from baseline to posttest in depressive symptoms, loneliness, appearance- and approval from others–contingent self-worth, social connectedness, appearance, social and performance self-esteem, social comparison orientation, and FoMO. At the daily level, we used the same analytical approach to test the effect of the intervention on participants' average daily levels of loneliness, social comparison, state self-esteem, and social comparison.

Results

Overview

As shown in Table 1, participants assigned to the control condition did not differ from those assigned to the intervention condition in any of the baseline measures.

Table 1. Equivalence between participants assigned to the control and intervention conditions.

Measure	Control (n=29)	Intervention (n=26)	Estimate[a]*(df)*	*P* value
Variable, mean (SD)				
SCO[b]	3.93 (0.08)	3.94 (0.11)	1.08 (51)	.29
PSE[c]	3.32 (0.10)	3.21 (0.12)	0.38 (51)	.71
SSE[d]	3.12 (0.17)	3.29 (0.20)	−0.87 (51)	.39
ASE[e]	3.02 (0.13)	3.13 (0.18)	−0.32 (51)	.75
Loneliness	2.08 (0.11)	1.92 (0.11)	1.06 (51)	.29
Connection	3.65 (0.18)	4.04 (0.18)	−0.92 (51)	.36
Appearance-contingent self-worth	4.96 (0.12)	4.90 (0.12)	0.68 (51)	.50
Approval from others–contingent self-worth	3.41 (0.22)	3.48 (0.23)	−0.94 (51)	.35
FoMO[f]	2.49 (0.14)	2.76 (0.15)	0.79 (51)	.31
BDI-II[g]	11.54 (8.63)	11.32 (10.15)	−0.71 (51)	.48
Demographics				
Women, n (%)	21 (72)	25 (96)	2.26 (4)	.13
White, n (%)	19 (65)	17 (65)	3.73 (4)	.44
Age (years), mean (SD)	19.47 (0.17)	19.10 (0.17)	1.52 (51)	.14

[a]Estimate refers to the between-participant 2-tailed *t* test coefficient for all scales and age and chi-square test coefficients for the women and White demographics.

[b]SCO: social comparison orientation.

[c]PSE: performance self-esteem.

[d]SSE: social self-esteem.

[e]ASE: appearance self-esteem.

[f]FoMO: fear of missing out.

[g]BDI-II: Beck Depression Inventory–second edition.

Daily Reports of Comparison

The participants completed 685 daily surveys. Participants reported interacting the most with memes and comedy content in 56.5% (387/685) of all daily surveys, political and societal issues in 23.9% (164/685) of all daily surveys, fitness and sports content in 23.5% (161/685) of all daily surveys, beauty-related content in 21.8% (149/685) of all daily surveys, food and cooking content in 19.9% (136/685) of all daily surveys, dieting and eating content in 14.3% (98/685) of all daily surveys; and vacation (68/685, 9.9%), art (47/685, 6.9%), and nature and animals (46/685, 6.7%) content in less than 10% of all daily surveys. Of the 55 participants who completed at least 1 daily survey, 53 (96%) reported engaging in social comparison on social media at least once during the daily survey period. Comparisons were reported in 63.6% (436/685) of observations. When a comparison was reported, participants most commonly reported feeling “about the same” after the comparison (245/436, 56.2% observations), followed by feeling “much worse” or “worse” (173/436, 39.7% observations). Participants reported feeling “better” or “much better” in only 4.1% (18/436) of the comparisons.

Multilevel analyses tested how the type of content was associated with the daily reports of comparison and comparison-related affect, regardless of condition (Table 2). To enhance the interpretability of the comparison direction variable, and because of our specific interest in comparisons that worsen affect, we dichotomized the outcome as 1=comparisons that worsened affect and 0=comparisons that did not alter or improved affect. The analyses were limited to the types of content that were reported in at least 10% of the observations. On days in which participants reported engagement with each fitness content, diet content, and beauty content, they reported greater comparisons and greater odds of engaging in comparisons that worsened affect relative to comparisons that did not alter or improved affect (within-person associations; beauty: odds ratio [OR] 1.86, 95% CI 1.02-3.37; fitness: OR 1.92, 95% CI 1.06-3.50; diet: OR 3.69, 95% CI 1.60-8.49). Participants who engaged with more political content or more meme content than their peers were less likely to report engaging in comparisons that worsened affect relative to comparisons that did not alter or improved affect (between-person associations). No other within- or between-person associations between the type of content engaged with and social comparisons were statistically significant.

Table 2. Associations between the type of content and comparison[a].

	Comparison			Comparison affect		
	b[b] (SE; 95% CI)	*P* value	β[c]	*b* (SE; 95% CI)	*P* value	β
Within-person association						
Fitness	*0.33 (0.13; 0.07 to 0.59)* [d]	*.01*	*.18*	*0.66 (0.31; 0.06 to 1.25)*	*.03*	*.15*
Diet	*0.73 (0.13; 0.47 to 0.98)*	*<.001*	*.32*	*1.31 (0.43; 0.47 to 2.01)*	*.002*	*.27*
Food	0.01 (0.10; −0.18 to 0.20)	.93	.00	0.35 (0.28; −0.20 to 0.90)	.21	.08
Politics	0.05 (0.11; −0.17 to 0.28)	.65	.03	−0.13 (0.37; −0.85 to 0.59)	.72	−.03
Memes	0.14 (0.09; −0.04 to 0.32)	.13	.09	0.09 (0.29; −0.49 to 0.66)	.77	.02
Beauty	*0.33 (0.10; 0.13 to 0.54)*	*.001*	*.18*	*0.62 (0.31; 0.02 to 1.22)*	*.04*	*.15*
Between-person association						
Fitness	−0.32 (0.28; −0.86 to 0.23)	.26	−.20	−0.88 (0.66; −2.17 to 0.41)	.18	−.23
Diet	0.00 (0.25; −0.50 to 0.49)	.99	.00	0.25 (0.77; −1.27 to 1.76)	.75	.05
Food	0.11 (0.39; −0.65 to 0.86)	.79	.05	−1.11 (0.91; −2.89 to 0.67)	.22	−.21
Politics	0.11 (0.26; −0.40 to 0.62)	.67	.07	*−2.02 (0.75; −3.48 to −0.56)*	*.01*	*−.55*
Memes	−0.43 (0.22; −0.86 to 0.01)	.05	−.29	*−1.75 (0.66; −3.04 to −0.46)*	*.01*	*−.51*
Beauty	−0.09 (0.27−0.63 to ; 0.45)	.74	−.05	0.48 (0.73; −0.95 to 1.91)	.51	.11

[a]The coefficients in the "comparison" column refer to the extent to which participants engaged in social comparisons by type of content. The coefficients in the "comparison affect" column correspond to the log odds that social comparisons were associated with negative affect relative to unchanged or positive affect. Negative coefficients reflect lower odds that participants felt more negative about themselves (relative to unchanged or more positive), and positive coefficients reflect greater odds that participants felt more negative about themselves (relative to unchanged or more positive).

[b]Unstandardized regression coefficient.

[c]Standardized regression coefficient.

[d]Significant results are shown in italics.

Intervention Acceptability

Of those who completed the intervention and posttest survey (25/55, 45%), all but 4% (1/25) of participants answered at least 75% of the items about the intervention. In response to the multiple-choice items, most participants indicated that they were *likely* to practice social savoring in the future (18/25, 72%), that they would recommend the skill to others (*maybe/yes*: 23/25, 92%), and that they preferred our animated video over a live-action skit (22/25, 88%).

In response to the open-ended items, most participants reported that social savoring was a new skill to them (21/25, 84%) and that they agreed with the message of the video (24/25, 96%). Most participants (22/25, 88%) responded positively to the open-ended item that asked for thoughts on the video, including that they liked the intervention video (eg, "I liked the cartoon"), found the video helpful to convey information (eg, "It nicely summarized the concept in an entertaining way"), and found the video easy to understand (eg, "The cartoon was easy to understand"). A total of 4% (1/25) of participants reported that they thought that the cartoon was "a little cheese-y but got the point across," and another 4% (1/25) indicated that they "thought it was weird that they weren't all people and were like monster people." When asked about the duration of the intervention, only 4% (1/23) of participants who responded to this open-ended question reported that it took longer than they would have liked, whereas the remaining participants (22/23, 96%) had generally positive reactions to the intervention duration, reporting—for example—that the intervention "did not take very long, [it was] very convenient" and "helped [them] reflect on [themselves] and [their] abilities" and that they "found [themselves] surprised that [they] could complete the skill so quickly and have it become normal practice."

Associations With the Reports of Social Savoring

On average (mean 2.01, SD 0.99; based on averaged response categories 1=0%-20%, 2=20%-40%, 3=40%-60%, 4=60%-80%, and 5=80%-100%), participants in the intervention condition reported engaging in social savoring for 20% to 40% of their time spent on social media during the intervention period. Nevertheless, there was variability in the extent to which participants reported engaging in social savoring such that participants indicated engaging in social savoring for 0% to 20% of their time in 45% of their reports, 20% to 40% of their time in 25% of their reports, 40% to 60% of their time in 16% of their reports, 60% to 80% of their time in 9% of their reports, and 80% to 100% of their time in 5% of their reports. Reports of social savoring increased in the days following the intervention (days 9 to 13; *b*=0.09, 95% CI 0.03-0.16, SE 0.03; *P*=.01), indicating greater social savoring engagement over time.

Multilevel models assessed whether the daily reports of social savoring were associated with the daily reports of social comparison, affect in response to social comparison, loneliness,

and state self-esteem. Associations were not significant for social comparison (*P*=.82), social comparison affect (*P*=.30), and loneliness (*P*=.19). For state self-esteem, within-person (*b*=4.88, 95% CI 1.03-8.74, SE 1.97; *P*=.01) and between-person associations (*b*=6.70, 95% CI 1.48-11.93, SE 2.67; *P*=.01) emerged. The within-person associations indicate that on days in which participants reported engaging in more social savoring, they also reported greater state self-esteem (β=.18). The between-person associations indicate that participants who engaged in more social savoring relative to their peers also reported greater state self-esteem (β=.41).

Effect of the Intervention on Pre-Post Assessments

We found main effects of time of assessment on social comparison and connectedness (refer to Table 3 for means): at posttest, participants reported lower levels of social comparison in the control (t_{27}=2.33; *P*=.01; η_p^2=1.00) and intervention (t_{24}=3.51; *P*=.002; η_p^2=0.34) conditions and higher levels of connectedness in the control (t_{27}=−6.85; *P*<.001; η_p^2=0.63) and intervention (t_{24}=−3.99; *P*=.001; η_p^2=0.40) conditions.

Two-way mixed ANOVAs (Table 3) indicated that there was a significant effect of the interaction between condition and time on the level of performance self-esteem. As shown in Figure 3, participants in the intervention condition reported higher levels of performance self-esteem at posttest than at baseline (t_{24}=−3.61; *P*=.001; η_p^2=0.35), whereas the levels of performance self-esteem among those in the control condition did not differ across time points (t_{27}=−0.42; *P*=.68; η_p^2=0.01). Although participants in the intervention condition also reported higher levels of appearance self-esteem at posttest than at baseline (t_{24}=−2.83; *P*=.01; η_p^2=0.25) and those in the control condition did not (t_{27}=−0.55; *P*=.59; η_p^2=0.01), the interaction between condition and time indicated that the between-group difference was not significant.

Table 3. Differences in the pretest and posttest scores on key variables.

Variable	Control, mean (SD)		Intervention, mean (SD)		Time × condition	
	Pretest	Posttest	Pretest	Posttest	Estimate	*P* value
SCO[a]	3.93 (0.41)	3.72 (0.53)	3.94 (0.54)	3.62 (0.52)	0.72	.40
PSE[b]	*3.32 (0.54)* [c]	*3.35 (0.50)*	*3.21 (0.58)*	*3.52 (0.43)*	*6.19*	*.02*
SSE[d]	3.12 (0.90)	3.15 (0.88)	3.29 (0.99)	3.42 (0.84)	0.40	.53
ASE[e]	3.02 (0.71)	3.07 (0.77)	3.13 (0.89)	3.41 (0.69)	3.62	.08
Loneliness	2.08 (0.57)	1.92 (0.56)	1.92 (0.53)	1.88 (0.44)	0.21	.65
Connection	3.65 (0.98)	4.19 (1.03)	4.04 (0.89)	4.48 (0.92)	0.51	.48
Appearance-contingent self-worth	4.96 (0.66)	4.96 (0.71)	4.90 (0.62)	4.88 (0.49)	0.01	.92
Approval of others–contingent self-worth	3.41 (1.15)	3.52 (1.19)	3.49 (1.14)	3.51 (1.35)	0.14	.71
FoMO[f]	2.49 (0.75)	2.42 (0.88)	2.76 (0.76)	2.64 (0.89)	0.09	.77
BDI-II[g]	11.54 (8.63)	11.32 (10.15)	12.72 (7.92)	11.08 (7.35)	0.89	.35

[a]SCO: social comparison orientation.

[b]PSE: performance self-esteem.

[c]Significant results are shown in italics.

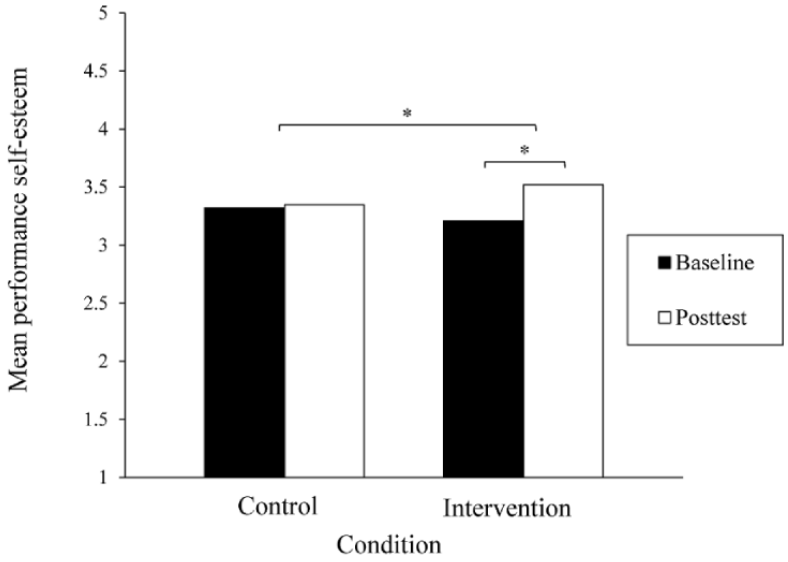
[d]SSE: social self-esteem.

[e]ASE: appearance self-esteem.

[f]FoMO: fear of missing out.

[g]BDI-II: Beck Depression Inventory–second edition.

Figure 3. Mean performance self-esteem by time and condition. Higher scores indicate higher levels of performance self-esteem. Significant differences are marked with an asterisk.



Effect of the Intervention on Daily Interactions With Social Media

To examine the effect of the intervention on participants' daily experiences with social media, we separately averaged their responses to the items on loneliness, social comparison, state self-esteem, and the impact of social comparison on days 2 to 8 (ie, before the introduction of the intervention) and days 9 to 14 (ie, after the introduction of the intervention). Thus, we created 2 composite scores for each daily variable: one referent to the days before the intervention and another referent to the days after the intervention. We then performed 2-way within-participant ANOVAs to test whether participants in the intervention condition differed from those in the control condition in their average daily experiences on days 2 to 8 and days 9 to 14.

As shown in Table 4, participants who watched the intervention video did not report any changes in loneliness (P=.24), state self-esteem (P=.92), comparisons on social media (P=.82), or feelings associated with comparisons (P=.76).

Table 4. Effects of the intervention on the daily reports of loneliness, self-esteem, social comparison, and comparison affect.

	State loneliness	State self-esteem	Social media comparison	Comparison affect[a]
Control				
Pretest, mean[b] (SD)	28.28 (21.26)	65.98 (15.33)	2.12 (0.64)	2.73 (0.47)
Posttest, mean[c] (SD)	23.85 (22.07)	65.84 (20.72)	2.01 (0.70)	2.69 (0.39)
F test (*df*)	1.88 (1)	0.14 (1)	1.54 (1)	0.27 (1)
P value	.18	.71	.23	.61
Intervention				
Pretest, mean (SD)	29.34 (20.30)	66.14 (15.17)	1.87 (0.48)	2.59 (0.38)
Posttest, mean (SD)	25.75 (17.96)	65.88 (18.00)	1.84 (0.61)	2.62 (0.37)
F test (*df*)	1.45 (1)	0.01 (1)	0.50 (1)	0.10 (1)
P value	.24	.92	.82	.76
Condition × time				
F test (*df*)	0.36 (1,53)	0.11 (1,53)	0.31 (1,53)	0.34 (1,50)
P value	.85	.74	.58	.56
η^2	0.00	0.00	0.01	0.01

[a]"Comparison affect" in this table refers to how positively participants felt about themselves after engaging in social comparisons; higher scores indicate feeling more positively, and lower scores indicate feeling more negatively.

[b]Pretest mean=average score on days 2 to 8 of the study.

[c]Posttest mean=average score on days 9 to 14 of the study.

Discussion

Principal Findings and Comparison With Prior Work

Overview

This pilot study assessed the acceptability, feasibility, and effectiveness of a novel social savoring intervention for reducing potentially harmful interactions with social media. Most participants reported positive perceptions of the web-based intervention and that they were likely to use the social savoring skill in the future. We found evidence of the effectiveness of the intervention in key psychosocial outcomes. Overall, we provide foundational research for future investigations on using social savoring to mitigate the negative effects of social media use.

Social Savoring Intervention

Our results show that this novel social savoring intervention is both acceptable and feasible. Participants generally reported a positive perception of the social savoring intervention in response to direct questions about their attitudes toward the intervention. That participants sustained high compliance with daily activities and increased their daily use of the (largely novel) social savoring skill after its introduction is further evidence of the acceptability and feasibility of the intervention. In addition, most participants expressed an interest in using the social savoring skill in the future and indicated that they would likely recommend this skill to a friend.

Our data provide preliminary evidence of the effectiveness of this intervention in promoting mental health. We found evidence of both between- and within-person associations between social savoring and self-esteem: participants who reported more social savoring than their peers also reported higher state self-esteem (between-person association), and participants who reported more social savoring than their own average on a given day reported higher state self-esteem on that day (within-person association). As participants reported savoring and self-esteem only once per day, the directionality of these associations is unclear. Social savoring involves focusing on pleasant experiences [35,36], and those who feel more positively about themselves may have an easier time engaging in this practice [55]. By contrast, savoring has previously been found to boost positive affect [56]. Future research using multiple same-day assessments can elucidate the directionality of these associations. Importantly, results from the posttest provide some support that the intervention was effective in increasing self-esteem in that performance self-esteem increased from pretest to posttest for the intervention group but not for the control group. The significant association in this domain, and not the others, could be an artifact of our sample of undergraduates coming from a highly selective university, making the intervention particularly effective in fostering the savoring of competence-related accomplishments. Alternatively, this finding may indicate that engaging in social savoring enhances positive self-evaluations via the prosocial nature of the activity.

Participants in both the control and intervention groups reported reduced social comparison at posttest compared with baseline. As demonstrated in prior research, asking participants to reflect and report on their psychological states can increase self-awareness and cause behavior changes [57]. The possible increases in self-monitoring in the control group may have inadvertently muted intervention effects.

Engaging in social savoring was associated with daily self-esteem but not with the reports of general social comparison, social comparison affect, loneliness, or depression. The lack of association with social comparison could indicate that social savoring does not necessarily reduce comparisons in general. Additional practice with social savoring may be necessary to shift people's initial tendency from engaging in a comparison to savoring the positive feelings of another person's experience. Alternatively, reducing the frequency of comparisons may not be necessary to buffer the negative effects of social comparisons. In fact, the potentially positive effects of social comparisons (eg, feelings of optimism, admiration, and inspiration) are consistent with the goals of social savoring [37].

Importantly, savoring was not associated with the consequences of social comparisons in that engaging in more savoring was unrelated to how participants felt about themselves after comparisons with social media content. This is consistent with our expectation that savoring operates differently from social comparisons because savoring shifts participants' focus to the positive experiences of others rather than to what those experiences mean for themselves. The lack of association between savoring and loneliness could be because social media content is not exclusively socially oriented; for example, it can be appearance oriented [16] or career oriented [15]. Thus, effects may have been absent when there was a mismatch between the target of savoring and the measured outcome. Future research should incorporate daily measures of well-being that are specific to various domains to assess potential differences.

Contrary to prior work [41], we did not find an association between social savoring and depression. However, this finding may be a consequence of a misalignment between the wording of the BDI-II and the duration of the study. Specifically, the BDI-II items asked participants to think about their experiences over the previous 2 weeks [50], which included both the time before and after learning the social savoring skill for those in the intervention condition. Although participants in the intervention condition descriptively reported lower levels of depression at posttest compared with baseline, it is possible that the perceived change was dampened by the inclusion of psychological experiences that occurred before the intervention took place.

Consistent with the evidence that social media is visually oriented and evokes appearance comparisons [16], participants in the intervention group showed higher appearance self-esteem in the posttest than in the pretest, whereas participants in the control condition did not; however, the difference between the 2 groups was not statistically significant. Given the finding that appearance-oriented content was especially likely to evoke comparisons and comparisons that worsened affect, it is possible that our sample size was too small to detect between-group differences in the effect of the intervention on participants' perceptions of the highly curated and attractive visual content

of others. Future studies with larger samples may be able to detect this difference if present.

Finally, the absence of baseline-to-posttest changes may be attributable to the fact that some measured domains tend to be stable, trait-level characteristics. Social comparison orientation, contingencies of self-worth, and FoMO may simply be too stable for a short-term intervention to evoke meaningful changes [53]. Our lack of effects could be because participants need to build their social savoring skill over time to impact trait-level characteristics or because our participants geared social savoring toward specific domains, such as responses to appearance-oriented content. Future research assessing longer-term social savoring practice and investigating more specific domains (eg, appearance esteem) is needed to test these conjectures.

Comparison Experiences

Nearly all participants engaged in social comparison on social media at least once daily during the data collection period. Participants most commonly reported that comparisons did not change their affect (245/436, 56.2%); comparisons that worsened affect were reported in 39.7% (173/436) of the observations, and comparisons that improved affect were quite rare (18/436, 4.1%). Although past research has confirmed that comparisons that improve affect are rare on social media, research has also found that comparisons that worsen affect are more common than those that do not alter it [12]. Our finding of less frequent comparisons that worsen affect relative to comparisons that do not alter affect could be owing to differences in methodology: whereas past research [12] has exclusively assessed appearance comparisons, we examined comparisons more generally. Past research with adolescents indicates that people tend to highlight their physical appearance in their social media posts [58], which may inflate the likelihood of engaging in comparisons that worsen affect, specifically in the appearance domain [59]. Because we assessed a range of comparison domains (beyond just appearance), we would expect to observe a comparatively lower number of comparisons that worsen affect. Furthermore, our results suggest that college students may be less likely to engage in comparisons that worsen affect in other domains, such as food and politics.

Participants reported most often engaging with meme content, although fitness and sports, food, political and societal issues, and beauty content were also reported in a notable minority of observations. Regardless of the condition, on days in which participants engaged with fitness, diet, and beauty content, they reported a greater likelihood of engaging in comparisons and comparisons that worsen affect. These types of content are unified in that they are appearance oriented, suggesting that these types of content may be particularly potent in facilitating upward appearance comparisons. Appearance comparisons that worsen affect may be especially harmful to well-being [12], raising concerns about how these comparisons affect the in-the-moment well-being of social media users. Importantly, these associations were not observed at the between-person level, indicating that participants who engaged with more appearance-related content, in general, did not necessarily also engage in more comparisons or report worse affect in response to comparisons. Nonetheless, future longitudinal research can better ascertain how these appearance-related comparisons accumulate in the relationship between these comparisons and well-being over time.

Participants who engaged with more politics- and meme-related content than their peers reported a lower likelihood of engaging in comparisons that worsened affect (and, therefore, a greater likelihood of experiencing comparisons that did not change or improved affect). It is possible that these types of content portray others' misfortunes, which could enhance the behavioral tendency to perceive oneself as better off than others. The findings of this study highlight a need to adopt a nuanced perspective that integrates various disciplines in studying how comparisons unfold for different types of content. Although viewing certain types of content may be beneficial in one domain (eg, memes associated with fewer comparisons that worsen affect), engagement with these types of content may be maladaptive in other domains (eg, perpetuating stereotypes).

Limitations and Future Directions

The findings of this study are limited to a small sample of primarily White and women college students enrolled in a selective private Southeastern university in the United States. We attempted to replicate these findings in a sample of community members recruited through social media sites (eg, Instagram [Meta Platforms Inc], Reddit [Reddit Inc], Facebook [Meta Platforms Inc], and Twitter [Twitter Inc]). However, recruitment through these platforms was limited, and it was difficult to gain sufficient interest and retention from the community members. Indeed, of the 574 community members who completed the baseline survey, 48 (8.4%) provided good quality data (ie, no "straight-lining," no nonsense responses to open-ended items, and no spam or repeated IP addresses), and only 11 (1.9%) completed the daily surveys to sufficient extent, a limited response rate that precluded comparisons with the present findings. These findings suggest the need for the development of innovative retention strategies if this intervention is to be implemented on a larger scale.

In addition, factors outside the content of the intervention appear to have influenced participants' self-reported experiences. Specifically, the evidence that loneliness and social comparison levels decreased in both intervention and control groups suggests that merely reporting their own social media use daily influenced participants' perceptions of these domains—and potentially how they interacted with social media. Another limitation is that we were unable to verify the true extent to which participants practiced the skills. Future studies using interventions such as the one used in this study may benefit from including a written practice during the introduction to the intervention, as done in the study by Hurley and Kwon [44], and daily written exercises to verify that participants spent time practicing social savoring.

Finally, the changes brought about by the COVID-19 pandemic may have shifted how participants interacted with social media such that social media use may have generally supported social connections. Given the restrictions imposed on participants at the time of the survey (eg, primarily web-based classes and restrictions on the size of gatherings), FoMO may have been

less common than it would have been before the pandemic and thus less flexible to change.

Despite these limitations, the initial evidence from this pilot study suggests that a web-based social savoring intervention may help minimize the potentially harmful consequences of social media use. A fruitful avenue for future work is to examine the effectiveness and acceptance of this intervention for people diagnosed with disorders characterized by higher levels of comparison with others (eg, eating disorders). Specifically, savoring the joys of others may be a helpful strategy to feel more positively about one's own skills and physical appearance.

Future studies should examine the effectiveness, feasibility, and acceptance of this intervention in a larger and more diverse sample. Given the evidence of an association between social comparisons on social media and eating disorders, depression, and anxiety [16,21,54,60], it would be valuable to examine the effectiveness of the present intervention among those with such psychopathologies. With changes such as those described earlier (eg, adding a daily written practice), a web-based intervention may be more easily delivered and accessed by a wider portion of the population.

Conclusions

In summary, a 7-minute animated intervention video that taught the skill of social savoring improved self-evaluation in performance domains. Participants valued this prosocial approach to engaging with social media content, as evidenced by the increased practice of this skill over time and their purported willingness to share this technique with friends. Such promising results with such minimal intervention suggest that the use of social savoring as more tightly integrated with social media engagement (eg, cued by certain social media content, behaviors, and platforms) may result in a more potent invention capable of enhancing prosocial feelings and social connections while mitigating the harm of social comparisons.

Acknowledgments

The authors are grateful for the generous support from the Charles Lafitte Foundation for Research in Psychology and Neuroscience.

Conflicts of Interest

None declared.

References

1. Auxier B, Anderson M. Social media use in 2021: a majority of Americans say they use YouTube and Facebook, while use of Instagram, Snapchat and TikTok is especially common among adults under 30. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2022-12-09]
2. McKenna KY, Green AS, Gleason ME. Relationship formation on the internet: what's the big attraction? J Soc Issues 2002 Dec 17;58(1):9-31 [FREE Full text] [doi: 10.1111/1540-4560.00246]
3. Blais JJ, Craig WM, Pepler D, Connolly J. Adolescents online: the importance of internet activity choices to salient relationships. J Youth Adolescence 2007 Dec 13;37(5):522-536. [doi: 10.1007/s10964-007-9262-7]
4. Valkenburg PM, Peter J. Preadolescents' and adolescents' online communication and their closeness to friends. Dev Psychol 2007 Mar;43(2):267-277. [doi: 10.1037/0012-1649.43.2.267] [Medline: 17352538]
5. Michikyan M, Subrahmanyam K. Social networking sites: implications for youth. In: Yan Z, editor. Encyclopedia of Cyber Behavior. Volumes I - III. Hershey, PA, USA: IGI Global; 2012:132-147.
6. Steers ML, Wickham RE, Acitelli LK. Seeing everyone else's highlight reels: how Facebook usage is linked to depressive symptoms. J Soc Clin Psychol 2014 Oct;33(8):701-731 [FREE Full text] [doi: 10.1521/jscp.2014.33.8.701]
7. Festinger L. A theory of social comparison processes. Hum Relat 1954 May 01;7(2):117-140 [FREE Full text] [doi: 10.1177/001872675400700202]
8. Meier A, Schäfer S. Positive side of social comparison on social network sites: how envy can drive inspiration on Instagram. Cyberpsychol Behav Soc Netw 2018 Jul;21(7):411-417. [doi: 10.1089/cyber.2017.0708] [Medline: 29995526]
9. Haferkamp N, Krämer NC. Social comparison 2.0: examining the effects of online profiles on social-networking sites. Cyberpsychol Behav Soc Netw 2011 May;14(5):309-314. [doi: 10.1089/cyber.2010.0120] [Medline: 21117976]
10. Vogel EA, Rose JP, Roberts LR, Eckles K. Social comparison, social media, and self-esteem. Psychol Pop Media Cult 2014;3(4):206-222 [FREE Full text] [doi: 10.1037/ppm0000047]
11. Vogel EA, Rose JP, Okdie BM, Eckles K, Franz B. Who compares and despairs? The effect of social comparison orientation on social media use and its outcomes. Pers Individ Dif 2015 Nov;86:249-256 [FREE Full text] [doi: 10.1016/j.paid.2015.06.026]
12. Fardouly J, Pinkus RT, Vartanian LR. The impact of appearance comparisons made through social media, traditional media, and in person in women's everyday lives. Body Image 2017 Mar;20:31-39. [doi: 10.1016/j.bodyim.2016.11.002] [Medline: 27907812]
13. Corning AF, Krumm AJ, Smitham LA. Differential social comparison processes in women with and without eating disorder symptoms. J Couns Psychol 2006 Jul;53(3):338-349 [FREE Full text] [doi: 10.1037/0022-0167.53.3.338]
14. Pinkasavage E, Arigo D, Schumacher LM. Social comparison, negative body image, and disordered eating behavior: the moderating role of coping style. Eat Behav 2015 Jan;16:72-77. [doi: 10.1016/j.eatbeh.2014.10.014] [Medline: 25464070]

15. Fukubayashi N, Fuji K. Social comparison on social media increases career frustration: a focus on the mitigating effect of companionship. Front Psychol 2021 Oct 15;12:720960 [FREE Full text] [doi: 10.3389/fpsyg.2021.720960] [Medline: 34721175]
16. Holland G, Tiggemann M. A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. Body Image 2016 Jun;17:100-110. [doi: 10.1016/j.bodyim.2016.02.008] [Medline: 26995158]
17. Myers TA, Crowther JH. Social comparison as a predictor of body dissatisfaction: a meta-analytic review. J Abnorm Psychol 2009 Nov;118(4):683-698. [doi: 10.1037/a0016763] [Medline: 19899839]
18. Baker N, Ferszt G, Breines JG. A qualitative study exploring female college students' Instagram use and body image. Cyberpsychol Behav Soc Netw 2019 Apr;22(4):277-282. [doi: 10.1089/cyber.2018.0420] [Medline: 30855190]
19. Javornik A, Marder B, Barhorst JB, McLean G, Rogers Y, Marshall P, et al. 'What lies behind the filter?' Uncovering the motivations for using augmented reality (AR) face filters on social media and their effect on well-being. Comput Human Behav 2022 Mar;128:107126 [FREE Full text] [doi: 10.1016/j.chb.2021.107126]
20. Chou HT, Edge N. "They are happier and having better lives than I am": the impact of using Facebook on perceptions of others' lives. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):117-121. [doi: 10.1089/cyber.2011.0324] [Medline: 22165917]
21. Seabrook EM, Kern ML, Rickard NS. Social networking sites, depression, and anxiety: a systematic review. JMIR Ment Health 2016 Nov 23;3(4):e50 [FREE Full text] [doi: 10.2196/mental.5842] [Medline: 27881357]
22. Fardouly J, Rapee RM. The impact of no-makeup selfies on young women's body image. Body Image 2019 Mar;28:128-134. [doi: 10.1016/j.bodyim.2019.01.006] [Medline: 30665030]
23. Slater A, Cole N, Fardouly J. The effect of exposure to parodies of thin-ideal images on young women's body image and mood. Body Image 2019 Jun;29:82-89. [doi: 10.1016/j.bodyim.2019.03.001] [Medline: 30870744]
24. Tiggemann M, Anderberg I. Social media is not real: the effect of 'Instagram vs reality' images on women's social comparison and body image. New Media Soc 2020 Dec;22(12):2183-2199 [FREE Full text] [doi: 10.1177/1461444819888720]
25. Mussweiler T, Rüter K, Epstude K. The ups and downs of social comparison: mechanisms of assimilation and contrast. J Pers Soc Psychol 2004 Dec;87(6):832-844. [doi: 10.1037/0022-3514.87.6.832] [Medline: 15598109]
26. Kang J, Liu B. A similarity mindset matters on social media: using algorithm-generated similarity metrics to foster assimilation in upward social comparison. Soc Media Soc 2019 Dec;5(4):205630511989088 [FREE Full text] [doi: 10.1177/2056305119890884]
27. Neff K. Self-compassion: an alternative conceptualization of a healthy attitude toward oneself. Self Identity 2003 Apr;2(2):85-101 [FREE Full text] [doi: 10.1080/15298860309032]
28. Neff KD, Dahm KA. Self-compassion: what it is, what it does, and how it relates to mindfulness. In: Ostafin BD, Robinson MD, Meier BP, editors. Handbook of Mindfulness and Self-Regulation. New York, NY, USA: Springer; Oct 13, 2015:121-137.
29. Baker ZG, Krieger H, LeRoy AS. Fear of missing out: relationships with depression, mindfulness, and physical symptoms. Transl Issues Psychol Sci 2016;2(3):275-282 [FREE Full text] [doi: 10.1037/tps0000075]
30. Przybylski AK, Murayama K, DeHaan CR, Gladwell V. Motivational, emotional, and behavioral correlates of fear of missing out. Comput Human Behav 2013 Jul;29(4):1841-1848 [FREE Full text] [doi: 10.1016/j.chb.2013.02.014]
31. Jones A, Hook M, Podduturi P, McKeen H, Beitzell E, Liss M. Mindfulness as a mediator in the relationship between social media engagement and depression in young adults. Pers Individ Dif 2022 Feb;185:111284 [FREE Full text] [doi: 10.1016/j.paid.2021.111284]
32. Keng SL, Ang Q. Effects of mindfulness on negative affect, body dissatisfaction, and disordered eating urges. Mindfulness 2019 Mar 20;10(9):1779-1791 [FREE Full text] [doi: 10.1007/s12671-019-01146-2]
33. Albertson ER, Neff KD, Dill-Shackleford KE. Self-compassion and body dissatisfaction in women: a randomized controlled trial of a brief meditation intervention. Mindfulness 2015;6(3):444-454 [FREE Full text] [doi: 10.1007/s12671-014-0277-3]
34. Barron AM, Krumrei-Mancuso EJ, Harriger JA. The effects of fitspiration and self-compassion Instagram posts on body image and self-compassion in men and women. Body Image 2021 Jun;37:14-27. [doi: 10.1016/j.bodyim.2021.01.003] [Medline: 33556914]
35. Bryant FB, Smith JL. Appreciating life in the midst of adversity: savoring in relation to mindfulness, reappraisal, and meaning. Psychol Inq 2015 Nov 24;26(4):315-321 [FREE Full text] [doi: 10.1080/1047840x.2015.1075351]
36. Bryant FB, Veroff J. Savoring: A New Model of Positive Experience. Mahwah, NJ, USA: Lawrence Erlbaum Associates; 2007.
37. Smith RH. Assimilative and contrastive emotional reactions to upward and downward social comparisons. In: Suls J, Wheeler L, editors. Handbook of Social Comparison: Theory and Research. New York, NY, USA: Springer; 2000:173-200.
38. Smith JL, Bryant FB. The benefits of savoring life: savoring as a moderator of the relationship between health and life satisfaction in older adults. Int J Aging Hum Dev 2016 Dec;84(1):3-23. [doi: 10.1177/0091415016669146] [Medline: 27655957]
39. Smith JL, Hanni AA. Effects of a savoring intervention on resilience and well-being of older adults. J Appl Gerontol 2019 Jan;38(1):137-152. [doi: 10.1177/0733464817693375] [Medline: 28380722]
40. Smith JL, Hollinger-Smith LH. Savoring, resilience, and psychological well-being in older adults. Aging Ment Health 2015;19(3):192-200. [doi: 10.1080/13607863.2014.986647] [Medline: 25471325]

41. Irvin KM, Bell DJ, Steinley D, Bartholow BD. The thrill of victory: savoring positive affect, psychophysiological reward processing, and symptoms of depression. Emotion 2022 Sep;22(6):1281-1293. [doi: 10.1037/emo0000914] [Medline: 33252936]
42. Silton RL, Kahrilas IJ, Skymba HV, Smith J, Bryant FB, Heller W. Regulating positive emotions: implications for promoting well-being in individuals with depression. Emotion 2020 Feb;20(1):93-97. [doi: 10.1037/emo0000675] [Medline: 31961185]
43. Moskowitz JT, Hult JR, Duncan LG, Cohn MA, Maurer S, Bussolari C, et al. A positive affect intervention for people experiencing health-related stress: development and non-randomized pilot test. J Health Psychol 2012 Jul;17(5):676-692 [FREE Full text] [doi: 10.1177/1359105311425275] [Medline: 22021272]
44. Hurley DB, Kwon P. Results of a study to increase savoring the moment: differential impact on positive and negative outcomes. J Happiness Stud 2011 Jul 2;13(4):579-588 [FREE Full text] [doi: 10.1007/s10902-011-9280-8]
45. Yu SC, Sheldon KM, Lan WP, Chen JH. Using social network sites to boost savoring: positive effects on positive emotions. Int J Environ Res Public Health 2020 Sep 02;17(17):6407 [FREE Full text] [doi: 10.3390/ijerph17176407] [Medline: 32887521]
46. Fredrickson BL. Promoting positive affect. In: Eid M, Larsen RJ, editors. The Science of Subjective Well-Being. New York, NY, USA: Guilford Press; 2008:449-468.
47. Gibbons FX, Buunk BP. Individual differences in social comparison: development of a scale of social comparison orientation. J Pers Soc Psychol 1999 Jan;76(1):129-142. [doi: 10.1037//0022-3514.76.1.129] [Medline: 9972558]
48. Andrade FC, Erwin SR, Burnell K, Jackson J, Nicholas J, Zucker NL. A social savoring intervention on social comparisons on social media. Open Science Framework. 2022 Jun 23. URL: https://www.osf.io/6ys4w [accessed 2022-06-22]
49. Heatherton TF, Polivy J. Development and validation of a scale for measuring state self-esteem. J Pers Soc Psychol 1991;60(6):895-910 [FREE Full text] [doi: 10.1037/0022-3514.60.6.895]
50. Beck AT, Steer RA, Ball R, Ranieri WF. Comparison of beck depression inventories -IA and -II in psychiatric outpatients. J Pers Assess 1996 Dec;67(3):588-597. [doi: 10.1207/s15327752jpa6703_13] [Medline: 8991972]
51. Russell D, Peplau LA, Ferguson ML. Developing a measure of loneliness. J Pers Assess 1978 Jun;42(3):290-294. [doi: 10.1207/s15327752jpa4203_11] [Medline: 660402]
52. Lee RM, Robbins SB. Measuring belongingness: the social connectedness and the social assurance scales. J Couns Psychol 1995 Apr;42(2):232-241 [FREE Full text] [doi: 10.1037/0022-0167.42.2.232]
53. Crocker J, Luhtanen RK, Cooper ML, Bouvrette A. Contingencies of self-worth in college students: theory and measurement. J Pers Soc Psychol 2003 Nov;85(5):894-908. [doi: 10.1037/0022-3514.85.5.894] [Medline: 14599252]
54. van Roekel E, Verhagen M, Engels RC, Scholte RH, Cacioppo S, Cacioppo JT. Trait and state levels of loneliness in early and late adolescents: examining the differential reactivity hypothesis. J Clin Child Adolesc Psychol 2018;47(6):888-899. [doi: 10.1080/15374416.2016.1146993] [Medline: 27191708]
55. Wood JV, Heimpel SA, Michela JL. Savoring versus dampening: self-esteem differences in regulating positive affect. J Pers Soc Psychol 2003 Sep;85(3):566-580. [doi: 10.1037/0022-3514.85.3.566] [Medline: 14498791]
56. Jose PE, Lim BT, Bryant FB. Does savoring increase happiness? A daily diary study. J Posit Psychol 2012 May;7(3):176-187 [FREE Full text] [doi: 10.1080/17439760.2012.671345]
57. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
58. Chua TH, Chang L. Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. Comput Human Behav 2016 Feb;55:190-197 [FREE Full text] [doi: 10.1016/j.chb.2015.09.011]
59. Fuller-Tyszkiewicz M. Body image states in everyday life: evidence from ecological momentary assessment methodology. Body Image 2019 Dec;31:245-272. [doi: 10.1016/j.bodyim.2019.02.010] [Medline: 30852080]
60. Feinstein BA, Hershenberg R, Bhatia V, Latack JA, Meuwly N, Davila J. Negative social comparison on Facebook and depressive symptoms: rumination as a mechanism. Psychol Pop Media Cult 2013;2(3):161-170 [FREE Full text] [doi: 10.1037/a0033111]

Abbreviations

BDI-II: Beck Depression Inventory–second edition
FoMO: fear of missing out
OR: odds ratio

Edited by J Torous; submitted 18.08.22; peer-reviewed by J Thrul, B Koronczai; comments to author 18.11.22; revised version received 12.01.23; accepted 30.01.23; published 28.04.23.

Please cite as:
Andrade FC, Erwin S, Burnell K, Jackson J, Storch M, Nicholas J, Zucker N
Intervening on Social Comparisons on Social Media: Electronic Daily Diary Pilot Study
JMIR Ment Health 2023;10:e42024
URL: https://mental.jmir.org/2023/1/e42024
doi:10.2196/42024
PMID:37115607

Original Paper

Evaluating a Smartphone App (MeT4VeT) to Support the Mental Health of UK Armed Forces Veterans: Feasibility Randomized Controlled Trial

Steven Parkes[1], BSc, MSc; Bethany Croak[1], BSc, MSc; Samantha K Brooks[1], BSc, PhD; Sharon A M Stevelink[1,2], BSc, MSc, PhD; Daniel Leightley[1], BSc, MSc, PhD; Nicola T Fear[1,3], BSc, MSc, DPhil; Laura Rafferty[1], BSc, PhD; Neil Greenberg[1], MD

[1]King's Centre for Military Health Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[2]Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[3]Academic Department of Military Mental Health, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

Corresponding Author:
Steven Parkes, BSc, MSc
King's Centre for Military Health Research
Institute of Psychiatry, Psychology & Neuroscience
King's College London
10 Cutcombe Road
London, SE5 9RJ
United Kingdom
Phone: 44 0207848 ext 5335
Email: steven.j.parkes@kcl.ac.uk

Abstract

Background: Previous research demonstrates that less than 50% of military veterans experiencing mental health difficulties seek formal support. Veterans often struggle to identify problems as mental health difficulties. In addition, they may fail to recognize the need for support before reaching a crisis point and face difficulties navigating care pathways to access support.

Objective: A feasibility trial was conducted to assess a novel digital smartphone app (Mental Health Toolkit for Veterans Project [MeT4VeT]) for UK Armed Forces (UKAF) veterans experiencing mental health difficulties. The trial aimed to explore the feasibility and acceptability of trial procedures for a later randomized controlled trial (RCT) and to assess the acceptability of the MeT4VeT app.

Methods: Participants were recruited at UK military medical centers, by advertising on social media, and through veteran third-sector organizations between February and November 2021, and assessed for eligibility (male, owned a smartphone, served at least 2 years in the UKAF, left the UKAF within the last 2 years, not undertaking formal mental health treatment). Eligible participants were assigned, on a 1:1 ratio, to either the intervention group (full app) or a control group (noninteractive app with signposting information). Three key objectives were determined a priori to assess the practicality of running an RCT including an assessment of recruitment and retention, evaluation of the technical app delivery and measurement processes, and acceptability and usability of the intervention.

Results: In total, 791 individuals completed the participant information sheet, of which 261 (33%) were ineligible, 377 (48%) declined or were unable to be contacted for consent, and 103 (13%) did not download the app or complete the baseline measures. Of this, 50 participants completed baseline measures and were randomly assigned to the intervention group (n=24) or the control group (n=26). The trial was effective at enabling both the technical delivery of the intervention and collection of outcome measures, with improvements in mental health demonstrated for the intervention group from baseline to the 3-month follow-up. Recruitment and retention challenges were highlighted with only 50 out of the 530 eligible participants enrolled in the trial. The acceptability and usability of the MeT4VeT app were generally supported, and it was reported to be a useful, accessible way for veterans to monitor and manage their mental health.

Conclusions: The results highlighted that further work is needed to refine recruitment processes and maintain engagement with the app. Following this, an RCT can be considered to robustly assess the ability of the app to positively affect mental health outcomes indicated within this trial.

Trial Registration: ClinicalTrials.gov NCT05993676; https://clinicaltrials.gov/ct2/show/NCT05993676

(JMIR Ment Health 2023;10:e46508) doi:10.2196/46508

KEYWORDS

military; veteran; mental health; military to civilian transition; digital health; mobile apps; smartphone; mobile phone; mobile health; mHealth; digital intervention; support; app; feasibility; acceptability; engagement; usability

Introduction

Most of the 14,000 individuals who leave the UK Armed Forces (UKAF) every year [1] transition to civilian life successfully; however, some experience difficulties with their mental health [2]. Approximately 6% of UK veterans are likely to have posttraumatic stress disorder (PTSD), 22% report symptoms of common mental disorders such as anxiety or depression, and 10% report alcohol misuse [2]. More than half of the veterans experiencing mental health difficulties do not seek professional help [3], with studies consistently demonstrating that veterans often prefer to manage their mental health problems alone [4]. Veterans with mental health difficulties are prone to negative life outcomes, such as lower employment rates [5,6], emphasizing the need to encourage mental health support in this population.

Previous research investigating barriers to help-seeking in male veterans [7] identified 3 core barriers: veterans struggled to define the problems they were experiencing as mental health difficulties; veterans did not recognize their need for help until their difficulty reached a crisis point; and once help was sought, veterans found it challenging to navigate the care pathway and make sense of the number of services available to them. Given the poor uptake of formal mental health support in this population and the drive to manage mental health independently, an app that enables veterans to actively manage their symptoms and encourages them to access formal care if self-help has been ineffective could be of substantial benefit.

The Mental Health Toolkit for Veterans Project (MeT4VeT) was initiated to develop a smartphone app to help transitioning UK veterans by identifying mental health difficulties, providing a degree of self-help, and assisting them in recognizing when more formal support is required [8]. The app was developed by King's Centre for Military Health Research, King's College London according to the UK Medical Research Council (MRC) complex intervention framework [9,10]. This included a plan for the app development process, the involvement of veterans and stakeholders from services providing veteran support, and a scoping review of published research evidence for existing toolkits (eg, apps and websites). A smartphone app was the preferred format because of its geographical reach, 24/7 availability, lack of requirement for face-to-face contact, and ability to be downloaded at no cost [11]. The app was designed according to the MRC complex intervention framework by drawing on existing theories of cognitive behavioral therapy (CBT), which have been used as the foundation for many self-help tools and have demonstrated effectiveness in treating common mental disorders [12].

The MeT4VeT app consisted of five main sections: (1) *People*, enabling participants to identify symptoms of mental health difficulties through veterans' stories; (2) *Tasks*, encouraging participants to set tasks to help them achieve goals in different areas of their life, for example, work, family and physical health; (3) *Tools*, providing participants with resources to independently manage their mental health; (4) *Tracking*, allowing participants to monitor their mental health and see their progression through the app; (5) *Notifications*, daily notifications to encourage engagement with the app. The elements of the app represent the intersection between core components of CBT, behavior change constructs, and potential means to overcome each of the barriers outlined in earlier research [7]. Core principles of CBT include the self-monitoring of thoughts, feelings, and behaviors (incorporated within the People section), behavioral activation and goal setting (included within the Tasks section), the use of techniques such as mindfulness and cognitive reframing (contained within the Tools section), and feedback on behavior (included in the Tracking section) [12].

This paper outlines a feasibility trial conducted in line with the MRC complex intervention framework [9,10]. The framework outlines the 4 phases in the development of an intervention and the trial focused on the second phase: "feasibility: assessing feasibility and acceptability of intervention and evaluation design to make decisions about progression to the next stage of evaluation" [10].

In line with this framework, a set of objectives were created to cover the following three core areas:

1. Assess the ability to recruit and retain participants throughout the trial period (28 days);
2. Evaluate technical app delivery and measurement processes, including the acceptability of 4 outcome measures: General Health Questionnaire (GHQ-12), PTSD Checklist-Civilian Version (PCL-C), Warwick Edinburgh Mental Well-Being Scale (WEMWBS), and World Health Organization Quality of Life Assessment (WHOQOL-BREF);
3. Examine the acceptability of the intervention, including engagement with and usability of the intervention.

Methods

Design

The trial had a parallel design. Participants were randomly assigned on a 1:1 ratio using block randomization to either an

intervention group, who received the full smartphone app, or a control group, who received a noninteractive app that contained only signposting information. Data were collected at baseline and follow-up periods of 1 month (28 days) and 3 months (84 days) postbaseline. These time periods were chosen to align with other self-help intervention trials targeting a change in GHQ-12 scores and upon which our initial power calculation was based [13]. The time points enabled an assessment of any change at the end of the trial period (28 days) and again at 3 months to ascertain whether the change was continued. All participants were asked to use the app for a minimum of 28 days and app notifications were sent to the intervention group to encourage engagement.

Participants

Participants were eligible for the study if they (1) are male, (2) own a smartphone, (3) served at least 2 years in the UKAF, (4) left the UKAF within the last 2 years, (5) are not currently undertaking formal mental health treatment (eg, mental health therapy conducted by a medical professional), and (6) indicated a degree of mental health distress assessed via a score of 2 or more on the GHQ-12 [14], similar to previous research [15]. Participants received compensation of a GBP £10 (US $12.85) Amazon voucher for completing each questionnaire (GBP £30 total, US $38.56), and participants in the intervention group who completed an interview received an additional GBP £10 (US $12.85) Amazon voucher.

Recruitment

Participant recruitment was attempted at 3 UK military medical centers between October 2020 and February 2021. Personnel undergo a medical check-up when they leave the military as part of the discharge process, which includes a mental health screening. Three participating medical centers were asked to identify eligible participants at these check-ups and invite them to participate in the study. However, this approach became unfeasible due to the COVID-19 pandemic. The recruitment strategy was then expanded to include the use of social media and advertising in veteran third-sector organizations between February and November 2021. The social media platforms Facebook (Meta Platforms Inc), Instagram (Meta Platforms Inc), and Twitter (Twitter Inc) were used to post both paid and free promotional advertisements, which included a link to the participant information sheet and consent form on the questionnaire platform Qualtrics (Qualtrics International Inc).

App Design

Full details regarding the app development and its components have been reported previously [8]. The intervention group received the MeT4VeT app with full functionality (Figure 1 and Table 1). This included a daily schedule of notifications for the first 7 days related to the main app sections, followed by targeted notifications depending on participants' usage of the app. For example, participants who completed at least 2 tasks in a week would receive the notification: "You're making great progress with the 'Tasks' section. Are there any other behaviors you'd like to work on?" compared with participants who completed less than 2 tasks in a week: "It's been a while since you have set some tasks, maybe add some more or change the ones you have set?" The control group received an app containing only signposting information about a range of statutory and third-sector organizations that focused on veterans' mental health. Both apps were available on Android (Google) through the Google Play Store (Google) and on iOS (Apple Inc) via the TestFlight (Apple Inc) app.

Figure 1. The Mental Health Toolkit for Veterans Project (MeT4VeT) app used by the intervention group.

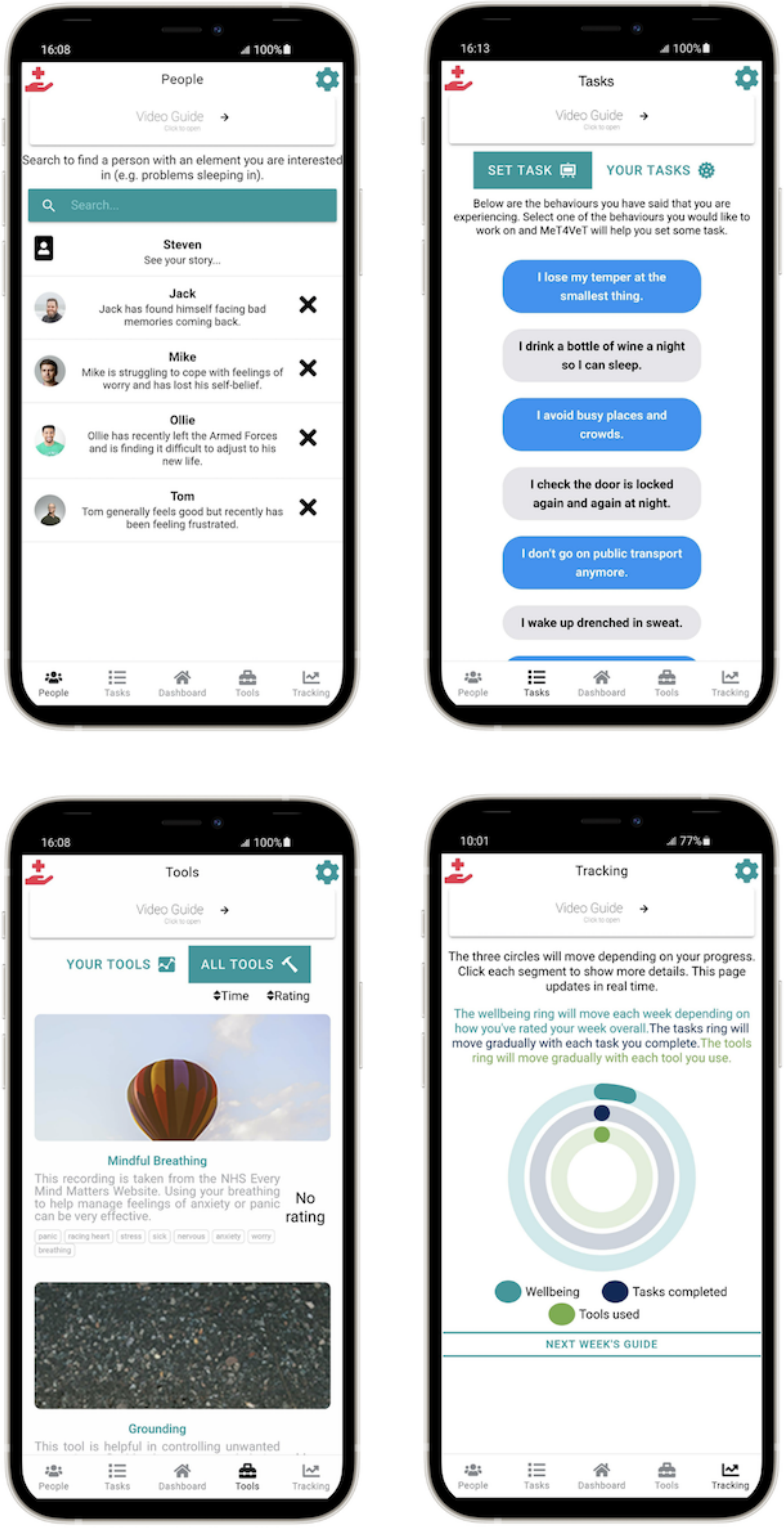


Table 1. App sections and content for the intervention and control groups.

Accessed by and app section	Content	Intervention	Control
First app launch and automatically displayed at follow-up points			
Questionnaires	Participants completed all questionnaires within the app after opening it for the first time and at subsequent follow-ups.	✓	✓
Navigation menu			
Dashboard	After the questionnaires were completed, the dashboard was the main landing page. It contained information on how long the participant had been using the app, their name (set by the participant), and general advice for using the app.	✓	
People	This section included 4 videos depicting fictional veterans discussing their mental health difficulties, corresponding to typical presentations of generalized anxiety disorder, depression, posttraumatic stress disorder, and a veteran who was struggling with mental health but did not necessarily meet a diagnostic threshold. After watching the videos, participants were asked to select the thoughts, feelings, and behaviors they related to in order to help them develop a picture of their mental health.	✓	
Tasks	Participants were encouraged to identify areas of their lives that were important to them, such as work, family, or physical health, and to set a series of small tasks to help make changes in these areas.	✓	
Tools	A range of mental health resources (eg, guided safe place imagery) were obtained from the National Health Service "Every Mind Matters" website [16]. These tools could either be accessed as a list or filtered to map onto the most appropriate ones to deal with the difficulties the veteran had selected in the earlier People section.	✓	
Tracking	Participants were able to monitor their mental health through weekly check-ins, which were viewable in the Tracking section along with their progress across the app.	✓	
Icons at the top of the app			
Signposting	A list of statutory and third-sector organizations that focused on veterans' mental health.	✓	✓
Settings	Information about the app, privacy policy, licenses, and questionnaire schedule. Participants were able to change their name, turn on or off notifications, and send feedback.	✓	✓
Passive			
Notifications	Participants were encouraged to engage with the app through a schedule of daily notifications.	✓	

Randomization

Once recruited into the study, participants were assigned a unique proxy identifier. This was linked to a randomization sequence generated in Stata (StataCorp) statistical software, which assigned participants on a 1:1 ratio to either the intervention or control group. All members of the research team were blind to group allocations except for authors SP, BC, and DL. SP and BC managed participant recruitment, conducted primary analyses, and had access to the raw study data. DL developed the MeT4VeT app and had access to the raw study data. Instructions for downloading the relevant app were sent to the participants via email. The apps could only be activated using a QR code or token code, which was sent to the participants by the research team upon successful registration to the trial.

Trial Objectives

Overview

A set of objectives were developed a priori to assess areas of uncertainty before progression to a randomized controlled trial (RCT). These were based on the CONSORT (Consolidated Standards of Reporting Trials) guidelines for conducting a feasibility trial [17] (Table 2).

Table 2. Objectives to assess progression to a randomized controlled trial.

Criteria	Objectives
Recruitment and retention	1. Assess the number of eligible or interested participants.
	2. Assess recruitment processes by calculating response rate.
	3. Explore retention by estimating 1- and 3-month follow-up rates.
	4. Determine the comparability of the sample to Ministry of Defence statistics for major subgroups of veterans (age, service, and rank).
Trial procedures and outcome measures	5. Evaluate the technical delivery of the app by gathering feedback from developers and participants.
	6. Examine the feasibility of using the app to collect outcome measures by assessing completion rates and times.
	7. Sample size estimation for the randomized controlled trial by assessing changes in mental health measures between groups.
Acceptability and usability	8. Explore the participants' interaction with core areas of the app (People, Tasks, Tools, and Tracking) via app usage data.
	9. Examine the participants' interaction with the app throughout the trial period via app usage data.
	10. Assess the usability of the app using the MAUQ[a].

[a]MAUQ: mHealth App Usability Questionnaire.

Recruitment and Retention

Recruitment rates throughout the trial were presented according to the CONSORT flow diagram [17], with a breakdown of reasons provided for participant dropout. Demographic information was captured from all participants to assess their similarity to those who had left the UKAF in the last 2 years, drawn from UK Ministry of Defence (MOD) statistics [18]. Information was captured regarding age, military service branch, military rank, and regular or reservist status (Table 3).

Table 3. Summary of measures across all time points.

Measure	Baseline	Baseline + 1 month	Baseline + 3 months
Demographics	✓		
Age			
Ethnicity			
Relationship status			
Military service branch			
Military rank			
Regular or reservist status			
Reason for leaving the military			
App acceptability		✓	
Engagement: app usage data			
Usability: MAUQ[a]			
Outcome measures	✓	✓	✓
GHQ-12[b]			
PCL-C[c]			
WEMWBS[d]			
WHOQOL-BREF[e]			

[a]MAUQ: mHealth App Usability Questionnaire.
[b]GHQ-12: General Health Questionnaire.
[c]PCL-C: PTSD Checklist-Civilian Version.
[d]WEMWBS: Warwick Edinburgh Mental Well-Being Scale.
[e]WHOQOL-BREF: World Health Organization Quality of Life Assessment.

Trial Procedures and Outcome Measures

Trial procedures were assessed by examining the technical delivery of the app and a review of the feedback option within the Settings section where participants were asked to note any technical issues as they occurred. The number of completed in-app surveys and the amount of time required to complete the surveys were also assessed.

Four outcome measures were trialed and completed by the participants in the Questionnaires section of the app. The GHQ-12 is a measure of common mental disorders such as anxiety and depression [14]. It is a 12-item self-report questionnaire measured on a 4-point scale. An example item is "Have you recently been feeling unhappy and depressed?" The GHQ-12 has demonstrated good psychometric properties (eg, Cronbach α=.91) [19].

To evaluate symptoms of PTSD, the PCL-C was used [20]. It is a 17-item self-report questionnaire, measured on a 5-point scale from (1) not at all to (5) extremely. An example item is "Over the past month how much have you been bothered by repeated, disturbing memories, thoughts or images of a stressful experience?" The PCL-C has been shown to have good psychometric properties (eg, Cronbach α=.96) [21].

The WEMWBS measures mental well-being within the previous 2 weeks and is a 14-item self-report questionnaire [22]. It is measured on a 5-point scale from (1) none of the time to (5) all of the time and an example item is "I've been thinking clearly." The WEMWBS has been shown to have good internal consistency (Cronbach α=.89) and test-retest reliability (Cronbach α=.83) [22].

Quality of life was assessed using the WHOQOL-BREF, a 26-item self-report questionnaire [23]. It is measured on a 5-point scale and an example item is "How satisfied are you with yourself?" The WHOQOL-BREF has been shown to have good psychometric properties (eg, Cronbach α>.80 on 3 domains) [24].

Acceptability and Usability

App acceptability was assessed through engagement with the app, obtained from the app usage data. Engagement is defined as usage of the intervention including the temporal patterns (frequency and duration) and depth (specific intervention content) [25]. Engagement has been linked to behavior change through specific mechanisms of action such as the content and delivery of the intervention [25]. App usage data were collected from Google Analytics via Firebase and included information such as the number of times the app was opened, the duration of each app use, and the number of times specific sections of the app were used. Participants' devices automatically reported the data to Google Analytics, although participants could choose to disable app tracking from their own devices outside of the MeT4VeT app. This type of automatic tracking of usage is a commonly used measure of engagement for digital behavior change interventions [25].

Usability data were obtained from the mHealth App Usability Questionnaire (MAUQ) [26], completed within the Questionnaires section of the app. The MAUQ is a 21-item self-report questionnaire measured on a 7-point scale ranging from strongly disagree (1) to strongly agree (7). Similar to a previous veteran app study [27], 5 questions were removed (resulting in a 16-item questionnaire), as the questions related to communicating with health care services through the app, which was not applicable in our study. Participant responses were aggregated into overall usability and 3 other domains: ease of use and satisfaction, interface and functionality, and usefulness [26]. The mean total score and the mean of each 3 domains were scored out of 7, with a score of 4 indicating a neutral response and higher scores indicating greater app usability. The MAUQ has been shown to have good psychometric properties (eg, Cronbach α>.80) for all 3 domains [26].

Data Analysis

Data were analyzed using Stata v17 (StataCorp). For the outcome measures trialed, independent sample *t* tests and chi-square tests (for categorical data) were used to assess changes in outcome measures between the groups at 3 time points (baseline, baseline + 1 month, and baseline + 3 months) (Table S1 in Multimedia Appendix 1). App usage data were reported using medians and IQR. The MAUQ was reported using means and SD. A power analysis calculation was performed using the GHQ-12 with a threshold score of 3, as the common thresholds used in previous research are 2 or 3 [15]. Statistical significance is reported using *P* values with statistical significance being determined as $P<.05$.

Ethics Approval

The feasibility trial was granted full ethical approval by the UK Ministry of Defence Research Ethics Committee (1074/MODREC/20) and the Research Governance Office at King's College London (DPRF-19/20-16015).

Results

Overview

Three main areas of the feasibility trial were assessed (recruitment and retention; trial procedures and outcome measures; and acceptability and usability) to provide data on areas of uncertainty prior to progression to an RCT (Table 4).

Table 4. Objectives to assess progression to a randomized controlled trial and results from the Mental Health Toolkit for Veterans Project (MeT4VeT) feasibility trial.

Criteria and objectives	Results
Recruitment and retention	
1. Assess the number of eligible or interested participants.	791 interested, 530 eligible (67%). Suggest opening criteria to women and those who have left the UK Armed Forces more than 2 years ago
2. Assess the recruitment processes by calculating the response rate.	Response rate of 29% (153/530) suggests refinements to minimize attrition
3. Explore retention by estimating 1- and 3-month follow-up rates.	Retention rates of 80% (40/50) at 1 month and 68% (34/50) at 3 months from baseline
4. Determine the comparability of the sample to UK MOD[a] statistics for major subgroups of veterans (age, service, and rank).	Comparison conducted, suggest refinements to target younger veterans and women to enhance comparability
Trial procedures and outcome measures	
5. Evaluate the technical delivery of the app by gathering feedback from developers and participants.	Minor feedback resolved
6. Examine the feasibility of using the app to collect outcome measures by assessing completion rates and times.	Completion rates and times calculated and deemed acceptable by the research team
7. Sample size estimation for the randomized controlled trial by assessing changes in mental health measures between groups.	Power calculation conducted. Potential effectiveness of measures identified
Acceptability and usability	
8. Explore the participants' interaction with core areas of the app (People, Tasks, Tools, and Tracking) via app usage data.	App usage data collected; suggest refinements to the Tools section of the app
9. Examine the participants' interaction with the app throughout the trial period via app usage data.	App usage data collected; suggest refinements to maintain engagement
10. Assess the usability of the app using the MAUQ[b].	Assessment conducted and deemed acceptable by the research team

[a]MOD: Ministry of Defence.

[b]MAUQ: mHealth App Usability Questionnaire.

Recruitment and Retention

Following recruitment, 791 people completed the participant information sheet and 261 did not meet the eligibility criteria (Figure 2). There were multiple reasons why interested participants were not eligible to participate in the study (Table 5). The most common reasons were undergoing mental health treatment and not having left the UKAF in the last 2 years. A total of 530 participants completed the consent form and 377 of those either declined or could not be contacted for recruitment into the study. Overall, 153 participants were recruited into the study, but 99 did not download the app and 4 did not complete the baseline measures. The final sample consisted of 50 participants randomly assigned to the intervention group (n=24) and control group (n=26). After 1 month, 10 participants were lost to follow-up (n=6 intervention group, n=4 control group) and after 3 months, another 6 were lost to follow-up (n=3 in both groups).

Figure 2. Participant recruitment flow diagram.

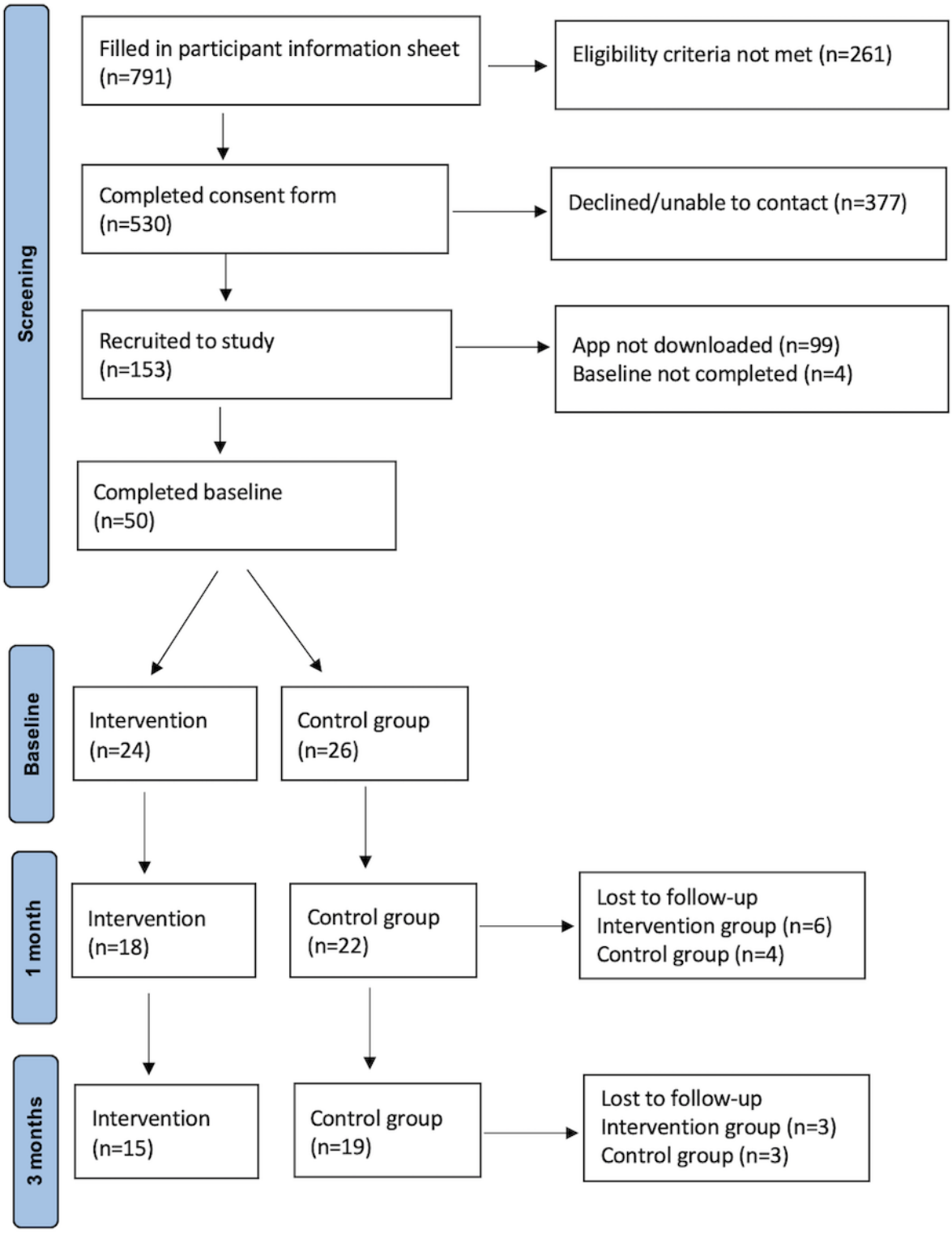


Table 5. Reasons why participants were not eligible to participate in the trial.

Criteria[a]	Participants (n=261), n
Currently undertaking mental health treatment	110
Did not leave the UK Armed Forces in the last 2 years	86
Score of <2 on the GHQ-12[b]	74
Incomplete response on the participant information sheet	16
Did not serve in the UK Armed Forces for at least 2 years	9
Not male	5

[a]Some participants were ineligible because of multiple criteria.

[b]GHQ-12: General Health Questionnaire.

At baseline, there were no significant differences between the groups for any participant characteristic (Table 6). The mean age of the participants was 41.7 years (95% CI 38.8-44.7). Participants were predominantly White (47/50, 94%) and married, cohabiting, or in a long-term relationship (41/50, 82%). Two-thirds of participants had served in the Army (32/50, 64%), followed by the Royal Navy or Royal Air Force (18/50, 36%). Participants were predominately noncommissioned officers (40/50, 80%) and regular serving personnel (45/50, 90%). Of the reported reasons for leaving the military, there were several responses in the "other" category (25/50, 50%) such as health problems and dissatisfaction with their military career. Over half of all participants used an Android device (27/50, 54%).

Table 6. Participant characteristics, military background, and device information.

Characteristics	Total (N=50)[a], n (%)	Intervention (n=24)[a], n (%)	Control (n=26)[a], n (%)
Age (years)			
Younger than 40	18 (36)	10 (42)	8 (31)
40 to 49	19 (38)	8 (33)	11 (42)
50 and older	13 (26)	6 (25)	7 (27)
Relationship status			
In a relationship	41 (82)	20 (83)	21 (81)
Not in a relationship	9 (18)	4 (17)	5 (19)
Service branch			
Army	32 (64)	14 (58)	18 (69)
Royal Navy or Royal Air Force	18 (36)	10 (42)	8 (31)
Military rank			
Officer	10 (20)	4 (17)	6 (23)
Nonofficer	40 (80)	20 (83)	20 (77)
Reason for leaving the Armed Forces			
Completed term of service	17 (34)	9 (37)	8 (31)
Impact of service life on family	8 (16)	4 (17)	4 (15)
Other	25 (50)	11 (46)	14 (54)
Device operating system			
Android	27 (54)	14 (58)	13 (50)
iOS	23 (46)	10 (42)	13 (50)

[a]When cells had less than 4 participants, categories were combined.

Participant characteristics were compared to the UK MOD personnel statistics [18] for male UKAF regulars leaving service from March 2020 to March 2022 (n=62,740). Statistics including reservists were not available by gender. The findings indicated an approximate mapping for rank: officer, 20% (11% MOD); other ranks, 80% (89% MOD); and service branch (Army, 64% [61% MOD], Royal Navy or Royal Air Force, 36% [39% MOD], Royal Navy or Royal Marines and Royal Air Force). Comparison of the age of participants highlights a far younger age demographic compared to the population included in this study: 36% (75% MOD) were younger than 40 years, 38% (17%

MOD) were 40 to 49 years, and 26% (8% MOD) were 50 years and older.

A power calculation was performed to determine the sample size needed for a future RCT based on the difference in GHQ-12 scores at baseline and follow-up (3 months). To detect a difference in the GHQ-12 means of 3 points between groups at the 3-month follow-up, with 80% power and a significance level of 5%, each group would need to have 39 participants (n=78 in total). To allow for an attrition rate of 32% at the 3-month follow-up (16/50) based on this feasibility trial, 115 participants would need to be recruited. To account for a response rate of 29% (153/530) based on this feasibility trial, recruiting 397 participants is recommended for a future RCT.

Trial Procedures and Outcome Measures

Technical delivery of the app was successful, with only minor resolvable technical issues identified by either the research team or the participants. Three participants (total N=50) provided technical feedback within the app; 1 needed support to clear notifications and 2 participants from the control group questioned whether there was additional content. The protocol was successful in collecting all measures outlined and the participants completed 124 surveys within the app, totaling 19 hours (median 21 minutes per participant).

At baseline, there were no significant differences between the groups on any outcome measures (Table S1 in Multimedia Appendix 1). For the intervention group, symptoms of mental health distress (GHQ-12) decreased significantly (P<.001) from preintervention (mean 7.0, SD 3.6) to the 3-month follow-up (mean 2.7, SD 3.6) with a mean difference of 4.3 (95% CI 1.9-6.7) (Table S4 in Multimedia Appendix 1). Symptoms of PTSD (PCL-C) decreased significantly (P=.03) from preintervention (mean 44.2, SD 15.0) to the 3-month follow-up (mean 33.2, SD 13.9) with a mean difference of 11.0 (95% CI 1.3-20.7). Well-being (WEMWBS) improved significantly (P=.04) from preintervention (mean 38.3, SD 9.2) to the 3-month follow-up (mean 44.9, SD 9.4) with a mean difference of −6.6 (95% CI −12.8 to −0.5). For the WHOQOL-BREF subdomain of psychological, quality of life increased significantly (P=.01) from preintervention (mean 11.1, SD 2.6) to the 3-month follow-up (mean 13.3, SD 2.4) with a mean difference of −2.2 (95% CI −3.9 to −0.5).

For the control group, there were no significant differences in any of the outcome measures from preintervention to the 3-month follow-up (Table S5 in Multimedia Appendix 1). Analyses of the outcome measures across all time points are available in Multimedia Appendix 1.

Acceptability

Over the trial period (28 days), the intervention group initialized the app a median of 8.5 times (IQR 3-18) and over a median period of 2.5 weeks (IQR 1-3.5) (Table 7). The median session duration was 20.7 seconds (IQR 17.2-34.3). The control group initialized the noninteractive app a median of 2 times (IQR 1-6) and over a median period of 1 week (IQR 1-1). The median session duration was 29 seconds (IQR 20-39.2).

Table 7. App usage per participant for both groups over the trial period (28 days).

App usage	Intervention (n=24), median (IQR)	Control (n=26), median (IQR)
Initializations[a]	8.5 (3-18)	2 (1-6)
Session count[b]	52 (12-112)	15.5 (7-27)
Session duration (seconds)	20.7 (17.2-34.3)	29 (20-39.2)
Server interactions[c]	9.5 (4-40.5)	5 (4-8)
Weeks active	2.5 (1-3.5)	1 (1-1)

[a]Number of times the app was opened as a new instance.

[b]Number of times the app was reopened after running in the background.

[c]Number of times the app interacted with the server.

The People section of the app was accessed most frequently by the intervention group with an average of 14 (IQR 4-45) visits per participant, followed by the Dashboard section with an average of 11 (IQR 6-21) visits per participant (Table 8). Other than the Questionnaires section, participants spent the most time on the People section with an average visit length of 2.8 minutes (median 169.3 seconds, IQR 54.2-430.4). Out of the 4 main sections of the app (People, Tasks, Tools, and Tracking), the Tracking section was accessed by the least number of participants (n=19, 79%), the Tasks section was used the least number of times (median 8.5, IQR 4-43) and the Tools section was used for the least amount of time (median 49.5, IQR 6.7-166.8).

Table 8. App usage for each section of the Mental Health Toolkit for Veterans Project (MeT4VeT) app in the intervention group over the 28-day trial period (n=24).

App screen	Ever accessed[a], n (%)	Number of times accessed per person, median (IQR)	Time on screen per person (seconds), median (IQR)
Questionnaires[b]	23 (96)	2 (1-3)	552.5 (280.7-725.1)
Dashboard	23 (96)	11 (6-21)	103.3 (62.2-139.2)
People	22 (92)	14 (4-45)	169.3 (54.2-430.4)
Tasks	22 (92)	8.5 (4-43)	107.6 (25.5-452.8)
Tools	20 (83)	9 (2.5-23)	49.5 (6.7-166.8)
Tracking	19 (79)	9 (4-21)	68.9 (15.1-150.8)
Settings	16 (67)	3 (1.5-4)	14.9 (0-22.8)
Signposting	15 (63)	2 (2-6)	9.9 (0-40.8)

[a]During the trial, Apple (developer of iOS) changed policies related to how developers could track and monitor the usage of an app. This required specific user content, which could be modified outside the app. It is therefore not possible to ascertain whether a user did not provide data because they were not using the app or whether they declined to share app usage statistics.

[b]One participant accessed the app but did not open the Questionnaires screen until day 29.

Usability

Participants in the intervention group rated the overall usability (mean 5.3, SD 1.5) on the MAUQ and each of the subdomains of ease of use and satisfaction (mean 5.5, SD 1.5), interface and functionality (mean 4.9, SD 1.6), and usefulness (mean 5.3, SD 1.8) of the app with positive (above neutral) responses. The control group had higher scores compared to the intervention group for overall usability (mean 5.6, SD 1.2), ease of use and satisfaction (mean 5.9, SD 1.1), interface and functionality (mean 5.5, SD 1.4), and lower scores for usefulness (mean 4.8, SD 1.6).

Discussion

Principal Findings

The aims of the trial were to collect data to determine the feasibility of running an RCT by assessing (1) recruitment and retention, (2) trial procedures and outcome measures, and (3) the acceptability of the MeT4VeT app.

Recruitment and Retention

The COVID-19 pandemic prevented effective recruitment through military bases and social media was used as an alternative recruitment strategy. This presented some challenges in obtaining valid responses and retaining participants, which likely accounts for the high dropout rate. These challenges are common in research using social media as a recruitment tool [28,29]. If a future RCT uses social media recruitment, further work will be required to reduce dropout rates or increase the provision of resources to support this degree of dropout. Recruitment may have also been impacted by stringent restrictions from Apple for publishing apps on their app store and another app (TestFlight) was required to be installed before MeT4VeT could be downloaded, which may have led to reduced engagement. Once participants were engaged in the trial, the retention rate was 68% (34/50) from baseline measures to follow-up, which is acceptable compared to other trials assessing interventions within a military veteran population [30,31]. A trial exploring a mental health monitoring app within this population indicated a retention rate of 81% from baseline to follow-up [30] and a trial evaluating an alcohol use reduction intervention reported a retention rate of 22% to follow-up [31].

The sample in this study is comparable to UK MOD statistics [18] in terms of service branch and rank division across the UKAF. However, our sample represented a substantially older demographic when compared to those leaving the UKAF between 2020 and 2022 [18]. To improve comparability, we suggest recruiting from a wider age range to include younger veterans and females.

Examination of the eligibility criteria showed that a large number of potential participants were not eligible as they had not left the UKAF within the last 2 years. Future studies could relax this criterion to enable veterans to enter the trial at any stage of transition. An even greater number of potential participants were not eligible to participate because they were currently undertaking mental health treatment, which could indicate a need for the development of an app to enable self-management to complement therapy within this population. The presence of a group of potential participants who did not meet the GHQ-12 cut-off but were interested in an app to manage their mental health suggests a potentially broader scope of interest to maintain positive mental health or prevent mental health distress.

Trial Procedures and Outcome Measures

The results of the trial indicate it would be feasible to conduct an RCT using the same technical delivery of the app and the same set of measurement processes. Four outcome measures were deemed acceptable for data collection, with the intervention indicating a potential improvement in mental health from preintervention to the 3-month follow-up, contrasting with no significant improvement in the GHQ-12 scores for the control group over the same time period. Due to the small sample size, it is difficult to draw any firm conclusions from these results, but they provide a promising signal for the MeT4VeT app and the use of the GHQ-12 as an outcome measure to be explored

in a future RCT with longer follow-up periods and a larger sample size. A provisional sample size for a later RCT of 397 was calculated.

Acceptability and Usability

The MAUQ scores indicated positive responses to the app, with scores ranging from 4.9 to 5.5, all of which were above a neutral response of 4. However, the scores were below those of a similar app designed to reduce veterans' alcohol consumption where scores of 5.7 to 5.9 were reported across the MAUQ domains [31]. Before a future RCT, work should be conducted using qualitative feedback on the app to refine its usability. Qualitative data were collected from this trial and will be reported in a future publication.

The app usage data indicated that the intervention group engaged with the MeT4VeT app more often (initializations and sessions) and for a longer duration than the control group (weeks active). When focusing on the 4 core sections of the app (People, Tasks, Tools, and Tracking), the People section was accessed most frequently and for the longest duration and it required active participation. The remaining sections were accessed less frequently by participants. This may have been because they could be accessed via other resources (eg, Tools), whereas the People section provided a novel resource (videos with veterans) not readily available elsewhere. A future iteration of the app may be best focused as a gateway into existing self-help tools, with continued development of the People, Tasks, and Tracking sections and replacement of the Tools section with links to existing resources. This aligns with the original intention of the app, to create a way to help people at the beginning of their journey to identify symptoms and track them over time to recognize when difficulties occur and when self-help, or more formal support, may be beneficial. The median duration of app usage in the intervention group was 2.5 weeks, which falls below the 4-week trial period. Before an RCT is initiated, further research is required to explore the reasons behind early disengagement with the app and potentially refine the notification schedule to encourage app usage throughout the full trial period.

Strengths

The strengths of this study include its rigorous design, with random assignment to groups and the inclusion of a control group. The MeT4VeT app development used a strong conceptual framework informed by the military mental health literature and previous research [7]. Furthermore, feedback on the app was provided throughout the app development process by stakeholders working on veterans' mental health at 4 stakeholder events, through interviews with veterans, and veteran involvement and engagement groups.

Limitations

The limitations of the study include the relatively small sample size, partly due to the COVID-19 pandemic, which impacted recruitment, and partly due to the loss of potential participants lacking interest in trial initiation. The generalizability of the results is affected by the limitations of the sample, which included only male veterans and those that had left the UKAF within the last 2 years. The development of the app was based on data drawn from research on male veterans [7], so the same barriers could not be assumed to be synonymous for the female veteran population. Future research should explore whether female veterans experience the same barriers and evaluate the use of the app in this population. The feasibility study tentatively demonstrated that MeT4VeT may be useful for those within 2 years of leaving the UKAF and whose departure from the UKAF was over a longer time period; further research is warranted to explore this.

Conclusions

The results of the trial suggest that although the study protocol was both feasible and acceptable, the recruitment process requires further refinement to provide an appropriate number of participants for a future RCT. The outcome measures used within the trial were acceptable for use in a future RCT to robustly examine the effectiveness of the app. The findings indicate that the MeT4VeT app shows some promise as a useful, accessible way for veterans to monitor and manage their mental health.

Acknowledgments

This study was funded by the Forces in Mind Trust (FiMT) using an endowment awarded by the National Lottery Community Fund (project FiMT17/1023K). The views expressed are those of the authors and not necessarily those of the FiMT, National Health Service, National Institute for Health and Care Research, Ministry of Defence, Society of Occupational Medicine, or Faculty of Occupational Medicine. For the purposes of open access, the authors have applied a Creative Commons attribution (CC-BY) license to any accepted author manuscript version arising from this submission.

Conflicts of Interest

SAMS is supported by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley National Health Service Foundation Trust and the NIHR Advanced Fellowship (NIHR300592). DL is a reservist in the UK Armed Forces; this work has been undertaken as part of his civilian employment. NTF is a member of the Emergency Services Senior Leaders Board and is partly funded by a grant from the UK Ministry of Defence. NTF is also a trustee of Help for Heroes, a charity supporting the well-being of service personnel, veterans, and their families. NG is a trustee of the Society of Occupational Medicine and the Faculty of Occupational Medicine and is the military lead for the Royal College of Psychiatrists.

Editorial notice: This randomized study was only retrospectively registered. The authors explained this is because it is a feasibility trial. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Outcome measures for all time points.
[DOCX File , 44 KB - mental_v10i1e46508_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist V 1.6.1.
[PDF File (Adobe PDF File), 1179 KB - mental_v10i1e46508_app2.pdf]

References

1. UK Armed Forces quarterly service personnel statistics. Ministry of Defence. 2022 Jan 01. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1059988/SPS_1_January_2022._10322.pdf [accessed 2022-05-19]
2. Stevelink SAM, Jones M, Hull L, Pernet D, MacCrimmon S, Goodwin L, et al. Mental health outcomes at the end of the British involvement in the Iraq and Afghanistan conflicts: a cohort study. Br J Psychiatry 2018;213(6):690-697 [FREE Full text] [doi: 10.1192/bjp.2018.175] [Medline: 30295216]
3. Stevelink SAM, Jones N, Jones M, Dyball D, Khera CK, Pernet D, et al. Do serving and ex-serving personnel of the UK armed forces seek help for perceived stress, emotional or mental health problems? Eur J Psychotraumatol 2019;10(1):1556552 [FREE Full text] [doi: 10.1080/20008198.2018.1556552] [Medline: 30693074]
4. Shepardson RL, Tapio J, Funderburk JS. Self-management strategies for stress and anxiety used by nontreatment seeking veteran primary care patients. Mil Med 2017;182(7):e1747-e1754 [FREE Full text] [doi: 10.7205/MILMED-D-16-00378] [Medline: 28810968]
5. Iversen A, Nikolaou V, Greenberg N, Unwin C, Hull L, Hotopf M, et al. What happens to British veterans when they leave the armed forces? Eur J Public Health 2005;15(2):175-184 [FREE Full text] [doi: 10.1093/eurpub/cki128] [Medline: 15755777]
6. Murphy D, Palmer E, Busuttil W. Exploring indices of multiple deprivation within a sample of veterans seeking help for mental health difficulties residing in England. J Epidemiol Public Health Rev 2017;1(6):1-6 [FREE Full text] [doi: 10.16966/2471-8211.132]
7. Rafferty LA, Wessely S, Stevelink SAM, Greenberg N. The journey to professional mental health support: a qualitative exploration of the barriers and facilitators impacting military veterans' engagement with mental health treatment. Eur J Psychotraumatol 2020;10(1):1700613 [FREE Full text] [doi: 10.1080/20008198.2019.1700613] [Medline: 33488992]
8. Parkes S, Croak B, Brooks SK, Stevelink SAM, Fear NT, Rafferty L, et al. MeT4VeT: development and assessment of a mental health toolkit for military veterans. Forces in Mind Trust. 2022. URL: https://s31949.pcdn.co/wp-content/uploads/20220708-MeT4VeT-FiMT-Report-Final.pdf [accessed 2022-07-19]
9. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]
10. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
11. Owen JE, Kuhn E, Jaworski BK, McGee-Vincent P, Juhasz K, Hoffman JE, et al. VA mobile apps for PTSD and related problems: public health resources for veterans and those who care for them. mHealth 2018;4(7):28 [FREE Full text] [doi: 10.21037/mhealth.2018.05.07] [Medline: 30148141]
12. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
13. Bower P, Richards D, Lovell K. The clinical and cost-effectiveness of self-help treatments for anxiety and depressive disorders in primary care: a systematic review. Br J Gen Pract 2001;51(471):838-845 [FREE Full text] [Medline: 11677710]
14. Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. Psychol Med 1979;9(1):139-145. [doi: 10.1017/s0033291700021644] [Medline: 424481]
15. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997;27(1):191-197. [doi: 10.1017/s0033291796004242] [Medline: 9122299]

16. Every Mind Matters. National Health Service. URL: https://www.nhs.uk/every-mind-matters/ [accessed 2022-05-19]
17. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
18. UK Armed Forces biannual diversity statistics: April 2022. Ministry of Defence. 2022 Jun 30. URL: https://www.gov.uk/government/statistics/uk-armed-forces-biannual-diversity-statistics-april-2022 [accessed 2023-01-13]
19. McCabe CJ, Thomas KJ, Brazier JE, Coleman P. Measuring the mental health status of a population: a comparison of the GHQ-12 and the SF-36 (MHI-5). Br J Psychiatry 1996;169(4):517-521. [doi: 10.1192/bjp.169.4.516] [Medline: 8894205]
20. Weathers F, Litz B, Herman D, Huska J, Keane T. The PTSD Checklist (PCL): reliability, validity, and diagnostic utility. 1993 Presented at: Annual Meeting of the International Society for Traumatic Stress Studies; Oct; San Antonio, Texas. [doi: 10.1037/t00503-000]
21. Keen SM, Kutter CJ, Niles BL, Krinsley KE. Psychometric properties of PTSD checklist in sample of male veterans. J Rehabil Res Dev 2008;45(3):465-474 [FREE Full text] [doi: 10.1682/jrrd.2007.09.0138] [Medline: 18629754]
22. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
23. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
24. Skevington SM, Lotfy M, O'Connell KA, WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004;13(2):299-310. [doi: 10.1023/B:QURE.0000018486.91360.00] [Medline: 15085902]
25. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
26. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
27. Williamson C, Dryden D, Palmer L, Rona R, Simms A, Fear NT, et al. An expert and veteran user assessment of the usability of an alcohol reduction app for military veterans, drinks:ration: a mixed-methods pilot study. Mil Behav Health 2022:1-14. [doi: 10.1080/21635781.2022.2151532]
28. Williamson C, Rona RJ, Simms A, Fear NT, Goodwin L, Murphy D, et al. Recruiting military veterans into alcohol misuse research: the role of social media and Facebook advertising. Telemed J E Health 2023;29(1):93-101. [doi: 10.1089/tmj.2021.0587] [Medline: 35544055]
29. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
30. Betthauser LM, Stearns-Yoder KA, McGarity S, Smith V, Place S, Brenner LA. Mobile app for mental health monitoring and clinical outreach in veterans: mixed methods feasibility and acceptability study. J Med Internet Res 2020;22(8):e15506 [FREE Full text] [doi: 10.2196/15506] [Medline: 32779572]
31. Leightley D, Williamson C, Rona RJ, Carr E, Shearer J, Davis JP, et al. Evaluating the efficacy of the drinks:ration mobile app to reduce alcohol consumption in a help-seeking military veteran population: randomized controlled trial. JMIR Mhealth Uhealth 2022;10(6):e38991 [FREE Full text] [doi: 10.2196/38991] [Medline: 35724966]

Abbreviations

CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
GHQ-12: General Health Questionnaire
MAUQ: mHealth App Usability Questionnaire
MeT4VeT: Mental Health Toolkit for Veterans Project
MOD: Ministry of Defence
MRC: Medical Research Council
PCL-C: PTSD Checklist-Civilian Version
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
UKAF: UK Armed Forces
WEMWBS: Warwick Edinburgh Mental Well-Being Scale
WHOQOL-BREF: World Health Organization Quality of Life Assessment

Edited by J Torous; submitted 14.02.23; peer-reviewed by L Bell, S Scott; comments to author 24.04.23; revised version received 07.06.23; accepted 08.06.23; published 28.08.23.

Please cite as:
Parkes S, Croak B, Brooks SK, Stevelink SAM, Leightley D, Fear NT, Rafferty L, Greenberg N
Evaluating a Smartphone App (MeT4VeT) to Support the Mental Health of UK Armed Forces Veterans: Feasibility Randomized Controlled Trial
JMIR Ment Health 2023;10:e46508
URL: https://mental.jmir.org/2023/1/e46508
doi:10.2196/46508
PMID:37639295

Original Paper

Effect of Individual Virtual Reality Cognitive Training Programs on Cognitive Function and Depression in Middle-Aged Women: Randomized Controlled Trial

Du-Ri Kim[1,2], MSc; EunSoo Moon[1,3,4], MD, PhD; Myung-Jun Shin[1,5,6], MD, PhD; Yeong-Ae Yang[7,8*], PhD; Jong-Hwan Park[1*], PhD

[1]Health Convergence Medicine Laboratory, Biomedical Research Institute, Pusan National University Hospital, Busan, Republic of Korea
[2]Department of Rehabilitation Science, Graduate School, Inje University, Gimhae, Republic of Korea
[3]Department of Psychiatry, Pusan National University School of Medicine, Yangsan, Republic of Korea
[4]Department of Psychiatry, Pusan National University Hospital, Busan, Republic of Korea
[5]Department of Rehabilitation Medicine, Pusan National University School of Medicine, Yangsan, Republic of Korea
[6]Department of Rehabilitation Medicine, Pusan National University Hospital, Busan, Republic of Korea
[7]Department of Occupational Therapy, College of Biomedical Science and Engineering, Inje University, Gimhae, Republic of Korea
[8]Institute of Aged Life Redesign, Inje University, Gimhae, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Jong-Hwan Park, PhD
Health Convergence Medicine Laboratory, Biomedical Research Institute, Pusan National University Hospital
179, Gudeok-ro, Seo-gu
Busan, 49241
Republic of Korea
Phone: 82 0512407071
Email: parkj@pusan.ac.kr

Abstract

Background: Given the increasing incidence of early-onset Alzheimer disease, strategies for early diagnosis and swift treatment interventions are crucial for mitigating cognitive problems in women and middle-aged individuals who face a high risk of cognitive impairment.

Objective: This study aimed to assess the effectiveness of individual cognitive training programs based on virtual reality (VR), a nonpharmacological intervention, on cognitive function and depression in middle-aged women at risk of cognitive impairment. It used VR technology, which has recently been recognized as a promising tool.

Methods: We administered a VR-based cognitive training program for 30 minutes daily, twice a week, for 12 weeks (24 sessions). This study included middle-aged women residing in older adults' welfare facilities in G-gu, Busan, from May to August 2021. A total of 60 participants were randomly divided into the training (n=30) and control (n=30) groups. Cognitive and depressive functions were assessed using the Korean versions of the Montreal Cognitive Assessment (K-MoCA), Digit Span Test (DST), Korean-Color Word Stroop Test (K-CWST), and Short Form of the Geriatric Depression Scale (SGDS-K) before the intervention. The training group underwent a VR-based cognitive training program, whereas the control group was educated to maintain regular daily activities. The same assessments were performed 12 weeks after treatment.

Results: A comparison of the mean scores before and after K-MoCA in the training group revealed a significant increase from 24.87 (SD 2.62) to 27.50 (SD 1.70; $P<.01$), indicating substantial cognitive improvement. Similarly, the mean DST forward scores increased significantly from 6.97 (SD 1.10) to 7.90 (SD 1.18; $P<.01$), suggesting enhanced short-term auditory memory and attention. The mean DST backward scores also showed a significant improvement from 4.10 (SD 0.71) to 4.77 (SD 1.2; $P=.01$). Notably, the mean SGDS-K scores decreased significantly from 3.97 (SD 2.51) to 2.13 (SD 1.87; $P<.01$), indicating a reduction in depression within the training group.

Conclusions: The VR-based cognitive training programs significantly enhanced cognitive function and reduced depression in middle-aged women. Consequently, these programs are considered beneficial nonpharmacological cognitive training interventions for middle-aged women at high risk of cognitive impairment.

Trial Registration: UMIN Clinical Trials Registry UMIN000049752; https://tinyurl.com/z5du989z

(JMIR Ment Health 2023;10:e48912) doi:10.2196/48912

KEYWORDS

cognitive function; depression; middle aged; virtual reality; women

Introduction

Cognitive aging begins in middle-aged women, leading to a decline in processing speed and memory, particularly in postmenopausal women. Additionally, the risk of estrogen loss and neurological decline following menopause is increased during this period, subsequently elevating the susceptibility of middle-aged women to Alzheimer disease compared with men [1,2]. Alzheimer disease is the most prevalent form of dementia and a significant global public health concern [3].

Recent investigations into early-onset Alzheimer disease (EOAD) have revealed a lack of domestic research on nonpharmaceutical approaches to EOAD, underscoring the pressing need for further research owing to the insufficient verification of long-term effects [4]. Given the increasing annual incidence of EOAD, there is an imperative demand for strategies geared toward early diagnosis and swift treatment interventions to combat cognitive issues in middle-aged individuals and women, both of whom face heightened risks of cognitive impairment. Recent discussions have emphasized the need for research aimed at evaluating the effectiveness of nonpharmacological cognitive interventions [5]. Cognitive training and behavioral interventions that do not involve pharmaceutical components have been shown to be effective in alleviating behavioral symptoms such as depression, apathy, wandering, sleep disturbances, and aggression. These interventions play a crucial role in preserving cognitive function and reducing cognitive decline [6]. The Lancet Commission's 2020 report examined the connection between depression and cognitive impairment [7]. This highlights the link between depression and the development of dementia through various psychological and physiological mechanisms, particularly during the early stages of dementia. Therefore, engaging in cognitive activities during middle age is recommended to reduce the risk of dementia through behavioral changes.

Novel virtual reality (VR)–based technologies have emerged to enhance the efficacy of cognitive training. This technology immerses the user in virtual environments and transmits various visual, auditory, and sensory stimuli through a head-mounted display (HMD) [8]. VR is a compelling approach for individuals with cognitive impairment, where regular training yields increased effectiveness, positioning it as a promising avenue for dementia treatment [9].

VR-based cognitive training harnesses multiple senses, uses engaging gaming formats, drives strong motivation, and improves cognitive function. These interventions have demonstrated favorable outcomes, particularly in memory, executive function, and various cognitive domains [10,11]. Nonetheless, given the individualized nature of age-related cognitive function decline, it is imperative to develop individually tailored cognitive training programs.

Consequently, this study sought to ascertain the effectiveness of VR-based individual cognitive training programs, which fall under the umbrella of nonpharmacological interventions, in addressing cognitive function and depression among middle-aged women at high risk of cognitive impairment. This endeavor has leveraged VR technology, which has recently been recognized as a promising tool in this context.

Methods

Recruitment

For 12 weeks, from May to August 2021, the objectives and methodology of this study were presented at 3 older adults' welfare facilities situated in G-gu. The recruitment of participants involved the issuance of recruitment notices to middle-aged women 40-65 years old who expressed a voluntary willingness to participate and met specific selection criteria. The inclusion criteria included individuals who (1) visited organizations and centers, (2) had no physical disabilities, (3) were able to read and write for evaluation, (4) had the ability to understand the purpose of this test and voluntarily agreed by signing the consent form, or (5) had a Mini-Mental State Examination (MMSE) score of 24-28 points. Patients (1) who had difficulty wearing an HMD and operating controller equipment owing to severe or unstable physical conditions and who had medical conditions that may interfere with the completion of clinical trials, (2) with other degenerative brain diseases or mental illnesses such as depression, (3) with drug and alcohol addiction, and (4) with hearing and vision impairment that could not be evaluated effectively were excluded. After determining the appropriate number of participants, based on the effect size calculated in previous studies involving VR-based cognitive training, only 34 patients in each group exhibited significant improvements in cognitive function [12].

Study Design

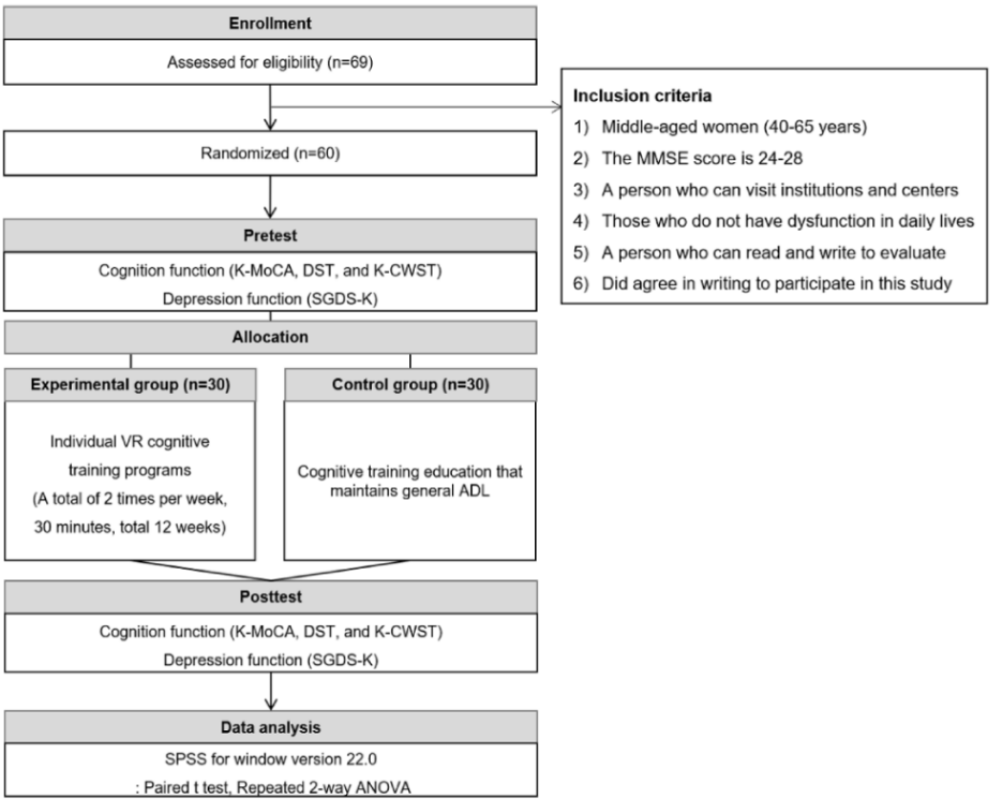
This study used a pre-post experimental design with a randomized control group to assess the impact of VR-based individual cognitive training programs on cognitive function and depression in middle-aged women. Middle-aged women between the ages of 40 and 65 years who voluntarily expressed interest in participating and met the specified selection criteria were recruited. Participants were randomly allocated to either the experimental group (n=30) or the control group (n=30) using block randomization facilitated by SAS programming (SAS Institute), maintaining an equal 1:1 ratio. The VR cognitive training program was structured as a comprehensive cognitive function training regimen. The experimental group participated in a VR-based individual cognitive training program twice a

week for 12 weeks, with each session lasting 30 minutes. The participants in the control group were instructed to continue their regular daily activities and cognitive training throughout the study (Table 1). Assessments were performed before the initiation of the VR-based training program (baseline) and after the intervention (follow-up). The procedural flow of the study is shown in Figure 1. This trial was registered under the identifier UMIN000049752 in the UMIN Clinical Trial Registration database. All educational and training activities in this study were conducted by researchers with a strong understanding of the equipment and occupational therapy and extensive experience in operating the equipment. To assess the effects of VR-based cognitive training programs on cognitive function and depression in middle-aged women, several standardized tests were used, including the Korean version of the Montreal Cognitive Assessment (K-MoCA), Digit Span Test (DST), Korean Color Word Stroop Test (K-CWST), and the Korean version of the Short Form of Geriatric Depression Scale (SGDS-K; Figure 1).

Table 1. Cognitive training education given to the control group.

Cognitive domain	Duration (weeks)	Cognitive training education during activities of daily living
Memory	1-3	• Habit of taking notes on important matters • Save it on your phone • Habit of putting things in the same place
Attention	4-6	• Focus on one thing • When concentrating, move to a quiet place and deal with it
Language	7-9	• Prepare in advance and practice loudly before the meeting • Listen carefully and follow the story
Mix and executive function	10-12	• Mental arithmetic • Memorizing well-known sayings • Eat food that helps your brain

Figure 1. Flowchart of the process of selecting the participants and measurement methods. ADL: activities of daily living; DST: Digit Span Test; K-CWST: Korean-Color Word Stroop Test; K-MoCA: The Korean version of the Montreal Cognitive Assessment; MMSE: Mini-Mental State Examination; SGDS-K: The Korean version of the Short Form of Geriatric Depression Scale; VR: virtual reality.



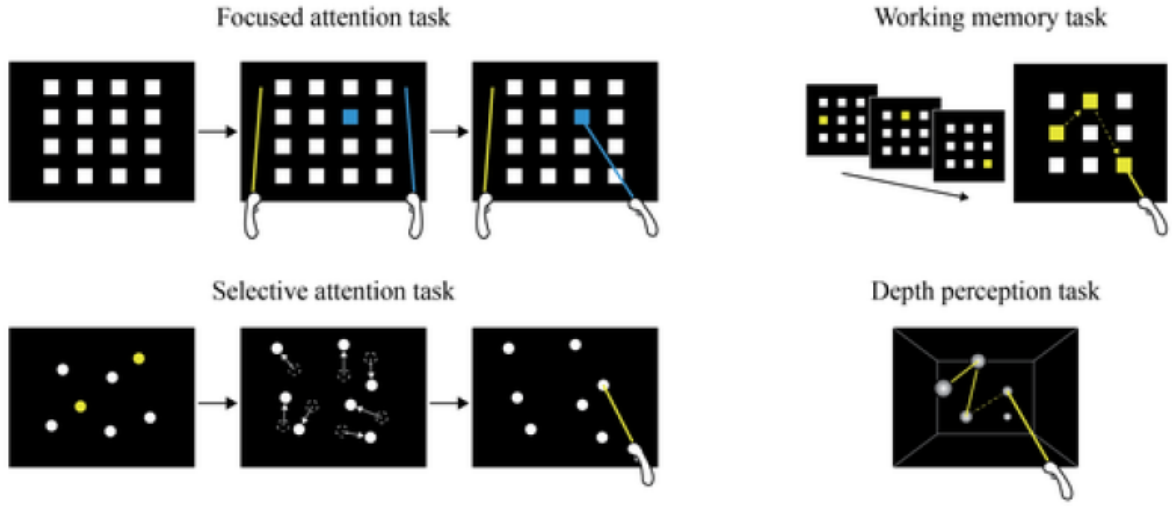
Measurements

VR-Based Cognitive Evaluation

The VR system was used to assess the participants' cognitive function and enhance the areas of cognitive deficiency through a VR-based cognitive training program. The assessment encompassed persistent attention, selective attention, working memory, and depth perception, all of which were conducted in a VR environment (Figures 2 and 3).

Figure 2. Virtual reality (VR) device system. PC: personal computer.

Figure 3. Cognitive training for each domain using virtual reality.



Immediate Object Location

The visual cognitive assessment test was conducted in a VR environment and comprised 2 phases: memory and recall. During the memory phase, participants were instructed to freely explore a 3D space and memorize the types and locations of objects within that space. Subsequently, in the recall phase, the participants were presented with a scene in which some of the objects from the memory phase had disappeared, and their challenge was to identify and reposition the objects they had observed in their original locations.

Recall

After a cognitive evaluation involving the adaptation of the Corsi block-tapping test to VR, participants were asked to recall a sequence of cubes displayed on the screen as they lit sequentially [13]. This task included both forward recall, in which participants remembered the exact order of cube activation, and backward recall, in which they recalled the reverse order of cube activation.

Text Reading

As part of the cognitive evaluation, 2 passages were presented to analyze participants' reading patterns. Participants were free to read and interact with the passages.

Tricky Ball

In the cognitive evaluation involving multiple balls of the same shape, several balls were illuminated with different colors. Subsequently, the balls were blinked and moved about, and the participants were tasked with tracking the balls that were illuminated in different colors without missing any. Participants then transitioned to sessions in both 2D and 3D settings to observe any differences between the 2 and select the illuminated balls while at rest.

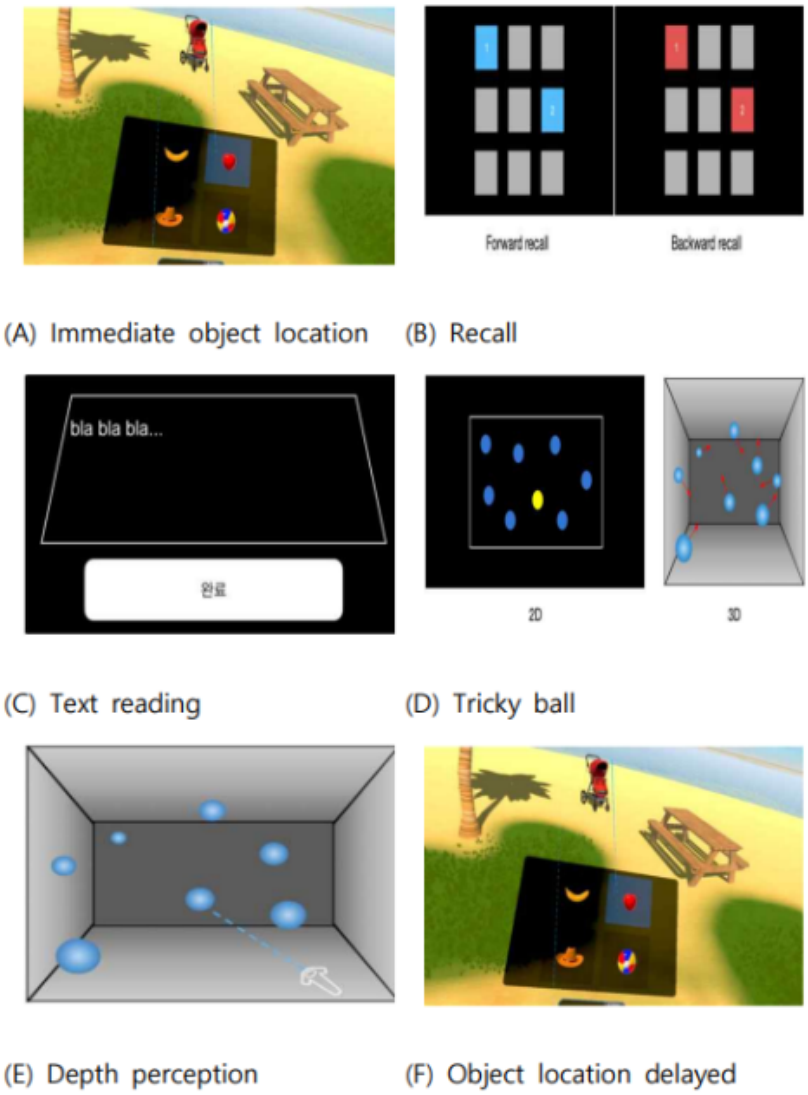
Depth Perception

Participants were required to make selections based on the order of proximity to the balls dispersed in space or the order of proximity to the balls as perceived in terms of depth. The outcomes of these selections were examined.

Delayed Object Location

Building on the cognitive evaluation conducted immediately after the object location assessment, the participants were tasked with identifying items that were not present in the initial beach scene (Figure 4).

Figure 4. Cognitive training using virtual reality: (A) immediate object location, (B) recall, (C) text reading, (D) tricky ball, (E) depth perception, and (F) object location delayed.



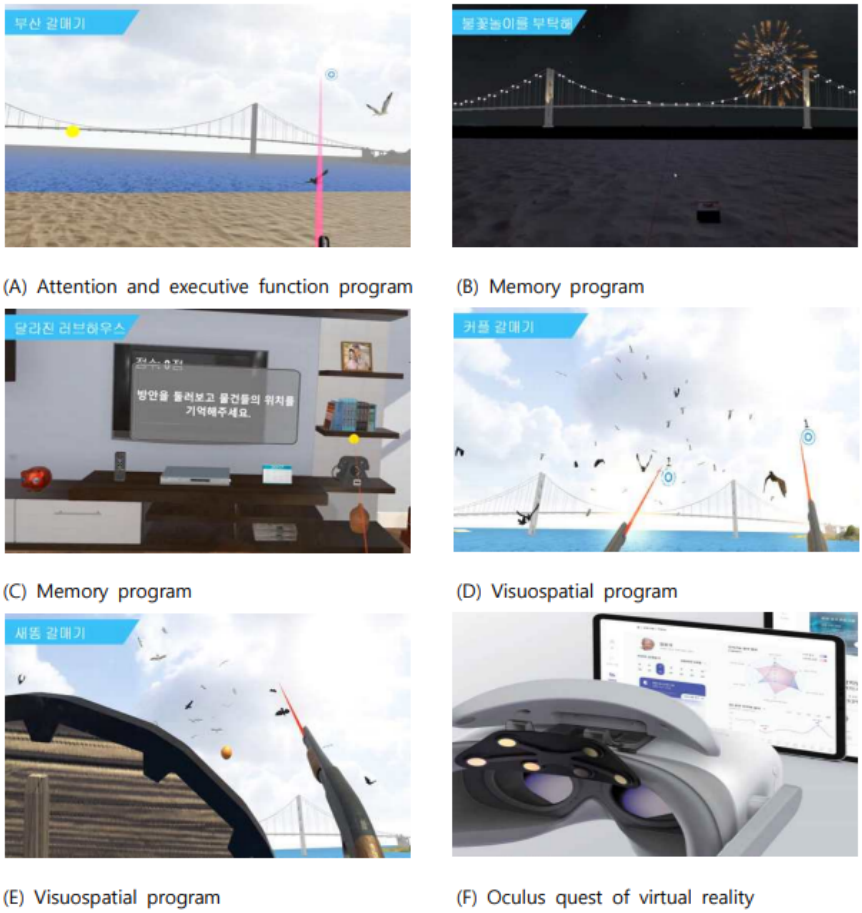
VR-Based Individual Cognitive Training Program

The VR-based cognitive training program was administered for 30 minutes per individual over 24 sessions. It was specifically designed to prevent dementia and enhance cognitive function [14,15]. VR-based cognitive training programs encompass 3 categories: attention and executive function, memory, and visuospatial function.

- Busan Seagull Content: This program focuses on improving attention and executive functions. Participants were trained to spot crows flying amid seagulls.
- Firework Content: Geared toward enhancing working memory, this program tasked individuals with recalling a sequence in which fireworks exploded.
- Changed Love-House Content: Designed to enhance spatial memory, this program involved individuals in finding items within a changing room.
- Coupled Seagull Content: To improve visuospatial skills, this program challenged participants to simultaneously catch birds with both hands.
- Bird's Poo Seagull Content: Targeting visuospatial function improvement, which required participants to catch crows with one hand while shielding themselves against flying bird excrement.

The VR-based cognitive training program was personalized to match each participant's cognitive function by adjusting the difficulty level accordingly (Figure 5).

Figure 5. Cognitive training programs using virtual reality: (A) attention and executive function program, (B) memory program, (C) memory program, (D) visuospatial program, (E) visuospatial program, and (F) oculus quest of virtual reality.



Cognitive Function

Cognitive function was assessed using the MMSE-2nd edition, adapted for use with older adults in Korea, based on the original MMSE developed by Folstein et al [16] and Kang et al [17]. K-MoCA was created by Kang et al [18] as a quicker alternative to the MoCA developed by Nasreddine et al [19]. The DST is a component of the Korean Wechsler Adult Intelligence Scale, which assesses short-term auditory memory and attention [20]. The K-CWST measures the ability to suppress automatic responses, selectively allocate attention, and shift between tasks, thus evaluating executive function [21].

Depression Function

Depression was assessed using the SGDS-K, an adaptation designed to evaluate depression in older adults based on the Geriatric Depression Scale, originally developed by Yesavage et al [22] and Cho et al [23].

Statistical Analysis

SPSS (version 22.0; IBM Corp) was used for statistical analysis of all data. Descriptive statistics, including mean (SD), were used to summarize the dependent variables within each group. A frequency analysis was conducted using descriptive statistics to compare the general characteristics of the study participants. The general characteristics of the participants were analyzed using an independent *t* test for continuous variables. A paired *t* test was used to assess changes in cognitive function and depression status in middle-aged women before and after the implementation of the VR-based individual cognitive training programs. Additionally, we conducted a repeated 2-way ANOVA to ascertain the average differences between the 2 groups. The significance level was set at $P<.05$ in all analyses.

Ethical Considerations

This study was approved by the institutional review board of Inje University (approval number 2021-05-062-005).

Results

Participants' General Characteristics

The study consisted of female participants, with a mean age of 55.17 (SD 5.81) years in the training group and 55.53 (SD 5.66) years in the control group. In the training group (n=30), 4 (13%) participants had diabetes, whereas in the control group (n=30), 2 (7%) participants had diabetes. High blood pressure was observed in 5 (17%) participants in the training group and 6 (20%) participants in the control group. Education levels averaged 12.7 (SD 2.65) years in the training group and 12.63 (SD 2.87) years in the control group. The selection criteria, as measured by the MMSE scores, were mean 26.8 (SD 1.16) points in the training group and mean 26.1 (SD 1.56) points in the control group. Notably, there were no significant differences in general characteristics between the 2 groups (Table 2). Regarding the K-MoCA scores, the training group averaged

24.87 (SD 2.62) points, while the control group averaged 24.1 (SD 1.75) points. In SGDS-K, the training group scored a mean of 3.97 (SD 2.51) points, while the control group scored a mean of 3.17 (SD 2.38) points. The homogeneity test conducted between the training and control groups revealed no significant differences (Table 3).

Table 2. Baseline sociodemographic characteristics.

Characteristics	Control (n=30)	Intervention (n=30)	*P* value[a]
Age (years), mean (SD)	55.53 (5.66)	55.17 (5.81)	.81
Alcohol consumed, n (%)	13 (43)	11 (37)	.61
Smoking, n (%)	0 (0)	0 (0)	—[b]
Diabetes mellitus, n (%)	2 (7)	4 (13)	.40
Hypertension, n (%)	6 (20)	5 (17)	.74
Education levels (years), mean (SD)	12.63 (2.87)	12.7 (2.65)	.93
MMSE[c] (points), mean (SD)	26.1 (1.56)	26.8 (1.16)	.05

[a]*P*<.05 is considered significant.

[b]Not applicable.

[c]MMSE: Mini-Mental State Examination.

Table 3. Baseline cognitive function and depression.

Function (points)	Control (n=30), mean (SD)	Intervention (n=30), mean (SD)	*P* value[a]
K-MoCA[b]	24.1 (1.75)	24.87 (2.62)	.19
DST[c] forward	6.53 (1.14)	6.97 (1.1)	.14
DST backward	4.4 (1.22)	4.1 (0.71)	.25
K-CWST WR[d]	111.37 (1.1)	111.77 (0.57)	.08
K-CWST CR[e]	105.37 (9.68)	105.9 (9.07)	.83
SGDS-K[f]	3.17 (2.38)	3.97 (2.51)	.21

[a]*P*<.05 is considered significant.

[b]K-MoCA: Korean version of the Montreal Cognitive Assessment.

[c]DST: Digit Span Test.

[d]K-CWST WR: Korean-Color Word Stroop Word Reading Test.

[e]K-CWST CR: Korean-Color Word Stroop Word Color.

[f]SGDS-K: Korean version of the Short Form of Geriatric Depression Scale.

Results of Analysis of Differences in Cognitive and Depression Function Before and After Intervention

Changes in cognitive function and depression before and after the VR-based individual cognitive training programs are presented in Table 4. A comparison of the mean scores before and after K-MoCA revealed a significant increase from 24.87 (SD 2.62) to 27.50 (SD 1.70) in the training group and from 24.10 (SD 1.75) to 24.50 (SD 2.67) in the control group. Similarly, when comparing the average scores before and after DST forward, a significant difference was observed, as the mean scores increased from 6.97 (SD 1.10) to 7.90 (SD 1.18) in the training group and from 6.53 (SD 1.14) to 6.60 (SD 1.16) in the control group. The same pattern was observed for DST backward, where the mean scores increased significantly from 4.10 (SD 0.71) to 4.77 (SD 1.22) in the training group and from 4.40 (SD 1.22) to 4.23 (SD 0.86) in the control group. In contrast, while the mean scores before and after K-CWST WR increased slightly from 111.77 (SD 0.57) to 111.97 (SD 0.18) in the training group and from 111.37 (SD 1.10) to 111.73 (SD 0.64) in the control group, no significant difference was detected. Likewise, no significant difference was observed in the average scores before and after K-CWST CR, with the mean scores increasing from 105.90 (SD 9.07) to 107.80 (SD 8.61) in the training group and from 105.37 (SD 9.68) to 107.20 (SD 7.96) in the control group. However, a significant difference was identified in the mean scores before and after SGDS-K, as they decreased from 3.97 (SD 2.51) to 2.13 (SD 1.87) in the training group and from 3.17 (SD 2.38) to 3.03 (SD 1.35) in the control group, signifying a notable reduction in depression.

Table 4. Changes in measurements of cognitive function and depression at baseline and after 12 weeks in the training and control groups.

Variables and groups	Pretest, mean (SD)	Posttest, mean (SD)	Difference	*t* test (*df*)	*P* value[a]
K-MoCA[b] (points)					.01
Control	24.1 (1.75)	24.5 (2.67)	0.4	–1.379 (29)	.18
Intervention	24.87 (2.62)	27.5 (1.7)	2.63	–5.046 (29)	.001
DST[c] forward (points)					.01
Control	6.53 (1.14)	6.6 (1.16)	0.07	–0.403 (29)	.69
Intervention	6.97 (1.1)	7.9 (1.18)	0.93	–4.157 (29)	.001
DST backward (points)					.01
Control	4.4 (1.22)	4.23 (0.86)	–0.17	0.841 (29)	.41
Intervention	4.1 (0.71)	4.77 (1.22)	0.67	–3.247 (29)	.003
K-CWST WR[d] (points)					.41
Control	111.37 (1.1)	111.73 (0.64)	0.36	–2.164 (29)	.04
Intervention	111.77 (0.57)	111.97 (0.18)	0.2	–1.795 (29)	.08
K-CWST CR[e] (points)					.96
Control	105.37 (9.68)	107.2 (7.96)	1.83	–1.515 (29)	.14
Intervention	105.9 (9.07)	107.8 (8.61)	1.9	–2.419 (29)	.02
SGDS-K[f] (points)					<.001
Control	3.17 (2.38)	3.03 (1.35)	–0.14	0.403 (29)	.69
Intervention	3.97 (2.51)	2.13 (1.87)	–1.84	5.061 (29)	.001

[a]*P*<.05 is considered significant.

[b]K-MoCA: Korean version of the Montreal Cognitive Assessment.

[c]DST: Digit Span Test.

[d]K-CWST WR: Korean-Color Word Stroop Word Reading Test.

[e]K-CWST CR: Korean-Color Word Stroop Word Color.

[f]SGDS-K: Korean version of the Short Form of Geriatric Depression Scale.

Discussion

This study investigated the impact of VR-based individual cognitive training programs on cognitive function and depressive symptoms in middle-aged women at risk of cognitive impairment. After the VR-based cognitive training programs conducted before and after this study, the scores in the training group showed a significant improvement compared with those in the control group on both the K-MoCA and DST assessments, confirming a noteworthy difference.

Previous research suggests that VR-based cognitive training is effective in delaying or preventing dementia among middle-aged individuals at high risk of developing the condition. This efficacy can be attributed to its ability to strengthen the cognitive reserve in the brain and stimulate neuroplasticity [24,25]. Moreover, VR-based cognitive training interventions are particularly suited to enhance executive functions such as attention and memory [26]. The VR-based cognitive training program used in this study included attention and executive function, work and spatial memory, and visuospatial ability training. Consequently, consistent participation in cognitive training contributes to the overall improvement in cognitive function.

Conventional rehabilitation training can become monotonous and uninspiring, potentially reducing the participants' ability to concentrate on training. Conversely, game-based cognitive training, as implemented in this VR program, effectively bolsters concentration by making the training more engaging and enjoyable.

In addition, the VR measurement system used for VR-based cognitive evaluations provides immediate visual feedback on cognitive function scores and areas of cognitive deficiency, thereby enhancing immersion and motivation within the VR environment. This domain-specific cognitive training intervention is believed to play a pivotal role in improving cognitive function.

Visual and auditory attention are essential for performing cognitive tasks. VR-based cognitive training interventions enhance connectivity between the frontal and occipital brain regions by engaging in temporal and spatial functions, resulting in heightened realism, presence, and immersion [27]. Consequently, cognitive training that consistently stimulates these interconnected brain regions through a visual, auditory, or spatial multisensory approach within a VR environment is

considered a crucial strategy to prevent cognitive function decline.

In the SGDS-K test, conducted to evaluate depressive symptoms after the intervention, scores in the training group exhibited a significant reduction compared with those in the control group, indicating a substantial difference. Previous research has demonstrated that behavioral activation and physical activity can lead to a reduction in depressive symptoms [28].

In this study, HMDs were used to induce immersion by aligning head movements in the direction of gaze. Furthermore, the continuous upper-limb movement facilitated by the controller expands the range of motion, contributing to increased physical activity. VR technology offers immediate feedback, fosters enjoyment, and piques interest, thereby enhancing the participants' motivation for rehabilitation. Consequently, the competitive and cooperative dynamics observed among participants during gaming activities were effective in reducing depression as they promoted motivation.

The VR-based cognitive training program, designed as an engaging gaming experience in a virtual environment, is expected to motivate participants by immersing them in an alternative reality. This approach encourages physical activity and behavior, potentially alleviating the symptoms of depression.

Limitations

This study has some limitations. First, it focused exclusively on women, and the sample size was relatively small, limiting the generalizability of the results. Moreover, individual variations in cognitive aging among middle-aged women do not account for all potential factors contributing to dementia risk. Conditions such as high blood pressure and diabetes can negatively affect various cognitive functions, including reaction time, processing speed, and working memory evaluation [29], and are known risk factors for dementia in middle-aged women. Additionally, factors such as increased weight, abdominal obesity, reduced physical strength, and poor overall health are associated with dementia risk [30,31]. Future research should aim to establish clinical significance by including a larger and more diverse participant pool while also developing interventions that take into account multiple risk factors for dementia. Furthermore, exploring whether VR-based cognitive training programs can effectively manage these risk factors in middle-aged women at risk of cognitive impairment could lead to more effective early prevention strategies.

Conclusions

This study aimed to examine the impact of VR-based individual cognitive training programs on cognitive function and depression in middle-aged women to address the risk of cognitive impairment. The findings of this study confirmed the effectiveness of VR-based cognitive training programs in enhancing cognitive function and reducing depression in middle-aged women at risk of cognitive impairment. This nonpharmacological cognitive training program is considered beneficial for improving cognitive function and alleviating depression in this population. This offers a promising avenue for early prevention of cognitive impairment in middle-aged women facing such risks.

Acknowledgments

This study was supported by a Biomedical Research Institute Grant (20210024) from Pusan National University Hospital.

Conflicts of Interest

None declared.

Authors' Contributions

YAY(otyya62@inje.ac.kr) and JHP (parkj@pusan.ac.kr) are the co-corresponding authors of this article.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1240 KB - mental_v10i1e48912_app1.pdf]

References

1. Karlamangla AS, Lachman ME, Han W, Huang M, Greendale GA. Evidence for cognitive aging in midlife women: study of women's health across the nation. PLoS One 2017;12(1):e0169008 [FREE Full text] [doi: 10.1371/journal.pone.0169008] [Medline: 28045986]
2. Barha CK, Liu-Ambrose T. Sex differences in exercise efficacy: is midlife a critical window for promoting healthy cognitive aging? FASEB J 2020;34(9):11329-11336 [FREE Full text] [doi: 10.1096/fj.202000857R] [Medline: 32761860]
3. Prince MJ, Wimo A, Guerchet MM, Ali GC, Wu YT, Prina M. World Alzheimer Report 2015: The Global Impact of Dementia : an Analysis of Prevalence, Incidence, Cost and Trends. London: Alzheimer's Disease International; 2015.
4. Won KA, Lee MA, Sim JY. Characteristics and non-pharmacological approach of early-onset Dementia(young-onset Dementia). Korea J Geriatr Occup Ther 2020;2(1):53-65.
5. Kim HJ. Introduction of non-pharmacological treatment for cognitive preservation in elderly dementia patients. J Occup Ther Aged Dement 2019;13(2):41-46. [doi: 10.34263/jsotad.2019.13.2.41]

6. Alzheimer's Association. 2013 Alzheimer's disease facts and figures. Alzheimers Dement 2013;9(2):208-245 [FREE Full text] [doi: 10.1016/j.jalz.2013.02.003] [Medline: 23507120]
7. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the lancet commission. Lancet 2020;396(10248):413-446 [FREE Full text] [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
8. Chua SIL, Tan NC, Wong WT, Allen JC, Quah JHM, Malhotra R, et al. Virtual reality for screening of cognitive function in older persons: comparative study. J Med Internet Res 2019;21(8):e14821 [FREE Full text] [doi: 10.2196/14821] [Medline: 31373274]
9. Diaz-Perez E, Florez-Lozano JA. [Virtual reality and dementia]. Rev Neurol 2018;66(10):344-352 [FREE Full text] [Medline: 29749595]
10. Kim MY, Lee KS, Choi JS, Kim HB, Park CI. Effectiveness of cognitive training based on virtual reality for the elderly. J Korean Acad Rehabil Med 2005;29(4):424-433 [FREE Full text]
11. Ge S, Zhu Z, Wu B, McConnell ES. Technology-based cognitive training and rehabilitation interventions for individuals with mild cognitive impairment: a systematic review. BMC Geriatr 2018;18(1):213 [FREE Full text] [doi: 10.1186/s12877-018-0893-1] [Medline: 30219036]
12. Liao YY, Tseng HY, Lin YJ, Wang CJ, Hsu WC. Using virtual reality-based training to improve cognitive function, instrumental activities of daily living and neural efficiency in older adults with mild cognitive impairment. Eur J Phys Rehabil Med 2020;56(1):47-57 [FREE Full text] [doi: 10.23736/S1973-9087.19.05899-4] [Medline: 31615196]
13. Berch DB, Krikorian R, Huha EM. The corsi block-tapping task: methodological and theoretical considerations. Brain Cogn 1998;38(3):317-338 [FREE Full text] [doi: 10.1006/brcg.1998.1039] [Medline: 9841789]
14. Kim DR, Kim GM, Song SH, Huh U, Chang JH, Tak YJ, et al. Effect of virtual reality-based cognition training on the cognitive function of the elderly adults in the community who use Dementia Care Center. J Occup Ther Aged Dement 2020;14(2):5-14. [doi: 10.34263/jsotad.2020.14.2.1]
15. Park JH, Liao Y, Kim DR, Song S, Lim JH, Park H, et al. Feasibility and tolerability of a culture-based Virtual Reality (VR) training program in patients with mild cognitive impairment: a randomized controlled pilot study. Int J Environ Res Public Health 2020;17(9):3030 [FREE Full text] [doi: 10.3390/ijerph17093030] [Medline: 32349413]
16. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12(3):189-198 [FREE Full text] [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
17. Kang Y, Na DL, Hahn S. A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. J Korean Neurol Assoc 1997:300-308 [FREE Full text]
18. Kang Y, Park J, Yu K, Lee B. A reliability, validity, and normative study of the Korean-Montreal Cognitive Assessment (K-MoCA) as an instrument for screening of vascular Cognitive Impairment (VCI). Korean J Clin Psychol 2009;28(2):549-562 [FREE Full text] [doi: 10.15842/kjcp.2009.28.2.013]
19. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53(4):695-699 [FREE Full text] [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
20. Yeom T, Park Y, Oh-Hashi K, Kim J, Lee Y. The Manual of Korean-Wechsler Adult Intelligence Scale. Seoul: Korea Guidance; 1992:105-119.
21. Kang Y, Jahng S, Na TL. Seoul Neuropsychological Screening Battery (SNSB-II). 2nd Edition. Incheon: Human Brain Research & Consulting Co; 2020.
22. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49 [FREE Full text] [doi: 10.1016/0022-3956(82)90033-4] [Medline: 7183759]
23. Cho MJ, Bae JN, Suh GH, Hahm BJ, Kim JK, Lee DW, et al. Validation of Geriatric Depression Scale, Korean version (GDS) in the assessment of DSM-III-R major depression. J Korean Neuropsychiatr Assoc 1999:48-63 [FREE Full text]
24. Doniger GM, Beeri MS, Bahar-Fuchs A, Gottlieb A, Tkachov A, Kenan H, et al. Virtual reality-based cognitive-motor training for middle-aged adults at high Alzheimer's disease risk: a randomized controlled trial. Alzheimers Dement (N Y) 2018;4:118-129 [FREE Full text] [doi: 10.1016/j.trci.2018.02.005] [Medline: 29955655]
25. Teo WP, Muthalib M, Yamin S, Hendy AM, Bramstedt K, Kotsopoulos E, et al. Does a combination of virtual reality, neuromodulation and neuroimaging provide a comprehensive platform for neurorehabilitation? - a narrative review of the literature. Front Hum Neurosci 2016;10:284 [FREE Full text] [doi: 10.3389/fnhum.2016.00284] [Medline: 27445739]
26. Gamito P, Oliveira J, Alves C, Santos N, Coelho C, Brito R. Virtual reality-based cognitive stimulation to improve cognitive functioning in community elderly: a controlled study. Cyberpsychol Behav Soc Netw 2020;23(3):150-156 [FREE Full text] [doi: 10.1089/cyber.2019.0271] [Medline: 32031888]
27. Kang JM, Kim N, Lee SY, Woo SK, Park G, Yeon BK, et al. Effect of cognitive training in fully immersive virtual reality on visuospatial function and frontal-occipital functional connectivity in predementia: randomized controlled trial. J Med Internet Res 2021;23(5):e24526 [FREE Full text] [doi: 10.2196/24526] [Medline: 33955835]

28. Soucy I, Provencher M, Fortier M, McFadden T. Efficacy of guided self-help behavioural activation and physical activity for depression: a randomized controlled trial. Cogn Behav Ther 2017;46(6):493-506 [FREE Full text] [doi: 10.1080/16506073.2017.1337806] [Medline: 28644740]
29. Pavlik VN, Hyman DJ, Doody R. Cardiovascular risk factors and cognitive function in adults 30-59 years of age (NHANES III). Neuroepidemiology 2005;24(1-2):42-50 [FREE Full text] [doi: 10.1159/000081049] [Medline: 15459509]
30. Ma Y, Ajnakina O, Steptoe A, Cadar D. Higher risk of dementia in english older individuals who are overweight or obese. Int J Epidemiol 2020;49(4):1353-1365 [FREE Full text] [doi: 10.1093/ije/dyaa099] [Medline: 32575116]
31. Hörder H, Johansson L, Guo X, Grimby G, Kern S, Östling S, et al. Midlife cardiovascular fitness and dementia: a 44-year longitudinal population study in women. Neurology 2018;90(15):e1298-e1305 [FREE Full text] [doi: 10.1212/WNL.0000000000005290] [Medline: 29540588]

Abbreviations

DST: Digit Span Test
EOAD: early-onset Alzheimer disease
HMD: head-mounted display
K-CWST: Korean-Color Word Stroop Test
K-MoCA: The Korean version of the Montreal Cognitive Assessment
MMSE: Mini-Mental State Examination
SGDS-K: The Korean version of the Short Form of Geriatric Depression Scale
VR: virtual reality

Edited by J Torous; submitted 11.05.23; peer-reviewed by A Wang, S Schüssler; comments to author 13.07.23; revised version received 04.09.23; accepted 12.09.23; published 25.10.23.

Please cite as:
Kim DR, Moon E, Shin MJ, Yang YA, Park JH
Effect of Individual Virtual Reality Cognitive Training Programs on Cognitive Function and Depression in Middle-Aged Women: Randomized Controlled Trial
JMIR Ment Health 2023;10:e48912
URL: https://mental.jmir.org/2023/1/e48912
doi:10.2196/48912
PMID:37878378

Original Paper

Momentary Self-esteem as a Process Underlying the Association Between Childhood Trauma and Psychosis: Experience Sampling Study

Maud Daemen[1,2], PhD; Therese van Amelsvoort[1], Prof Dr; GROUP investigators[3]; Ulrich Reininghaus[1,2,4], Prof Dr

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands

[2]Department of Public Mental Health, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany

[3]See Acknowledgments

[4]Centre for Epidemiology and Public Health, Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

Corresponding Author:
Maud Daemen, PhD
Department of Psychiatry and Neuropsychology
School for Mental Health and Neuroscience
Maastricht University
Vijverdalseweg 1
Maastricht, 6226 NB
Netherlands
Phone: 31 (0)43 38838
Email: maud.daemen@maastrichtuniversity.nl

Abstract

Background: Exposure to childhood trauma is associated with an increased risk of developing and maintaining psychotic symptoms later in life. Self-esteem might be an important psychological process underlying the association between childhood trauma and psychosis, but there is only limited evidence to support this claim, especially in daily life.

Objective: In this study, we aimed to investigate whether exposure to childhood trauma (physical, emotional, and sexual abuse and physical and emotional neglect) moderates the cross-sectional and temporal associations between self-esteem and psychotic experiences in patients with psychotic disorders, their first-degree relatives, and controls.

Methods: We assessed momentary self-esteem and psychotic experiences in daily life using the experience sampling method in 139 patients with psychotic disorders, 118 first-degree relatives of patients with psychotic disorders, and 111 controls. Childhood trauma was measured using the Childhood Trauma Questionnaire. We fitted linear mixed models and added two-way and three-way interaction terms to test the hypotheses.

Results: The association between momentary self-esteem and psychotic experiences in daily life was modified by prior exposure to high versus low levels of several types of childhood trauma, that is, physical (χ^2_2=24.9, family-wise error-corrected P<.001) and sexual abuse (χ^2_2=15.9, P<.001) and physical neglect (χ^2_2=116.7, P<.001). Specifically, momentary self-esteem was associated with more intense psychotic experiences in patients exposed to high versus low levels of physical neglect, in relatives exposed to high versus low levels of physical abuse, and in relatives and controls exposed to high versus low levels of sexual abuse. When investigating temporal order, the results showed no evidence that childhood trauma modified the temporal associations between self-esteem at t_{n-1} and psychotic experiences at t_n or those between psychotic experiences at t_{n-1} and self-esteem at t_n.

Conclusions: The association between self-esteem and psychotic experiences in daily life was found to be stronger in those exposed to high versus low levels of several types of childhood trauma (ie, physical abuse, sexual abuse, and physical neglect).

(JMIR Ment Health 2023;10:e34147) doi:10.2196/34147

KEYWORDS

psychosis; self-esteem; childhood trauma; childhood adversity; experience sampling method; ecological momentary assessment

Introduction

Background

Exposure to childhood trauma can have persistent adverse effects on an individual's well-being, social development, and physical and mental health [1]. Childhood trauma refers to potentially harmful experiences of physical, emotional, and sexual abuse and physical and emotional neglect during childhood [2]. Accumulating evidence suggests that childhood trauma is associated with psychotic disorders [3-6] and increases the risk of developing and maintaining psychotic symptoms later in life [3,7,8]. The study by Norman et al [9] showed that the prevalence of childhood trauma in patients with psychotic disorders is very common; 82 of their 100 participants (patients receiving treatment from an early psychosis clinic) reported exposure to childhood trauma. Similarly, other studies have shown that patients with psychotic disorders reported having experienced disproportionate levels of childhood trauma [3,7,8,10-12]. Furthermore, the results of the meta-analysis conducted by Varese et al [3] suggested that 1 out of 3 patients might not have developed psychosis if adversity were to be eliminated as a risk factor (assuming causality), which indicates that childhood trauma plays a prominent role in the development of psychosis. It has also been demonstrated that psychotic symptoms are more severe in patients exposed to childhood trauma [9,13]. A prospective cohort study indicated that the incidence of psychotic experiences decreased significantly when exposure to trauma ceased over the course of the study [14]. Finally, a systematic review revealed that exposure to childhood adversity was associated with the persistence of psychotic symptoms in both the general population and clinical studies [15], suggesting that childhood adversity may play an important role in the maintenance of psychotic symptoms in patients with enduring psychotic disorders.

Evidence for familial liability to psychosis [16-18] shows that first-degree relatives of patients with psychosis have an increased risk of developing psychotic disorders [17,19]. In addition, although exposure to childhood trauma is higher in patients with psychotic disorders, patients and their siblings share a degree of exposure to childhood trauma [20], as they also share many sociodemographic, parental, and developmental characteristics [21].

Several putative psychological mechanisms and processes have been posited to be involved in the association between childhood trauma and psychosis [22,23]. Self-esteem is one such process. A systematic review indicated that exposure to childhood trauma may contribute to low self-esteem [24]. Low self-esteem has also been found to be common in patients with psychotic disorders [25-28]. Moreover, self-esteem has been shown to be involved in the development and maintenance of psychotic symptoms [28-32]. Lower levels of momentary self-esteem have been found to be associated with an increased intensity of psychotic experiences in daily life [33]. Prior exposure to childhood trauma may affect cognition [34], which contributes to the development of a complex system consisting of negative views of one's self, the world, and the future [35]. Some researchers have suggested that self-esteem is involved in the pathway from negative life events (eg, childhood trauma) to psychosis, but evidence to support this claim remains limited [29,36,37]. A few studies have investigated self-esteem and closely related processes such as negative self-schemas, or negative beliefs about self and others, as an underlying process in the association between childhood trauma and psychotic experiences [38-42]. In contrast, other researchers have suggested low self-esteem to be a product of an individual's experience of psychosis, for example, owing to hospitalization, the loss of their social role or rank, or stigmatization [43,44].

Self-esteem and psychotic experiences are often assessed using cross-sectional measures, with global scores [38-41]. Another way of collecting data on these variables is through the experience sampling method (ESM). ESM assesses moment-to-moment variations in thoughts, feelings, and behaviors in daily life. It generates longitudinal data with a very limited recall bias and high ecological validity [45].

Objective

Overall, there is limited evidence that exposure to childhood trauma moderates the association between self-esteem and psychotic experiences, especially in daily life. Exposure to childhood trauma plays an important role in the development and maintenance of psychosis [17-19,46], but unfortunately, preventing childhood trauma remains challenging. By investigating whether self-esteem is a relevant underlying process in the association between childhood trauma and psychotic experiences, we might be able to identify individuals exposed to (high levels of) childhood trauma in the early stages and target self-esteem to prevent a full-blown psychotic disorder.

Therefore, we aimed to investigate whether cross-sectional and temporal associations of momentary self-esteem and psychotic experiences in daily life were modified by prior exposure to childhood trauma (ie, physical, emotional, and sexual abuse and physical and emotional neglect) in patients with enduring psychotic disorders, their first-degree relatives, and controls. Specifically, using data from the Genetic Risk and Outcome in Psychosis (GROUP) study [47], a large multicenter study of patients with psychotic disorders, their first-degree relatives, and controls without a family history of psychotic disorders, we aimed to test the following hypotheses:

1. In patients, their first-degree relatives, and controls, the magnitude of associations of self-esteem and psychotic experiences in daily life (both measured with the ESM) is stronger in individuals exposed to high levels of each type of childhood trauma (ie, physical, emotional, and sexual abuse and physical and emotional neglect) than in those exposed to low levels of childhood trauma (measured using the Childhood Trauma Questionnaire [CTQ]; hypothesis 1).
2. The difference in the magnitude of associations of self-esteem and psychotic experiences in daily life between those exposed to high levels of childhood trauma and those exposed to low levels of childhood trauma is (1) greater in patients than in controls, (2) greater in relatives than in controls, and (3) greater in patients than in relatives (hypothesis 2).

3. In patients, their first-degree relatives, and controls, the temporal association (1) between self-esteem at t_{n-1} and psychotic experiences at t_n and (2) between psychotic experiences at t_{n-1} and self-esteem at t_n in daily life (both measured with the ESM) is stronger in individuals exposed to high levels of each type of childhood trauma than those exposed to low levels of childhood trauma (measured using the CTQ; hypothesis 3).
4. The difference in the magnitude of temporal associations of (1) self-esteem and psychotic experiences and (2) psychotic experiences and self-esteem in daily life between those exposed to high levels of childhood trauma and those exposed to low levels of childhood trauma is (1) greater in patients than in controls, (2) greater in relatives than in controls, and (3) greater in patients than in relatives (hypothesis 4).

Methods

Sample

A sample of patients with psychotic disorders, their first-degree relatives, and controls without a family history of psychotic disorders was recruited in the GROUP study [47], a large longitudinal multicenter study in the Netherlands and Belgium. Individuals with psychotic disorders were recruited from regional psychosis care facilities or academic centers in selected geographical areas. The patients' relatives were contacted after obtaining written informed consent. Participants in the control group were recruited by contacting random addresses in the same geographical areas as the patients. To be eligible, participants had to be between 16 and 50 years old and their command of Dutch language had to be sufficient. As an additional criterion, patients had to be diagnosed with a nonaffective psychotic disorder according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria [48]. In the control group, individuals with a family history of psychotic disorders were excluded.

Ethics Approval

Ethics approval was provided by the Ethical Review Board of the University Medical Centre Utrecht (METC: 0 4/003- O), and all participants gave written informed consent. For this analyses, only participants who completed the CTQ and, in line with previous ESM studies [49], at least one-third (33.3%) of the ESM assessments were included.

Data Collection

Basic Sample Characteristics

Using a sociodemographic schedule, data on age, gender, ethnicity, marital status, and educational level (adapted from the Dutch Standard Classification of Education) [50] were collected.

Childhood Trauma

The Dutch version of the 25-item CTQ-short version [2] was used to assess childhood trauma at the baseline. The 25 items inquire about 5 types of trauma in childhood (emotional, physical, and sexual abuse and emotional and physical neglect). All 5 types of trauma are covered with 5 items rated on a 5-point Likert scale (1=never true to 5=very often true). The CTQ–short version has shown to be reliable and to provide adequate content coverage. There is also good evidence of criterion validity in both adolescent patients with a psychiatric disorder and individuals from a normative community sample [2]. For the analyses, the mean score for each type of trauma was used.

ESM Procedure

The ESM was used at the 6-year follow-up assessment to collect data on self-esteem and psychotic experiences in daily life. Participants were offered an ESM briefing session that provided detailed instructions on the ESM procedure. Participants received a dedicated digital device (ie, PsyMate, ECS International B.V.), which emitted a beep at 10 semirandom times a day, within 90-minute time blocks, between 7:30 AM and 10:30 PM for 6 consecutive days. Each time the PsyMate emitted a beep, the participants were asked to complete an ESM questionnaire immediately after the beep. A debriefing session was scheduled after 6 days. Research has shown that the ESM in samples of patients with psychosis and their relatives is feasible, reliable, and valid [51-54].

ESM Measures

To assess momentary self-esteem, the mean score of the following 2 ESM items was used: "I like myself" and "I doubt myself" (reversed) [55] (following the approach by Geldhof et al [56] to analyze multilevel reliability; within-person α=.22; between-person α=.76). To assess the intensity of psychotic experiences, the mean score of the following 8 ESM items was used: "My thoughts are influenced by others," "I can't get these thoughts out of my head," "I feel unreal," "My thoughts can't be expressed in words," "I feel suspicious," "I hear voices," "I see things that aren't really there," and "I am afraid I will lose control" [57,58] (within-person α=.64; between-person α=.85). All the items were measured on a 7-point Likert scale, ranging from 1 ("not at all") to 7 ("very much").

Statistical Analysis

For this analyses, we used release 7.0 of the overall GROUP database and release 2.0 of the ESM data. The "mixed" command in Stata 13.0 (StataCorp) was used to fit linear mixed models. Because ESM data include multiple observations within each participant, they were treated as nested within participants and participants nested within families. Restricted maximum likelihood estimation was used to fit the models and estimated the associations between momentary self-esteem as the continuous independent variable and psychotic experiences as the outcome variable. We controlled for the potential confounders of age, gender, ethnicity (people of color or White), level of education, and marital status.

We then added two-way (self-esteem × abuse, self-esteem × group, and abuse × group) and three-way (self-esteem × abuse × group) interaction terms to test whether associations between self-esteem and psychotic experiences were modified by prior exposure to high (mean + 1 SD of continuous CTQ scores) versus low levels (mean − 1 SD of continuous CTQ scores) [59,60] of each type of childhood abuse to each group (patients, relatives, and controls). We standardized the continuous ESM

and CTQ variables (mean 0, SD 1) to interpret significant three-way interaction terms [61]. Wald tests were used to test the hypothesis that the interaction effects were equal to zero. For the three-way interaction models, the significance levels of the Wald tests were adjusted to correct for type-I errors using family-wise error-corrected *P* value, which were computed by multiplying the unadjusted *P* value by the total number of tests. The "lincom" command was used to compute linear combinations of coefficients to test hypotheses 1 and 2. Next, we conducted time-lagged analyses to test hypotheses 3 and 4. To do so, we generated within-subject lagged variables of self-esteem and psychotic experiences (at t_{n-1} and t_n) and fitted linear mixed models.

Results

Basic Characteristics

The third-wave (6-year follow-up) data of the GROUP study was completed by 486 participants (patients: n=194, 39.9%; relatives: n=169, 34.8%; and controls: n=123, 25.3%). Participants who did not complete the CTQ in the first wave (30/486, 6.2%) and a minimum of one-third of the ESM assessments in the third wave (88/486, 18.1%) were excluded from the analysis, resulting in a sample of 368 participants (patients: n=139, 37.8%; relatives: n=118, 32.1%; and controls: n=111, 30.2%). The basic characteristics of the excluded participants of the third-wave sample were broadly similar to those of the included participants (Multimedia Appendix 1). Overall, the basic characteristics of the included patients, their relatives, and controls were similar, except that included participants of all 3 groups were more often White compared with the excluded participants. Compared with the relatives and controls, the patients were younger and had a lower educational level. The patient group consisted of more men and were more often unmarried compared with the other 2 groups (Table 1). Patients reported higher levels of all types of childhood trauma (emotional abuse, *P*<.001; physical abuse, *P*=.02; sexual abuse, *P*=.02; emotional neglect, *P*=.01; physical neglect, *P*=.01) than the controls (Table 2). Patients also reported higher levels of emotional abuse (adjusted β=.58, 95% CI 0.35-0.80; *P*<.001), physical abuse (adjusted β=.25, 95% CI 0.06-0.44; *P*=.01), and sexual abuse (adjusted β=.41, 95% CI 0.20-0.63; *P*<.001) than their first-degree relatives, where adjusted β are standardized regression coefficients. Although patients reported higher levels of emotional and physical neglect, the differences between the patients and relatives were not statistically significant (patients: adjusted β=.22, 95% CI −0.02 to 0.46; *P*=.07 and relatives: adjusted β=.08, 95% CI −0.13 to 0.29; *P*=.47; Table 3). The levels of all types of childhood trauma were similar in controls and relatives.

Table 1. Basic characteristics.

	Patients (n=139)	Relatives (n=118)	Controls (n=111)	Test statistics		*P* value
				F test (*df*)	χ^2 (*df*)	
Age (years), mean (SD)	34.5 (8.3)	35.5 (8.7)	41.0 (11.5)	15.87 (2)	N/A[a]	<.001
Gender (patients: n=135; relatives: n=118; control: n=111), n (%)[b]				N/A	34.8 (2)	<.001
Men	88 (65.2)	47 (39.8)	33 (29.7)			
Women	47 (34.8)	71 (60.2)	78 (70.3)			
Ethnicity (patients: n=139; relatives: n=118; control: n=111), n (%)				N/A	5.2 (2)	.08
White	123 (88.5)	107 (90.7)	107 (96.4)			
People of color	16 (11.5)	11 (9.3)	4 (3.6)			
Level of education (patients: n=139; relatives: n=118; control: n=111), n (%)				N/A	42.2 (4)	<.001
Low	35 (25.2)	8 (6.8)	3 (2.7)			
Middle	60 (43.2)	45 (38.1)	42 (37.8)			
High	44 (31.6)	65 (55.1)	66 (59.5)			
Marital status (patients: n=139; relatives: n=118; control: n=93), n (%)[c]				N/A	71.1 (4)	<.001
Not married	101 (72.7)	37 (31.4)	25 (26.9)			
Married or living together	28 (20.1)	76 (64.4)	62 (66.7)			
Divorced	10 (7.2)	5 (4.2)	6 (6.4)			
Current antipsychotic use (patients: n=100), n (%)[d]	98 (98)	N/A	N/A	N/A	N/A	N/A
Childhood trauma, mean (SD)						
Emotional abuse[e]	9.5 (4.2)	7.5 (3.1)	7.1 (3.4)	15.51 (2)	N/A	<.001
Physical abuse[e]	6.6 (2.9)	5.7 (2.2)	5.5 (1.6)	8.53 (2)	N/A	<.001
Sexual abuse[e]	6.3 (3.1)	5.6 (2.6)	5.9 (3.0)	1.93 (2)	N/A	.15
Emotional neglect[f]	11.4 (4.1)	10.6 (4.1)	9.2 (3.9)	9.25 (2)	N/A	<.001
Physical neglect	7.2 (2.4)	6.6 (2.5)	6.0 (1.9)	9.16 (2)	N/A	<.001

[a]N/A: not applicable.
[b]Missing values: 4/368, 0.01%.
[c]Missing values: 18/368, 4.9%.
[d]Missing values: 39/139, 28.1%.
[e]Missing values: 2/368, 0.5%.
[f]Missing values: 3/368, 0.8%.

Table 2. Categorical Childhood Trauma Questionnaire severity scores by group.

	Patients (n=139), n (%)	Relatives (n=118), n (%)	Controls (n=111), n (%)	χ^2 (*df*)	*P* value
Emotional abuse[a] (patients: n=139; relatives: n=118; control: n=109)				32.8 (6)	<.001
Severe to extreme	13 (9.3)	4 (3.4)	3 (2.8)		
Moderate to severe	19 (13.7)	2 (1.7)	6 (5.5)		
Mild to moderate	34 (24.5)	22 (18.6)	13 (11.9)		
None	73 (52.5)	90 (76.3)	87 (79.8)		
Physical abuse[a] (patients: n=138; relatives: n=118; control: n=110)				15.5 (6)	.02
Severe to extreme	8 (5.8)	3 (2.5)	3 (2.7)		
Moderate to severe	11 (8)	2 (1.7)	1 (0.9)		
Mild to moderate	7 (5.1)	5 (4.3)	2 (1.8)		
None	112 (81.1)	108 (91.5)	104 (94.6)		
Sexual abuse[a] (patients: n=139; relatives: n=118; control: n=109)				15.5 (6)	.02
Severe to extreme	9 (6.5)	4 (3.4)	5 (4.6)		
Moderate to severe	14 (10.0)	2 (1.7)	6 (5.5)		
Mild to moderate	19 (13.7)	11 (9.3)	6 (5.5)		
None	97 (69.8)	101 (85.6)	92 (84.4)		
Emotional neglect[b] (patients: n=138; relatives: n=118; control: n=109)				17.6 (6)	.01
Severe to extreme	10 (7.2)	9 (7.6)	6 (5.5)		
Moderate to severe	20 (14.5)	11 (9.3)	4 (3.7)		
Mild to moderate	60 (43.5)	44 (37.3)	35 (32.1)		
None	48 (34.8)	54 (45.8)	64 (58.7)		
Physical neglect (patients: n=139; relatives: n=118; control: n=111)				15.9 (6)	.01
Severe to extreme	6 (4.3)	6 (5.1)	2 (1.8)		
Moderate to severe	18 (13.0)	6 (5.1)	4 (3.6)		
Mild to moderate	21 (15.1)	13 (11.0)	9 (8.1)		
None	94 (67.6)	93 (78.8)	96 (86.5)		

[a]Missing values: 2/368, 0.5%.

[b]Missing values: 3/268, 0.8%.

Table 3. Differences in scores of Childhood Trauma Questionnaire severity between groups.

	Patients versus controls		Relatives versus controls		Patients versus relatives	
	Adjusted β (95% CI)[a]	*P* value	Adjusted β (95% CI)	*P* value	Adjusted β (95% CI)	*P* value
Emotional abuse	.67 (0.42 to 0.92)	<.001	.09 (−0.14 to 0.32)	.44	.58 (0.35 to 0.80)	<.001
Physical abuse	.30 (0.09 to 0.51)	.005	.05 (−0.14 to 0.25)	.58	.25 (0.06 to 0.44)	.01
Sexual abuse	.42 (0.19 to 0.66)	<.001	.01 (−0.21 to 0.23)	.92	.41 (0.20 to 0.63)	<.001
Emotional neglect	.51 (0.25 to 0.78)	<.001	.30 (0.05 to 0.54)	.02	.22 (−0.02 to 0.46)	.07
Physical neglect	.26 (0.03 to 0.50)	.03	.19 (−0.02 to 0.40)	.08	.08 (−0.13 to 0.29)	.47

[a]CI adjusted for age, gender, ethnicity, education, and marital status.

Cross-sectional Association Between Self-esteem and Psychotic Experiences by Childhood Trauma and Group

As shown in Table 4, after controlling for age, gender, ethnicity, education, and marital status, we found strong evidence that the association between self-esteem and psychotic experiences was modified by exposure to childhood trauma, as indicated by the statistically significant interaction effect of self-esteem × CTQ total score × group (χ^2_2=9.0, *P*=.01). We found statistically significant interaction effects (all *P*≤.001) of self-esteem × physical abuse × group, self-esteem × sexual abuse × group, and self-esteem × physical neglect × group. However, there was no evidence of interaction effects of self-esteem × emotional abuse × group (χ^2_2=5.4, *P*=.07) and self-esteem × emotional neglect × group (χ^2_2=1.8, *P*=.41).

Table 4. Associations between momentary self-esteem and psychotic experiences by childhood trauma in patients, relatives, and controls[a].

	Outcome: psychotic experiences							
	Patients		Relatives		Controls		Wald test for interaction	
	Adjusted β[b] (95% CI)	*P* value	Adjusted β (95% CI)	*P* value	Adjusted β (95% CI)	*P* value	χ^2 (*df*)	*P* value[c]
Momentary SE[d] × childhood trauma × group[e]							9.0 (2)	.01
Level of childhood trauma								
High (mean + 1 SD)	−.20 (−0.25 to −0.15)	<.001	−.29 (−0.32 to −0.27)	<.001	−.15 (−0.19 to −0.12)	<.001		
Average (mean)	−.17 (−0.20 to −0.14)	<.001	−.27 (−0.29 to −0.25)	<.001	−.18 (−0.20 to −0.15)	<.001		
Low (mean − 1 SD)	−.14 (−0.18 to −0.10)	<.001	−.25 (−0.28 to −0.22)	<.001	−.20 (−0.24 to −0.16)	<.001		
High versus low	−.06 (−0.13 to 0.01)	.12	−.04 (−0.08 to −0.01)	.03	.05 (−0.00 to 0.10)	.07		
Momentary SE × emotional abuse × group[e]							5.4 (2)	.07
Momentary SE × physical abuse × group[e]							24.9 (2)	<.001
Level of physical abuse								
High (mean + 1 SD)	−.16 (−0.22 to −0.10)	<.001	−.32 (−0.35 to −0.30)	<.001	−.08 (−0.12 to −0.04)	<.001		
Average (mean)	−.16 (−0.18 to −0.13)	<.001	−.26 (−0.29 to −0.24)	<.001	−.11 (−0.14 to −0.08)	<.001		
Low (mean − 1 SD)	−.16 (−0.21 to −0.11)	<.001	−.21 (−0.24 to −0.17)	<.001	−.14 (−0.18 to −0.09)	<.001		
High versus low	0 (−0.09 to 0.09)	.99	−.12 (−0.15 to −0.08)	<.001	.06 (−0.00 to 0.12)	.058		
Momentary SE × sexual abuse × group[e]							15.9 (2)	<.001
Level of sexual abuse								
High (mean + 1 SD)	−.11 (−0.17 to −0.06)	<.001	−.29 (−0.32 to −0.27)	<.001	−.08 (−0.12 to −0.04)	<.001		
Average (mean)	−.15 (−0.18 to −0.13)	<.001	−.27 (−0.29 to −0.25)	<.001	−.11 (−0.14 to −0.08)	<.001		
Low (mean − 1 SD)	−.19 (−0.24 to −0.15)	<.001	−.25 (−0.28 to −0.22)	<.001	−.13 (−0.17 to −0.09)	<.001		
High versus low	.08 (−0.01 to 0.16)	.07	−.05 (−0.08 to −0.02)	.001	.05 (0.00 to 0.10)	.049		
Momentary SE × emotional neglect × group							1.8 (2)	.41
Momentary SE × physical neglect × group[e]							116.7 (2)	<.001
Level of physical neglect								
High (mean + 1 SD)	−.23 (−0.28 to −0.19)	<.001	−.29 (−0.32 to −0.27)	<.001	−.08 (−0.12 to −0.04)	<.001		
Average (mean)	−.17 (−0.20 to −0.14)	<.001	−.28 (−0.30 to −0.25)	<.001	−.11 (−0.14 to −0.08)	<.001		
Low (mean − 1 SD)	−.11 (−0.14 to −0.07)	<.001	−.26 (−0.29 to −0.23)	<.001	−.13 (−0.17 to −0.09)	<.001		
High versus low	−.13 (−0.19 to −0.07)	<.001	−.03 (−0.07 to 0.01)	.14	.05 (−0.01 to 0.11)	.11		

[a]Adjusted for age, gender, ethnicity, level of education, and marital status.

[b]Adjusted β: standardized regression coefficients (continuous independent variables were standardized [mean 0, SD 1] for interpreting significant three-way interaction terms and examining the difference in associations between high [mean + 1 SD], average [mean], and low [mean − 1 SD] levels of childhood trauma within and across groups [patients, relatives, and controls]).

[c]Family-wise error-corrected *P* values were computed by multiplying the unadjusted *P* value by the total number of tests to adjust the significance levels of Wald tests for three-way interactions.

[d]SE: self-esteem.

[e]Three-way interaction included in the following model (with y_{ij} psychotic experiences as the outcome variable): $y_{ij} = \beta_0 + \beta_1(\text{self-esteem}) + \beta_2(\text{childhood trauma}_j) + \beta_3(\text{group}_j) + \beta_4(\text{self-esteem}_{ij} \times \text{childhood trauma}_j) + \beta_5(\text{self-esteem}_{ij} \times \text{group}_j) + \beta_6(\text{childhood trauma}_j \times \text{group}_j) + \beta_7(\text{self-esteem}_{ij} \times \text{childhood trauma}_j \times \text{group}_j) + \varepsilon_{ij}$ (full model not shown and available upon request).

Within-Group Comparison—Hypothesis 1

Lower levels of momentary self-esteem were associated with more intense psychotic experiences in relatives exposed to high levels of childhood trauma in general compared with those exposed to low levels of childhood trauma (adjusted $\beta_{high\ vs\ low}$=−.04; *P*=.03). There was no evidence that this association was stronger in patients or controls exposed to high versus low levels of childhood trauma in general. Specifically, in relatives, we found a stronger association between lower self-esteem and more intense psychotic experiences for those exposed to high versus low levels of physical abuse (adjusted $\beta_{high\ vs\ low}$=−.12; *P*<.001); however, we did not find evidence that this association was modified in patients or controls (Figure S1 in Multimedia Appendix 2). Furthermore, lower self-esteem was associated with more intense psychotic experiences in relatives and controls exposed to high versus low levels of sexual abuse (relatives: adjusted $\beta_{high\ vs\ low}$=−.05; *P*<.001 and controls: adjusted $\beta_{high\ vs\ low}$=.05; *P*=.049) but not in patients (Figure S2 in Multimedia Appendix 2). Finally, in patients, we found a stronger association between lower self-esteem and more intense psychotic experiences in those exposed to high versus low levels of physical neglect (adjusted $\beta_{high\ vs\ low}$=−.13; *P*<.001; Figure S3 in Multimedia Appendix 2). However, there was no evidence that this association was modified in relatives and controls.

Between-Group Comparison—Hypothesis 2

Next, we examined the differences in the magnitude of associations of self-esteem and psychotic experiences between those exposed to high versus low levels of childhood trauma in general, physical and sexual abuse, and physical neglect across groups. These differences in magnitude were only examined if both groups showed significant (*P*<.05) within-group associations (Table 5). When comparing relatives with controls, we found differences in the magnitude of associations between self-esteem and psychotic experiences between those exposed to high versus low levels of sexual abuse (adjusted $\beta_{high\ vs\ low}$=−.10; *P*<.001), with differences in magnitude of association being greater in relatives.

Table 5. Difference in associations in those exposed to high versus low levels of childhood trauma across groups (Δ high vs low).

	Outcome: psychotic experiences					
	Patients versus controls		Relatives versus controls		Patients versus relatives	
	Adjusted β (95% CI)	*P* value	Adjusted β (95% CI)	*P* value	Adjusted β (95% CI)	*P* value
Self-esteem						
Δ High versus low childhood trauma across groups	N/A[a]	N/A	N/A	N/A	N/A	N/A
Δ High versus low physical abuse across groups	N/A	N/A	N/A	N/A	N/A	N/A
Δ High versus low sexual abuse across groups	N/A	N/A	−.10 (−0.16 to −0.04)	<.001	N/A	N/A
Δ High versus low physical neglect across groups	N/A	N/A	N/A	N/A	N/A	N/A

[a]N/A: not applicable.

Temporal Association Between Self-esteem and Psychotic Experiences by Childhood Trauma and Group

Findings regarding the temporal associations between momentary self-esteem and psychotic experiences in patients, relatives, and controls are shown in Tables 6 and 7. After controlling for age, gender, ethnicity, education, and marital status, we found no evidence that associations between self-esteem at t_{n-1} and psychotic experiences at t_n or between psychotic experiences at t_{n-1} and self-esteem at t_n were modified by exposure to any of the types of childhood trauma (ie, emotional, physical, and sexual abuse and emotional and physical neglect) within (hypothesis 3) and across (hypothesis 4) groups (ie, patients, relatives, and controls).

Table 6. Effect of self-esteem and psychotic symptoms at t_{n-1} on psychotic experiences at t_n by group and modified by childhood trauma[a].

	Patients		Relatives		Controls		Wald test for interaction	
	Adjusted β[b] (95% CI)	*P* value	Adjusted β (95% CI)	*P* value	Adjusted β (95% CI)	*P* value	χ^2 (*df*)	*P* value[c]
Self-esteem t_{n-1}								
Emotional abuse							0.07 (2)	.80
High (mean + 1 SD)	−.04 (−0.10 to 0.03)	.28	−.11 (−0.14 to −0.08)	<.001	−.02 (−0.07 to 0.03)	.48		
Average (mean)	−.03 (−0.07 to −0.00)	.05	−.13 (−0.16 to −0.11)	<.001	−.02 (−0.06 to 0.01)	.17		
Low (mean − 1 SD)	−.03 (−0.08 to 0.02)	.20	−.15 (−0.19 to −0.11)	<.001	−.03 (−0.07 to 0.02)	.20		
High versus low	−.00 (−0.10 to 0.09)	.92	.04 (−0.01 to 0.08)	.10	.01 (−0.06 to 0.08)	.77		
Physical abuse							0.58 (2)	.45
High (mean + 1 SD)	.01 (−0.07 to 0.09)	.82	−.14 (−0.17 to 0.11)	<.001	−.02 (−0.06 to 0.03)	.49		
Average (mean)	−.03 (−0.06 to 0.01)	.14	−.12 (−0.14 to −0.09)	<.001	−.02 (−0.06 to 0.01)	.15		
Low (mean − 1 SD)	−.06 (−0.11 to −0.00)	.03	−.10 (−0.13 to −0.06)	<.001	−.03 (−0.08 to 0.01)	.17		
High versus low	.07 (−0.05 to 0.18)	.25	−.04 (−0.09 to 0.00)	.06	.02 (−0.05 to −0.08)	.60		
Sexual abuse							0.01 (2)	.91
High (mean + 1 SD)	−.03 (−0.09 to 0.03)	.32	−.12 (−0.14 to −0.09)	<.001	−.02 (−0.06 to 0.02)	.34		
Average (mean)	−.03 (−0.06 to −0.00)	.046	−.13 (−0.15 to −0.10)	<.001	−.02 (−0.06 to 0.01)	.14		
Low (mean − 1 SD)	−.03 (−0.08 to 0.01)	.15	−.14 (−0.17 to −0.10)	<.001	−.03 (−0.07 to 0.01)	.19		
High versus low	0 (−0.08 to 0.09)	.93	.02 (−0.01 to 0.05)	.27	.01 (−0.04 to 0.06)	.72		
Emotional neglect							0.04 (2)	.83
High (mean + 1 SD)	−.02 (−0.07 to 0.02)	.29	−.11 (−0.14 to −0.08)	<.001	−.02 (−0.07 to 0.03)	.40		
Average (mean)	−.03 (−0.06 to −0.00)	.04	−.13 (−0.15 to −0.10)	<.001	−.02 (−0.06 to 0.01)	.15		
Low (mean − 1 SD)	−.04 (−0.08 to 0.00)	.07	−.15 (−0.19 to −0.11)	<.001	−.03 (−0.07 to 0.02)	.22		
High versus low	.01 (−0.05 to 0.07)	.64	.04 (−0.0 to 0.09)	.11	0 (−0.06 to 0.07)	.89		
Physical neglect							0.18 (2)	.67
High (mean + 1 SD)	−.02 (−0.07 to 0.03)	.41	−.11 (−0.14 to.08)	<.001	−.03 (−0.08 to 0.02)	.29		
Average (mean)	−.03 (−0.06 to 0.00)	.05	−.13 (−0.15 to −0.10)	<.001	−.03 (−0.06 to 0.01)	.13		
Low (mean − 1 SD)	−.04 (−0.08 to 0.00)	.07	−.14 (−0.18 to −0.10)	<.001	−.02 (−0.07 to −0.02)	.32		
High versus low	.02 (−0.05 to 0.09)	.60	.03 (−0.02 to 0.08)	.28	−.00 (−0.08 to 0.07)	.94		

[a]Adjusted for age, sex, ethnicity, level of education, and marital status.

[b]Adjusted β, standardized regression coefficients (continuous independent variables were standardized [mean 0, SD 1] for interpreting significant three-way interaction terms and examining the difference in associations between high [mean + 1 SD], average [mean], and low [mean − 1 SD] levels of childhood trauma within and across groups [patients, relatives, and controls]).

[c]Family-wise error-corrected *P* values were computed by multiplying the unadjusted *P* value by the total number of tests to adjust significance levels of Wald tests for three-way interactions.

Table 7. Effect of self-esteem and psychotic symptoms at t_{n-1} on self-esteem at t_n by group and modified by childhood trauma[a].

	Patients		Relatives		Controls		Wald test for interaction	
	Adjusted β[b] (95% CI)	*P* value	Adjusted β (95% CI)	*P* value	Adjusted β (95% CI)	*P* value	χ^2 (*df*)	*P* value[c]
Psychotic Experiences t_{n-1}								
Emotional abuse							0.52 (2)	.47
High (mean + 1 SD)	−.09 (−0.22 to 0.05)	.20	−.09 (−0.12 to −0.06)	<.001	−.13 (−0.27 to 0.00)	.06		
Average (mean)	−.09 (−0.14 to −0.04)	.001	−.11 (−0.14 to −0.08)	<.001	−.08 (−0.15 to −0.02)	.02		
Low (mean − 1 SD)	−.09 (−0.19 to −0.00)	.046	−.13 (−0.18 to −0.09)	<.001	−.03 (−0.13 to −0.07)	.51		
High versus low	.01 (−0.20 to 0.21)	.96	.04 (−0.01 to 0.10)	.09	−.10 (−0.29 to −0.10)	.33		
Physical abuse							0.04 (2)	.84
High (mean + 1 SD)	−.07 (−0.21 to 0.07)	.32	−.10 (−0.13 to −0.07)	<.001	−.04 (−0.15 to 0.07)	.46		
Average (mean)	−.09 (−0.14 to −0.03)	.002	−.10 (−0.13 to −0.08)	<.001	−.07 (−0.13 to −0.01)	.03		
Low (mean − 1 SD)	−.10 (−0.20 to −0.01)	.03	−.11 (−0.15 to −0.07)	<.001	−.10 (−0.20 to −0.00)	.04		
High versus low	.03 (−0.17 to 0.24)	.75	−.01 (−0.03 to 0.06)	.48	.06 (−0.11 to 0.22)	.48		
Sexual abuse							0.02 (2)	.88
High (mean + 1 SD)	−.07 (−0.19 to 0.05)	.24	−.10 (−0.13 to −0.07)	<.001	−.06 (−0.15 to −0.02)	.15		
Average (mean)	−.09 (−0.14 to −0.04)	<.001	−.10 (−0.13 to −0.07)	<.001	−.07 (−0.13 to −0.01)	.03		
Low (mean − 1 SD)	−.11 (−0.20 to −0.01)	.03	−.10 (−0.14 to −0.07)	<.001	−.08 (−0.16 to 0.00)	.055		
High versus low	.03 (−0.16 to 0.23)	.72	.00 (−0.04 to 0.04)	.85	.02 (−0.09 to 0.13)	.75		
Emotional neglect							0.00 (2)	.99
High (mean + 1 SD)	−.10 (−0.17 to −0.02)	.01	−.10 (−0.14 to −0.06)	<.001	−.08 (−0.19 to 0.04)	.18		
Average (mean)	−.09 (−0.14 to −0.04)	<.001	−.10 (−0.13 to −0.08)	<.001	−.07 (−0.14 to −0.01)	.03		
Low (mean − 1 SD)	−.08 (−0.15 to −0.02)	.01	−.11 (−0.15 to −0.06)	<.001	−.07 (−0.16 to −0.03)	.18		
High versus low	−.02 (−0.12 to 0.08)	.76	.01 (−0.05 to 0.07)	.83	−.01 (−0.18 to 0.15)	.87		
Physical neglect							0.48 (2)	.49
High (mean + 1 SD)	−.09 (−0.16 to −0.01)	.02	−.10 (−0.13 to −0.07)	<.001	−.00 (−0.19 to 0.18)	.96		
Average (mean)	−.09 (−0.14 to −0.04)	<.001	−.11 (−0.14 to −0.08)	<.001	−.06 (−0.13 to 0.02)	.12		
Low (mean − 1 SD)	−.09 (−0.16 to −0.03)	.006	−.12 (−0.16 to −0.07)	<.001	−.11 (−0.22 to 0.00)	.05		
High versus low	.01 (−0.09 to 0.11)	.89	.02 (−0.03 to 0.07)	.41	−.11 (−0.16 to 0.37)	.43		

[a]Adjusted for age, sex, ethnicity, level of education, and marital status.

[b]Adjusted β, standardized regression coefficients (continuous independent variables were standardized [mean 0, SD 1] for interpreting significant three-way interaction terms and examining the difference in associations between high [mean + 1 SD], average [mean], and low [mean−1 SD] levels of childhood trauma within and across groups [patients, relatives, and controls]).

[c]Family-wise error-corrected *P* values were computed by multiplying the unadjusted *P* value by the total number of tests to adjust significance levels of Wald tests for three-way interactions.

Discussion

Principal Findings

Using an experience sampling design, the results of this study showed strong evidence that associations between momentary low self-esteem and increased intensity of psychotic experiences in daily life were modified by several types of childhood trauma, that is, physical and sexual abuse and physical neglect. For physical abuse, this was only the case for relatives of patients with psychotic disorders, and for physical neglect, this was only the case for patients. Sexual abuse modified the association in relatives and controls, but not in patients. Emotional abuse and

emotional neglect did not modify the association between self-esteem and psychotic experiences in any group. When investigating temporal order, we found no evidence that childhood trauma modified the temporal associations between self-esteem at t_{n-1} and psychotic experiences at t_n or those between psychotic experiences at t_{n-1} and self-esteem at t_n.

Methodological Considerations

Several limitations of this study should be considered when interpreting its findings. First, only a selection of the baseline sample made it through the third wave. For our analyses, we had to exclude 118 individuals because they did not complete either a sufficient number of ESM assessments or the CTQ. Possibly, these assessments were too burdensome and, therefore, might have led to selection bias. However, when comparing the excluded individuals with those in the analytical sample, who participated in the third wave of GROUP assessments, in terms of basic sample characteristics, the included and excluded participants were comparable to a great extent, except for ethnicity (included participants were more often White in all 3 groups). Second, the CTQ, which was used to measure exposure to several types of trauma during childhood, is a retrospective self-report measure. It has been argued that the CTQ is prone to recall bias and that the manifestation of psychotic symptoms might affect the ratings of this measure [62,63]. However, Gayer-Anderson et al [64] demonstrated the accuracy, strength of agreement, and convergent validity to be broadly similar between patients with first-episode psychosis and controls. Notably, although measured differently, in our control group, the prevalence of several childhood trauma types was higher than that in the general population in the Netherlands [65]. However, if we only look at the categories “severe to extreme” and “moderate to severe” in our sample, the percentages are quite similar, except for sexual abuse. Possibly related to this, we found that lower levels of self-esteem were associated with more intense psychotic experiences in controls exposed to high versus low levels of sexual abuse; this was not the case for any of the other types of childhood trauma (in controls). This might indicate that higher levels of sexual abuse in this group might have affected the outcome. In addition, all ESM assessments of self-esteem and psychotic experiences were based on subjective self-reports, which might have led to bias. However, the ESM has been found to be a feasible, reliable, and valid assessment method in various populations [45,51,66], including patients with psychosis [51,52,58].

Moreover, momentary self-esteem was measured using only 2 items. Using fewer items in ESM research is quite common because it minimizes reactivity to the assessment method [67]. The construct of self-esteem involves both positive and negative self-esteem [68,69]. Therefore, we used 1 item that measured the positive dimension and 1 that measured the negative dimension of self-esteem, which is in line with previous ESM studies investigating self-esteem [55,70]. Because of this heterogeneity of the self-esteem construct and, hence, ESM items, the internal consistency of the 2 momentary self-esteem items that we used was as expected (within-person α=.22; between-person α=.76) [71]. In previous analyses, we investigated the extent to which the 2 ESM items adequately tapped into the self-esteem construct in a subsample of the current sample. We found that the convergent validity between the 2 momentary self-esteem items and the Rosenberg Self-Esteem Scale [69] was fair, with the intraclass correlation coefficient being 0.41 [33]. It would have been interesting, however, to triangulate momentary ESM measures with explicit (Rosenberg Self-Esteem Scale) and implicit measures [72] of self-esteem to corroborate our findings. In addition, for future research, we would recommend using multiple items to measure both the positive and negative dimensions of momentary self-esteem.

Next, we controlled for potential a priori confounders, such as age, gender, ethnicity, education level, and marital status. We did not include medication use as a confounder because 98% (98/100) of the patients were using medication. Nevertheless, unmeasured confounders, such as other childhood adversities, comorbidities, the impact of illness chronicity, and molecular genetic measures, were not considered and may have, therefore, influenced the findings.

Finally, we standardized the continuous childhood trauma variables because it allowed us to interpret associations at higher versus lower levels of childhood trauma [60,61]. Notably, this implies that by using continuous and standardized variables for childhood trauma, no discrete distinction was made between those exposed and not exposed, rather between those exposed to higher and lower levels of each type of childhood trauma. Therefore, these results could possibly be an underestimation of the reality.

Comparison With Previous Research

There is well-established evidence that exposure to childhood trauma is a risk factor for developing a psychotic disorder [3-8,10-12], and self-esteem has been proposed to be involved in the pathway from childhood trauma to psychosis [38-42]. In addition, mediation models linking childhood trauma and self-esteem in pathways to psychosis have been previously proposed and tested [41,73]. However, evidence regarding whether prior exposure to childhood trauma (emotional, physical, and sexual abuse and emotional and physical neglect) moderates the association between momentary self-esteem and psychotic experiences in daily life remains limited.

As hypothesized (hypothesis 1), and extending previous findings [33], we found evidence that exposure to physical and sexual abuse and physical neglect modified the association between momentary self-esteem and psychotic experiences in daily life. However, the results also suggested that exposure to emotional abuse and emotional neglect did not modify the association between momentary self-esteem and psychotic experiences in daily life. This is unexpected, as previous studies demonstrated that when comparing the 5 types of childhood trauma, emotional abuse and neglect are the types that affect self-esteem the most [74,75]. Therefore, we would expect that especially these types of trauma would have a particular bearing on the association between self-esteem and psychosis. However, the timing, chronicity, and severity of emotional abuse and neglect are related to the extent to which maltreatment affects developmental trajectories [76]. For example, when maltreatment is less severe or ends early in childhood, it is

possible that its impact might fade with time in some individuals [77,78]. Indeed, our data show that the vast majority of the participants experienced none or mild to moderate levels of emotional abuse and neglect. Another explanation for these unexpected results might be that cross-sectional modeling of associations between momentary self-esteem and psychotic experiences does not consider moment-to-moment variation in these measures. However, when we evaluated these associations using time lags, we observed that exposure to childhood trauma did not modify the temporal association between self-esteem and psychotic experiences and vice versa.

On the basis of previous literature [38-42], we hypothesized that the effect of exposure to childhood trauma would modify the association between low self-esteem and increased intensity of psychotic experiences in daily life most substantially in patients, followed by their relatives. We found that physical neglect modified the association between self-esteem and psychotic experiences in patients; physical abuse modified this association only in relatives; and sexual abuse modified this association in relatives and controls, with the impact being stronger in relatives than in controls. It is possible that the controls exposed to childhood trauma may have better coping strategies [79] compared with first-degree relatives of patients with psychotic disorders who have a familial liability to psychosis. Therefore, controls might be more resilient to lower levels of self-esteem and psychotic experiences in daily life [79]. In line with this, there is also evidence that controls exposed to high levels of childhood trauma are more resilient to daily life stress, compared with patients with first-episode psychosis, individuals with at-risk mental state, and help-seeking service users [22,23]. Moreover, previous literature showed that resilience at baseline was lower in individuals who developed psychosis at follow-up than in those who did not [80], suggesting that resilience is a protective factor in the formation of psychosis [81].

In addition, relatives have a familial liability to psychosis [16,18], and they share a degree of exposure to childhood trauma with their siblings with psychotic disorders [20], which may explain the stronger associations in relatives than in controls. Patients were recruited via treating clinicians, which implies that they all received a form of (standard) mental health care, including medication. Treatments such as psychoeducation, cognitive behavioral therapy, and antipsychotic medication have been shown to reduce symptoms and prevent relapse [82-84]. It has been demonstrated that the therapies target distorted beliefs about delusions and hallucinations, thereby decreasing the negative consequences of psychotic symptoms [83-85]. Moreover, the results of a meta-analysis showed that the interpretation of these beliefs is addressed during treatment by considering psychological mechanisms and processes that might contribute to the formation and maintenance of psychotic symptoms, such as emotions, arousal, attachment, interpersonal issues, trauma, and self-esteem [86]. For example, the results of a randomized clinical trial showed that treatment, such as eye movement desensitization and reprocessing and prolonged exposure, in patients with psychotic disorders reduced trauma symptoms and psychotic symptoms [87], also at 6-month follow-up. This potentially implies that the influence of trauma on patients might become less impactful over time owing to the effects of the treatment they had received. Furthermore, 98% (98/100) of patients were using antipsychotic medication. Emotional flattening might be a negative symptom of schizophrenia, but it is also a side effect of medication. Emotional flattening interferes with expressiveness, and this often leads to problems in interpersonal interactions, which in turn leads to more withdrawal from (social) activities [88]. Selective perception and selective memory are important features of low self-esteem, and as a consequence, negative convictions about one's self are repeatedly confirmed in new situations [89]. However, this will not occur often if patients who are using antipsychotics socially withdraw and therefore experience fewer social interactions. Potentially, directly or indirectly, medication use and the effects of illness chronicity might have a flattening effect on self-esteem. Therapy and medication use might limit the impact of exposure to traumatic experiences on the association between self-esteem and psychotic experiences in daily life for patients with enduring psychotic disorders.

Overall, the cross-sectional findings indicate that it might possibly be relevant to develop and implement early and low-level interventions that directly target self-esteem, especially for individuals exposed to physical and sexual abuse and physical neglect. Thus, the intensity of psychotic experiences in daily life can be reduced. We are currently evaluating the ecological momentary intervention (EMI) "SELFIE" [90]. The next step would be to implement these types of EMIs in routine mental health care.

Previous research has investigated the temporal order of self-esteem and psychotic experiences in daily life and found that self-esteem preceded psychotic symptoms only in controls and psychotic experiences had a temporal effect on self-esteem in patients with psychotic disorders, their first-degree relatives, and controls [33]. In this study, we investigated whether prior exposure to childhood trauma modified the temporal associations between self-esteem at t_{n-1} and psychotic experiences at t_n and between psychotic experiences at t_{n-1} and self-esteem at t_n. However, the results showed no evidence of these associations. Nevertheless, we would not rule out the hypothesis of a temporal order entirely because, in this study, we investigated the interaction between prior exposure to childhood trauma, psychotic experiences at t_n, and momentary self-esteem at the previous time point (t_{n-1}) and vice versa. Assessments took place within 90-minute time blocks. It is possible that longer time lags would have yielded different results, as the occurrence of psychosis might be preceded by weeks, months, or even years of psychological and behavioral abnormalities [91]. Perhaps self-esteem is a process that needs more time to unfold to be succeeded by psychotic experiences.

Conclusions

Taken together, our findings suggest that the association between self-esteem and psychotic experiences in daily life is stronger in those exposed to high versus low levels of several types of childhood trauma. Hence, self-esteem might be a psychological process that links childhood trauma to psychosis. The results showed that this is especially the case in first-degree relatives

and, to a lesser extent, in patients with psychotic disorders. Although we did not find evidence of temporal associations, the cross-sectional results indicate that improving self-esteem may potentially reduce the intensity of psychotic experiences in daily life. This underlines the importance of developing and evaluating early and low-level EMIs that directly target self-esteem to reduce the intensity of psychotic experiences in daily life. We currently evaluate such an intervention in the SELFIE study [90].

Acknowledgments

The authors are grateful for the generosity of the time and effort of the patients, their families, and control participants. The authors would also like to thank all research personnel involved in the Genetic Risk and Outcome in Psychosis (GROUP) project, in particular Joyce van Baaren, Erwin Veermans, Ger Driessen, Truda Driesen, Karin Pos, Erna van 't Hag, Jessica de Nijs, and Atiqul Islam.

This work was supported by a research grant from the Netherlands Organization for Health Research and Development (ZonMw; grant 60-63600-98-346) and a Heisenberg Professorship from the German Research Foundation (grant 389624707) to Ulrich Reininghaus (UR). The infrastructure for the GROUP study is funded through the Geestkracht program of the Dutch Health Research Council (Zon-Mw; grant 10-000-1001) and matching funds from participating pharmaceutical companies (Lundbeck, Astra Zeneca, Eli Lilly, and Janssen-Cilag), universities, and mental health care organizations (Amsterdam: Academic Psychiatric Centre of the Academic Medical Center and the mental health institutions GGZ Ingeest, Arkin, Dijk en Duin, GGZ Rivierduinen, Erasmus Medical Centre, and GGZ Noord Holland Noord. Groningen: University Medical Center Groningen and the mental health institutions Lentis; GGZ Friesland; GGZ Drenthe; Dimence; Mediant; GGNet Warnsveld; Yulius Dordrecht; and Parnassia psycho-medical center, The Hague. Maastricht: Maastricht University Medical Centre and the mental health institutions GGzE, GGZ Breburg, GGZ Oost-Brabant, Vincent van Gogh voor Geestelijke Gezondheid, Mondriaan, Virenze riagg, Zuyderland GGZ, MET ggz, Universitair Centrum Sint-Jozef Kortenberg, CAPRI University of Antwerp, PC Ziekeren Sint-Truiden, PZ Sancta Maria Sint-Truiden, GGZ Overpelt, and OPZ Rekem. Utrecht: University Medical Center Utrecht and the mental health institutions Altrecht, GGZ Centraal, and Delta).

The members of the *GROUP investigators* are as follows:

Therese van Amelsvoort[a], Agna A Bartels-Velthuis[b], Richard Bruggeman[b,c], Wiepke Cahn[d,e], Lieuwe de Haan[f,g], Frederike Schirmbeck[f,g], Claudia J P Simons[a,h], Jim van Os[d,i]

[a]Department of Psychiatry and Neuropsychology, Maastricht University Medical Center, School for Mental Health and Neuroscience, Maastricht, The Netherlands

[b]University of Groningen, University Medical Center Groningen, University Center for Psychiatry, Rob Giel Research center, Groningen, The Netherlands

[c]Department of Clinical and Developmental Neuropsychology, University of Groningen, Groningen, The Netherlands

[d]Department of Psychiatry, University Medical Center Utrecht, Brain Centre Rudolf Magnus, Utrecht University, Utrecht, The Netherlands

[e]Altrecht, General Mental Health Care, Utrecht, The Netherlands

[f]Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

[g]Arkin Institute for Mental Health, Amsterdam, The Netherlands

[h]GGzE Institute for Mental Health Care, Eindhoven, The Netherlands

[i]Department of Psychosis Studies, King's College London, King's Health Partners, Institute of Psychiatry, London, United Kingdom

Conflicts of Interest

None declared.

Multimedia Appendix 1
Basic characteristics of the excluded participants (participants with less than one-third experience sampling method assessments or no Childhood Trauma Questionnaire results were excluded).
[DOCX File , 14 KB - mental_v10i1e34147_app1.docx]

Multimedia Appendix 2
Associations between self-esteem psychotic experiences at high (mean + 1 SD) versus low (mean − 1 SD) levels of physical abuse, sexual abuse, and physical neglect in patients, relatives, and controls.
[DOCX File , 138 KB - mental_v10i1e34147_app2.docx]

References

1. Duhig M, Patterson S, Connell M, Foley S, Capra C, Dark F, et al. The prevalence and correlates of childhood trauma in patients with early psychosis. Aust N Z J Psychiatry 2015 Jul 26;49(7):651-659. [doi: 10.1177/0004867415575379] [Medline: 25722463]
2. Bernstein DP, Stein JA, Newcomb MD, Walker E, Pogge D, Ahluvalia T, et al. Development and validation of a brief screening version of the childhood trauma questionnaire. Child Abuse Negl 2003 Feb;27(2):169-190. [doi: 10.1016/s0145-2134(02)00541-0] [Medline: 12615092]
3. Varese F, Smeets F, Drukker M, Lieverse R, Lataster T, Viechtbauer W, et al. Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies. Schizophr Bull 2012 Jun;38(4):661-671 [FREE Full text] [doi: 10.1093/schbul/sbs050] [Medline: 22461484]
4. B Folk J, Tully L, Blacker D, Liles B, Bolden K, Tryon V, et al. Uncharted waters: treating trauma symptoms in the context of early psychosis. J Clin Med 2019 Sep 12;8(9):1456 [FREE Full text] [doi: 10.3390/jcm8091456] [Medline: 31547447]
5. Stanton KJ, Denietolis B, Goodwin BJ, Dvir Y. Childhood trauma and psychosis: an updated review. Child Adolesc Psychiatr Clin N Am 2020 Jan;29(1):115-129. [doi: 10.1016/j.chc.2019.08.004] [Medline: 31708041]
6. Frissen A, Lieverse R, Marcelis M, Drukker M, Delespaul P, GROUP Investigators. Psychotic disorder and educational achievement: a family-based analysis. Soc Psychiatry Psychiatr Epidemiol 2015 Oct 23;50(10):1511-1518 [FREE Full text] [doi: 10.1007/s00127-015-1082-6] [Medline: 26099998]
7. Kraan T, van Dam DS, Velthorst E, de Ruigh EL, Nieman DH, Durston S, et al. Childhood trauma and clinical outcome in patients at ultra-high risk of transition to psychosis. Schizophr Res 2015 Dec;169(1-3):193-198. [doi: 10.1016/j.schres.2015.10.030] [Medline: 26585219]
8. Kraan TC, Ising HK, Fokkema M, Velthorst E, van den Berg DP, Kerkhoven M, et al. The effect of childhood adversity on 4-year outcome in individuals at ultra high risk for psychosis in the Dutch Early Detection Intervention Evaluation (EDIE-NL) Trial. Psychiatry Res 2017 Jan;247:55-62. [doi: 10.1016/j.psychres.2016.11.014] [Medline: 27863320]
9. Norman RE, Byambaa M, De R, Butchart A, Scott J, Vos T. The long-term health consequences of child physical abuse, emotional abuse, and neglect: a systematic review and meta-analysis. PLoS Med 2012;9(11):e1001349 [FREE Full text] [doi: 10.1371/journal.pmed.1001349] [Medline: 23209385]
10. Frissen A, Lieverse R, Drukker M, van Winkel R, Delespaul P, GROUP Investigators. Childhood trauma and childhood urbanicity in relation to psychotic disorder. Soc Psychiatry Psychiatr Epidemiol 2015 Oct 21;50(10):1481-1488 [FREE Full text] [doi: 10.1007/s00127-015-1049-7] [Medline: 25895686]
11. Matheson SL, Shepherd AM, Pinchbeck RM, Laurens KR, Carr VJ. Childhood adversity in schizophrenia: a systematic meta-analysis. Psychol Med 2012 Apr 30;43(2):225-238. [doi: 10.1017/s0033291712000785]
12. Kraan T, Velthorst E, Smit F, de Haan L, van der Gaag M. Trauma and recent life events in individuals at ultra high risk for psychosis: review and meta-analysis. Schizophr Res 2015 Feb;161(2-3):143-149. [doi: 10.1016/j.schres.2014.11.026] [Medline: 25499046]
13. Bendall S, Hulbert CA, Alvarez-Jimenez M, Allott K, McGorry PD, Jackson HJ. Testing a model of the relationship between childhood sexual abuse and psychosis in a first-episode psychosis group: the role of hallucinations and delusions, posttraumatic intrusions, and selective attention. J Nerv Ment Dis 2013 Nov;201(11):941-947. [doi: 10.1097/NMD.0000000000000033] [Medline: 24177480]
14. Kelleher I, Keeley H, Corcoran P, Ramsay H, Wasserman C, Carli V, et al. Childhood trauma and psychosis in a prospective cohort study: cause, effect, and directionality. Am J Psychiatry 2013 Jul;170(7):734-741. [doi: 10.1176/appi.ajp.2012.12091169] [Medline: 23599019]
15. Trotta A, Murray RM, Fisher HL. The impact of childhood adversity on the persistence of psychotic symptoms: a systematic review and meta-analysis. Psychol Med 2015 Apr 23;45(12):2481-2498. [doi: 10.1017/s0033291715000574]
16. Guloksuz S, Pries L, Delespaul P, Kenis G, Luykx JJ, Lin BD, Genetic RiskOutcome of Psychosis (GROUP) investigators, et al. Examining the independent and joint effects of molecular genetic liability and environmental exposures in schizophrenia: results from the EUGEI study. World Psychiatry 2019 Jun 06;18(2):173-182 [FREE Full text] [doi: 10.1002/wps.20629] [Medline: 31059627]
17. van Os J, van der Steen Y, Islam MA, Gülöksüz S, Rutten BP, Simons CJ. Evidence that polygenic risk for psychotic disorder is expressed in the domain of neurodevelopment, emotion regulation and attribution of salience. Psychol Med 2017 Apr 24;47(14):2421-2437. [doi: 10.1017/s0033291717000915]
18. Islam MA, Khan MF, Quee PJ, Snieder H, van den Heuvel ER, Bruggeman R, GROUP Investigators. Familial liability to psychosis is a risk factor for multimorbidity in people with psychotic disorders and their unaffected siblings. Eur Psychiatry 2017 Sep;45:81-89. [doi: 10.1016/j.eurpsy.2017.05.001] [Medline: 28750277]

19. Linscott RJ, van Os J. An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. Psychol Med 2012 Jul 31;43(6):1133-1149. [doi: 10.1017/s0033291712001626]
20. Heins M, Simons C, Lataster T, Pfeifer S, Versmissen D, Lardinois M, et al. Childhood trauma and psychosis: a case-control and case-sibling comparison across different levels of genetic liability, psychopathology, and type of trauma. Am J Psychiatry 2011 Dec;168(12):1286-1294. [doi: 10.1176/appi.ajp.2011.10101531] [Medline: 21955935]
21. Collip D, Nicolson NA, Lardinois M, Lataster T, van Os J, Myin-Germeys I. Daily cortisol, stress reactivity and psychotic experiences in individuals at above average genetic risk for psychosis. Psychol Med 2011 May 16;41(11):2305-2315. [doi: 10.1017/s0033291711000602]
22. Rauschenberg C, van Os J, Cremers D, Goedhart M, Schieveld JN, Reininghaus U. Stress sensitivity as a putative mechanism linking childhood trauma and psychopathology in youth's daily life. Acta Psychiatr Scand 2017 Oct;136(4):373-388. [doi: 10.1111/acps.12775] [Medline: 28758672]
23. Reininghaus U, Gayer-Anderson C, Valmaggia L, Kempton MJ, Calem M, Onyejiaka A, et al. Psychological processes underlying the association between childhood trauma and psychosis in daily life: an experience sampling study. Psychol Med 2016 Jul 12;46(13):2799-2813. [doi: 10.1017/s003329171600146x]
24. Williams J, Bucci S, Berry K, Varese F. Psychological mediators of the association between childhood adversities and psychosis: a systematic review. Clin Psychol Rev 2018 Nov;65:175-196. [doi: 10.1016/j.cpr.2018.05.009] [Medline: 30243100]
25. Krabbendam L, Janssen I, Bak M, Bijl RV, de Graaf R, van Os J. Neuroticism and low self-esteem as risk factors for psychosis. Soc Psychiatry Psychiatr Epidemiol 2002 Jan 1;37(1):1-6. [doi: 10.1007/s127-002-8207-y] [Medline: 11924745]
26. Ciufolini S, Morgan C, Morgan K, Fearon P, Boydell J, Hutchinson G, et al. Self esteem and self agency in first episode psychosis: ethnic variation and relationship with clinical presentation. Psychiatry Res 2015 Jun 30;227(2-3):213-218. [doi: 10.1016/j.psychres.2015.03.030] [Medline: 25868868]
27. Lecomte T, Leclerc C, Wykes T. Symptom fluctuations, self-esteem, and cohesion during group cognitive behaviour therapy for early psychosis. Psychol Psychother 2018 Mar 14;91(1):15-26. [doi: 10.1111/papt.12139] [Medline: 28707407]
28. Smith B, Fowler DG, Freeman D, Bebbington P, Bashforth H, Garety P, et al. Emotion and psychosis: links between depression, self-esteem, negative schematic beliefs and delusions and hallucinations. Schizophr Res 2006 Sep;86(1-3):181-188 [FREE Full text] [doi: 10.1016/j.schres.2006.06.018] [Medline: 16857346]
29. Garety PA, Kuipers E, Fowler D, Freeman D, Bebbington PE. A cognitive model of the positive symptoms of psychosis. Psychol Med 2001 Feb 18;31(2):189-195 [FREE Full text] [doi: 10.1017/s0033291701003312] [Medline: 11232907]
30. Bentall RP, Kinderman P, Kaney S. The self, attributional processes and abnormal beliefs: towards a model of persecutory delusions. Behav Res Ther 1994 Mar;32(3):331-341. [doi: 10.1016/0005-7967(94)90131-7] [Medline: 8192633]
31. Kinderman P, Bentall RP. Self-discrepancies and persecutory delusions: evidence for a model of paranoid ideation. J Abnormal Psychol 1996 Feb;105(1):106-113. [doi: 10.1037/0021-843x.105.1.106]
32. Kesting M, Bredenpohl M, Klenke J, Westermann S, Lincoln TM. The impact of social stress on self-esteem and paranoid ideation. J Behav Ther Exp Psychiatry 2013 Mar;44(1):122-128. [doi: 10.1016/j.jbtep.2012.07.010] [Medline: 22960709]
33. Daemen M, van Amelsvoort T, Group Investigators, Reininghaus U. Self-esteem and psychosis in daily life: an experience sampling study. J Psychopathol Clin Sci 2022 Feb;131(2):182-197. [doi: 10.1037/abn0000722] [Medline: 35230860]
34. Mansueto G, van Nierop M, Schruers K, GROUP Investigators, Alizadeh BZ, Bartels-Velthuis AA, et al. The role of cognitive functioning in the relationship between childhood trauma and a mixed phenotype of affective-anxious-psychotic symptoms in psychotic disorders. Schizophr Res 2018 Feb;192:262-268. [doi: 10.1016/j.schres.2017.04.003] [Medline: 28416093]
35. Beck AT. The evolution of the cognitive model of depression and its neurobiological correlates. Am J Psychiatry 2008 Aug;165(8):969-977. [doi: 10.1176/appi.ajp.2008.08050721] [Medline: 18628348]
36. GARETY PA, BEBBINGTON P, FOWLER D, FREEMAN D, KUIPERS E. Implications for neurobiological research of cognitive models of psychosis: a theoretical paper. Psychol Med 2007 Mar 05;37(10):1377-1391. [doi: 10.1017/s003329170700013x]
37. Morgan C, Gayer-Anderson C. Childhood adversities and psychosis: evidence, challenges, implications. World Psychiatry 2016 Jun 05;15(2):93-102 [FREE Full text] [doi: 10.1002/wps.20330] [Medline: 27265690]
38. Fisher HL, Schreier A, Zammit S, Maughan B, Munafò MR, Lewis G, et al. Pathways between childhood victimization and psychosis-like symptoms in the ALSPAC birth cohort. Schizophr Bull 2013 Sep 01;39(5):1045-1055 [FREE Full text] [doi: 10.1093/schbul/sbs088] [Medline: 22941743]
39. Morgan C, Reininghaus U, Fearon P, Hutchinson G, Morgan K, Dazzan P, et al. Modelling the interplay between childhood and adult adversity in pathways to psychosis: initial evidence from the AESOP study. Psychol Med 2014 Jan;44(2):407-419 [FREE Full text] [doi: 10.1017/S0033291713000767] [Medline: 23590972]
40. Gracie A, Freeman D, Green S, Garety PA, Kuipers E, Hardy A, et al. The association between traumatic experience, paranoia and hallucinations: a test of the predictions of psychological models. Acta Psychiatr Scand 2007 Oct;116(4):280-289. [doi: 10.1111/j.1600-0447.2007.01011.x] [Medline: 17803758]

41. Fisher HL, Appiah-Kusi E, Grant C. Anxiety and negative self-schemas mediate the association between childhood maltreatment and paranoia. Psychiatry Res 2012 Apr 30;196(2-3):323-324 [FREE Full text] [doi: 10.1016/j.psychres.2011.12.004] [Medline: 22390830]
42. Evans GJ, Reid G, Preston P, Palmier-Claus J, Sellwood W. Trauma and psychosis: the mediating role of self-concept clarity and dissociation. Psychiatry Res 2015 Aug 30;228(3):626-632. [doi: 10.1016/j.psychres.2015.04.053] [Medline: 26099655]
43. Birchwood M, Iqbal Z. Depression and suicidal thinking in psychosis: a cognitive approach. In: Outcome and Innovation in Psychological Treatment of Schizophrenia. Hoboken, New Jersey: Wiley; 1998.
44. Iqbal Z, Birchwood M, Chadwick P, Trower P. Cognitive approach to depression and suicidal thinking in psychosis. Br J Psychiatry 2018 Jan 02;177(6):522-528. [doi: 10.1192/bjp.177.6.522]
45. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. J Nerv Ment Dis 1987 Sep;175(9):526-536. [doi: 10.1097/00005053-198709000-00004] [Medline: 3655778]
46. Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, et al. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. Arch Gen Psychiatry 1999 Feb 01;56(2):162-168. [doi: 10.1001/archpsyc.56.2.162] [Medline: 10025441]
47. Korver N, Quee PJ, Boos HB, Simons CJ, de Haan L, GROUP investigators. Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. Int J Methods Psychiatr Res 2012 Sep 15;21(3):205-221 [FREE Full text] [doi: 10.1002/mpr.1352] [Medline: 22419500]
48. Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV). Washington, D.C., United States: American Psychiatric Association; 1994.
49. Delespaul P. Assessing Schizophrenia in Daily Life : The Experience Sampling Method. Maastricht: Datawyse / Universitaire Pers Maastricht; 1995.
50. Standaard Onderwijsindeling 2006. Centraal Bureau voor de Statistiek. URL: https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/onderwijs-en-beroepen/standaard-onderwijsindeling--soi--/standaard-onderwijsindeling-2006 [accessed 2021-04-07]
51. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J. Experience sampling research in psychopathology: opening the black box of daily life. Psychol Med 2009 Feb 12;39(9):1533-1547. [doi: 10.1017/s0033291708004947]
52. Palmier-Claus JE, Dunn G, Lewis SW. Emotional and symptomatic reactivity to stress in individuals at ultra-high risk of developing psychosis. Psychol Med 2011 Nov 09;42(5):1003-1012. [doi: 10.1017/s0033291711001929]
53. Myin-Germeys I, van Os J, Schwartz JE, Stone AA, Delespaul PA. Emotional reactivity to daily life stress in psychosis. Arch Gen Psychiatry 2001 Dec;58(12):1137-1144. [doi: 10.1001/archpsyc.58.12.1137] [Medline: 11735842]
54. Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul PA, et al. Experience sampling research in individuals with mental illness: reflections and guidance. Acta Psychiatr Scand 2011 Jan;123(1):12-20. [doi: 10.1111/j.1600-0447.2010.01596.x] [Medline: 20712828]
55. Hermans K, van der Steen Y, Kasanova Z, van Winkel R, Reininghaus U, Lataster T, et al. Temporal dynamics of suspiciousness and hallucinations in clinical high risk and first episode psychosis. Psychiatry Res 2020 Aug;290:113039. [doi: 10.1016/j.psychres.2020.113039] [Medline: 32460186]
56. Geldhof GJ, Preacher KJ, Zyphur MJ. Reliability estimation in a multilevel confirmatory factor analysis framework. Psychol Methods 2014 Mar;19(1):72-91. [doi: 10.1037/a0032138] [Medline: 23646988]
57. Oorschot M, Kwapil T, Delespaul P, Myin-Germeys I. Momentary assessment research in psychosis. Psychol Assess 2009 Dec;21(4):498-505. [doi: 10.1037/a0017077] [Medline: 19947784]
58. Myin-Germeys I, Delespaul P, van Os J. Behavioural sensitization to daily life stress in psychosis. Psychol Med 2005 May;35(5):733-741. [doi: 10.1017/s0033291704004179] [Medline: 15918350]
59. Aiken L, West S, Reno R. Multiple Regression Testing and Interpreting Interactions. Thousand Oaks, California, United States: SAGE Publications; 1991.
60. Cohen J, Cohen P, West S, Aiken L. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Milton Park, Abingdon-on-Thames, Oxfordshire United Kingdom: Taylor & Francis; 1975.
61. Dawson JF, Richter AW. Probing three-way interactions in moderated multiple regression: development and application of a slope difference test. J Appl Psychol 2006 Jul;91(4):917-926. [doi: 10.1037/0021-9010.91.4.917] [Medline: 16834514]
62. Fisher HL, Craig TK, Fearon P, Morgan K, Dazzan P, Lappin J, et al. Reliability and comparability of psychosis patients' retrospective reports of childhood abuse. Schizophr Bull 2011 May 23;37(3):546-553 [FREE Full text] [doi: 10.1093/schbul/sbp103] [Medline: 19776204]
63. Maughan B, Rutter M. Retrospective reporting of childhood adversity: issues in assessing long-term recall. J Pers Disord 1997 Mar;11(1):19-33. [doi: 10.1521/pedi.1997.11.1.19] [Medline: 9113820]
64. Gayer-Anderson C, Reininghaus U, Paetzold I, Hubbard K, Beards S, Mondelli V, et al. A comparison between self-report and interviewer-rated retrospective reports of childhood abuse among individuals with first-episode psychosis and population-based controls. J Psychiatr Res 2020 Apr;123:145-150 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.02.002] [Medline: 32065950]

65. Euser S, Alink LR, Pannebakker F, Vogels T, Bakermans-Kranenburg MJ, Van IJzendoorn MH. The prevalence of child maltreatment in the Netherlands across a 5-year period. Child Abuse Negl 2013 Oct;37(10):841-851. [doi: 10.1016/j.chiabu.2013.07.004] [Medline: 23938018]
66. Reininghaus U, Depp CA, Myin-Germeys I. Ecological interventionist causal models in psychosis: targeting psychological mechanisms in daily life. Schizophr Bull 2016 Mar;42(2):264-269 [FREE Full text] [doi: 10.1093/schbul/sbv193] [Medline: 26707864]
67. Gabriel AS, Podsakoff NP, Beal DJ, Scott BA, Sonnentag S, Trougakos JP, et al. Experience sampling methods: a discussion of critical trends and considerations for scholarly advancement. Organiz Res Method 2018 Oct 07;22(4):969-1006. [doi: 10.1177/1094428118802626]
68. Udachina A, Bentall RP, Varese F, Rowse G. Stress sensitivity in paranoia: poor-me paranoia protects against the unpleasant effects of social stress. Psychol Med 2017 Jun 05;47(16):2834-2843. [doi: 10.1017/s0033291717001362]
69. Rosenberg M. Society and the Adolescent Self-Image. Princeton, New Jersey: Princeton University Press; 1965.
70. Palmier-Claus JE, Dunn G, Morrison AP, Lewis SW. The role of metacognitive beliefs in stress sensitisation, self-esteem variability, and the generation of paranoia. Cogn Neuropsychiatry 2011 Nov;16(6):530-546. [doi: 10.1080/13546805.2011.561583] [Medline: 22098083]
71. Udachina A, Thewissen V, Myin-Germeys I, Fitzpatrick S, O'kane A, Bentall RP. Understanding the relationships between self-esteem, experiential avoidance, and paranoia: structural equation modelling and experience sampling studies. J Nerv Ment Dis 2009 Sep;197(9):661-668. [doi: 10.1097/NMD.0b013e3181b3b2ef] [Medline: 19752645]
72. Greenwald AG, McGhee DE, Schwartz JL. Measuring individual differences in implicit cognition: the implicit association test. J Personal Social Psychol 1998;74(6):1464-1480. [doi: 10.1037/0022-3514.74.6.1464]
73. Bebbington P, Jonas S, Kuipers E, King M, Cooper C, Brugha T, et al. Childhood sexual abuse and psychosis: data from a cross-sectional national psychiatric survey in England. Br J Psychiatry 2011 Jul 02;199(1):29-37. [doi: 10.1192/bjp.bp.110.083642] [Medline: 21508437]
74. Briere J, Runtz M. Differential adult symptomatology associated with three types of child abuse histories. Child Abuse Negl 1990 Jan;14(3):357-364. [doi: 10.1016/0145-2134(90)90007-g] [Medline: 2207804]
75. Finzi-Dottan R, Karu T. From emotional abuse in childhood to psychopathology in adulthood: a path mediated by immature defense mechanisms and self-esteem. J Nerv Ment Dis 2006 Aug;194(8):616-621. [doi: 10.1097/01.nmd.0000230654.49933.23] [Medline: 16909071]
76. Toth SL, Cicchetti D. A developmental psychopathology perspective on child maltreatment. Introduction. Child Maltreat 2013 Aug 25;18(3):135-139 [FREE Full text] [doi: 10.1177/1077559513500380] [Medline: 23886641]
77. Thornberry TP, Ireland TO, Smith CA. The importance of timing: the varying impact of childhood and adolescent maltreatment onmultiple problem outcomes. Dev Psychopathol 2002 Jan 14;13(4):957-979. [doi: 10.1017/s0954579401004114]
78. Iwaniec D, Larkin E, Higgins S. Research review: risk and resilience in cases of emotional abuse. Child Fam Soc Work 2006 Feb;11(1):73-82. [doi: 10.1111/j.1365-2206.2006.00398.x]
79. Villasana M, Alonso-Tapia J, Ruiz M. A model for assessing coping and its relation to resilience in adolescence from the perspective of “person–situation interaction”. Personal Indiv Differ 2016 Aug;98:250-256. [doi: 10.1016/j.paid.2016.04.053]
80. Kim KR, Song YY, Park JY, Lee EH, Lee M, Lee SY, et al. The relationship between psychosocial functioning and resilience and negative symptoms in individuals at ultra-high risk for psychosis. Aust N Z J Psychiatry 2013 Aug 09;47(8):762-771. [doi: 10.1177/0004867413488218] [Medline: 23661784]
81. Patel V, Goodman A. Researching protective and promotive factors in mental health. Int J Epidemiol 2007 Aug 23;36(4):703-707. [doi: 10.1093/ije/dym147] [Medline: 17646185]
82. Lincoln TM, Wilhelm K, Nestoriuc Y. Effectiveness of psychoeducation for relapse, symptoms, knowledge, adherence and functioning in psychotic disorders: a meta-analysis. Schizophr Res 2007 Nov;96(1-3):232-245. [doi: 10.1016/j.schres.2007.07.022] [Medline: 17826034]
83. Hutton P, Taylor PJ. Cognitive behavioural therapy for psychosis prevention: a systematic review and meta-analysis. Psychol Med 2013 Mar 22;44(3):449-468. [doi: 10.1017/s0033291713000354]
84. Crossley NA, Constante M, McGuire P, Power P. Efficacy of atypical v. typical antipsychotics in the treatment of early psychosis: meta-analysis. Br J Psychiatry 2010 Jun 02;196(6):434-439 [FREE Full text] [doi: 10.1192/bjp.bp.109.066217] [Medline: 20513851]
85. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
86. Sitko K, Bewick BM, Owens D, Masterson C. Meta-analysis and meta-regression of cognitive behavioral therapy for psychosis (CBTp) across time: the effectiveness of CBTp has improved for delusions. Schizophrenia Bull Open 2020 Jan;1(1):sgaa023. [doi: 10.1093/schizbullopen/sgaa023]
87. van den Berg DP, de Bont PA, van der Vleugel BM, de Roos C, de Jongh A, Van Minnen A, et al. Prolonged exposure vs eye movement desensitization and reprocessing vs waiting list for posttraumatic stress disorder in patients with a psychotic disorder: a randomized clinical trial. JAMA Psychiatry 2015 Mar 01;72(3):259-267. [doi: 10.1001/jamapsychiatry.2014.2637] [Medline: 25607833]

88. Haverkampf CJ. Antipsychotics: emotional flattening vs apathy. J Psychiatry Psychother Commun 2013;2(2):31-32.
89. de Neef M. Build Your Confidence with CBT 6 Simple Steps to Be Happier, More Successful and Fulfilled. UK: Open University Press; 2015.
90. Daemen M, Postma MR, Lindauer R, Hoes-van der Meulen I, Nieman D, Delespaul P, et al. Efficacy of a transdiagnostic ecological momentary intervention for improving self-esteem (SELFIE) in youth exposed to childhood adversity: study protocol for a multi-center randomized controlled trial. Trials 2021 Sep 19;22(1):641 [FREE Full text] [doi: 10.1186/s13063-021-05585-y] [Medline: 34538272]
91. Larson MK, Walker EF, Compton MT. Early signs, diagnosis and therapeutics of the prodromal phase of schizophrenia and related psychotic disorders. Expert Rev Neurother 2010 Aug 09;10(8):1347-1359 [FREE Full text] [doi: 10.1586/ern.10.93] [Medline: 20662758]

Abbreviations

CTQ: Childhood Trauma Questionnaire
EMI: ecological momentary intervention
ESM: experience sampling method
GROUP: Genetic Risk and Outcome in Psychosis

Edited by J Torous; submitted 08.10.21; peer-reviewed by JM Suelves, J Thrul; comments to author 05.03.22; revised version received 28.07.22; accepted 09.12.22; published 05.04.23.

Please cite as:
Daemen M, van Amelsvoort T, GROUP investigators, Reininghaus U
Momentary Self-esteem as a Process Underlying the Association Between Childhood Trauma and Psychosis: Experience Sampling Study
JMIR Ment Health 2023;10:e34147
URL: https://mental.jmir.org/2023/1/e34147
doi:10.2196/34147
PMID:37018034

Original Paper

Understanding What Drives Long-term Engagement in Digital Mental Health Interventions: Secondary Causal Analysis of the Relationship Between Social Networking and Therapy Engagement

Shaunagh O'Sullivan[1,2], BA, MPsychSc; Niels van Berkel[3], PhD; Vassilis Kostakos[4], BSc, PhD; Lianne Schmaal[1,2], BA, MSc, PhD; Simon D'Alfonso[4], BASc (Hons), PhD; Lee Valentine[1,2], BA, MSocWk, PhD; Sarah Bendall[1,2], BA, PGDipClinPsych, MA, PhD; Barnaby Nelson[1,2], MPsych, PhD; John F Gleeson[5,6], BA (Hons), MPsych, PhD; Mario Alvarez-Jimenez[1,2], DClinPsy, MAResearchMeth, PhD

[1]Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia

[2]Orygen, Parkville, Australia

[3]Department of Computer Science, Aalborg University, Aalborg, Denmark

[4]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[5]Healthy Brain and Mind Research Centre, Australian Catholic University, Melbourne, Australia

[6]School of Behavioural and Health Sciences, Australian Catholic University, Melbourne, Australia

Corresponding Author:
Shaunagh O'Sullivan, BA, MPsychSc
Orygen
35 Poplar Road
Parkville, 3052
Australia
Phone: 61 428282470
Email: shaunagh.osullivan@orygen.org.au

Abstract

Background: Low engagement rates with digital mental health interventions are a major challenge in the field. Multicomponent digital interventions aim to improve engagement by adding components such as social networks. Although social networks may be engaging, they may not be sufficient to improve clinical outcomes or lead users to engage with key therapeutic components. Therefore, we need to understand what components drive engagement with digital mental health interventions overall and what drives engagement with key therapeutic components.

Objective: Horyzons was an 18-month digital mental health intervention for young people recovering from first-episode psychosis, incorporating therapeutic content and a private social network. However, it is unclear whether use of the social network leads to subsequent use of therapeutic content or vice versa. This study aimed to determine the causal relationship between the social networking and therapeutic components of Horyzons.

Methods: Participants comprised 82 young people (16-27 years) recovering from first-episode psychosis. Multiple convergent cross mapping was used to test causality, as a secondary analysis of the Horyzons intervention. Multiple convergent cross mapping tested the direction of the relationship between each pair of social and therapeutic system usage variables on Horyzons, using longitudinal usage data.

Results: Results indicated that the social networking aspects of Horyzons were most engaging. Posting on the social network drove engagement with all therapeutic components (r=0.06-0.36). Reacting to social network posts drove engagement with all therapeutic components (r=0.39-0.65). Commenting on social network posts drove engagement with most therapeutic components (r=0.11-0.18). Liking social network posts drove engagement with most therapeutic components (r=0.09-0.17). However, starting a therapy pathway led to commenting on social network posts (r=0.05) and liking social network posts (r=0.06), and completing a therapy action led to commenting on social network posts (r=0.14) and liking social network posts (r=0.15).

Conclusions: The online social network was a key driver of long-term engagement with the Horyzons intervention and fostered engagement with key therapeutic components and ingredients of the intervention. Online social networks can be further leveraged

to engage young people with therapeutic content to ensure treatment effects are maintained and to create virtuous cycles between all intervention components to maintain engagement.

Trial Registration: Australian New Zealand Clinical Trials Registry: ACTRN12614000009617; https://www.australianclinicaltrials.gov.au/anzctr/trial/ACTRN12614000009617

(JMIR Ment Health 2023;10:e44812) doi:10.2196/44812

KEYWORDS

digital intervention; digital health; youth mental health; psychotic disorders; usage metrics; log data; social networking

Introduction

Background

Poor engagement with digital mental health interventions is a major challenge in the field [1]. For example, Fleming and colleagues [2] found that there is a 15-day retention rate of only 3.9% and a 30-day retention rate of only 3.3% for the use of mental health apps in the general population and low completion or usage rates beyond 6 weeks (0.5% to 28.6%) in digital interventions targeting depression and anxiety.

A recent development in the field is the personalization of digital mental health interventions to enhance the user experience and engagement [3-6]. These include the use of machine learning algorithms [3,4] and prompts or reminders [5,6] to tailor interventions to individual users needs and patterns of engagement. Another novel example is the use of multicomponent interventions that increase flexibility to personalize and cater for the needs of different individuals, thereby promoting engagement [1,7]. These interventions capitalize on human support to promote engagement and clinical effects and on social media to promote sustained engagement and social support and connectedness [8-10]. Although there are promising indications that additional features in multicomponent digital mental health interventions may increase long-term engagement, little is known about which components are key for efficacy and which components contribute to engagement.

To examine the relationship between engagement and effectiveness, our recent study used a novel method to analyze the relationship between patterns of use of a multicomponent digital platform and treatment effects [11]. Specifically, we characterized the use of social networking and therapeutic aspects of Horyzons, which was an 18-month intervention aiming to improve social functioning, vocational recovery, and relapse prevention in young people recovering from first-episode psychosis (FEP) after receiving 2 years of specialized care [8,12,13]. It included therapy content; a private social network; and peer, clinical, and vocational support. We identified 3 user profiles: (1) low usage, (2) maintained usage of social networking elements, and (3) maintained usage of both social networking and therapeutic elements of the intervention. Results indicated that engagement with the therapy components of the platform was needed for improved outcomes in terms of social functioning, overall psychiatric symptom severity, and negative psychotic symptoms and that engaging with the social networking components of the platform alone was not sufficient to bring about clinical benefits. However, social networking may be key for engagement, as all sustained use of the intervention included use of the social network [11,14]. However, analyses such as those used in this recent study are not designed to measure drivers of engagement and causal relationships between intervention components. Thus, we need novel methods to analyze what drives engagement overall and what drives engagement with the subset of specific components associated with therapeutic mechanisms.

The field of human computer interaction (HCI) utilizes empirical dynamic modeling (EDM) as a nonlinear analysis method to investigate human use of technology from time series data, where interactions with technology change over time and may be bidirectional, and associations can be inferred from the data rather than predetermined hypotheses [15]. Convergent cross mapping (CCM) is a type of analysis based on EDM that tests causal relationships between variables, by establishing if states of the causal variable can be recovered from the time series of the affected variable [16]. Researchers then extended this approach to multiple CCM (MCCM), which accounted for multiple relationships or users of technology in the analysis [17]. As relationships between system use variables can be complex and nonlinear, linear analysis methods and those that make correlation-based inferences are not appropriate to investigate what aspects of usage lead to subsequent usage of intervention components [15].

The field of HCI has started to investigate causal and potentially bidirectional relationships between users and technology using MCCM, to improve and design better interactive technology [17]. It has been suggested that interaction with technology should be considered as a complex dynamic system, whereby our use of a technological system is affected by our mental state and our mental state is affected by our use of a technological system [17]. Usage behavior has previously been analyzed to determine the causal relationship between smartphone application use and emotional states [18]. This study found that application use drives user emotions in most cases but that user emotions also drive some aspects of application use [19]. Determining causality is important so that developers and researchers can take appropriate actions to ensure users access engaging information efficiently and effectively and to maximize the use of all intervention components. Furthermore, if we know what intervention components are linked to better outcomes, we can optimize interventions to enhance engagement with the therapeutic or active ingredients of these interventions.

Objective

Horyzons is an example of a unique long-term multicomponent digital intervention aiming to maintain long-term treatment

effects and engagement. Our previous research found a relationship between improved outcomes and combined use of the therapy and social networking components of Horyzons [11]. However, little is known about whether aspects of the social network aspect drive use with aspects of therapy content or vice versa. This study, therefore, aimed to determine the causal relationship between different aspects of system use and whether use of the social network leads to subsequent use of therapeutic content by applying the MCCM method.

Methods

Study Design

Horyzons was a single-blind, 18-month, randomized controlled trial (RCT) conducted with young people recovering from FEP, following 2 years of specialist early intervention treatment. Participants were randomly allocated to either treatment as usual (TAU) or TAU as well as 18 months of access to the Horyzons intervention [12]. Horyzons was underpinned by the moderated online social therapy (MOST) model, which integrates (1) interactive online therapy ("Pathways and Steps"), (2) peer-to-peer online social networking, (3) peer moderation, and (4) expert clinical and vocational support [20].

Ethics Approval

Ethics approval was obtained from the Melbourne Health Research Ethics Committee (2013.146).

Participants

Participants included 86 young people allocated to the Horyzons intervention. These participants were recruited after receiving 2 years of specialized care from the Early Psychosis Prevention and Intervention Centre (EPPIC) at Orygen, in Melbourne Australia between October 2013 and January 2017. EPPIC is a specialist FEP service that provides 18 months to 24 months of specialist early intervention for FEP to young people aged 15 years to 24 years [21,22].

Of the participants allocated to the intervention arm of the RCT, 5% (4/86) did not engage with the Horyzons platform. As there was no valid usage data, these participants were excluded from the analysis. The remaining 82 intervention participants were aged between 16 years and 27 years at randomization (mean 21, SD 2.88 years). As participants were recruited following their completion of early intervention treatment, this accounts for the age range extending to 27 years old. According to the Horyzons RCT eligibility criteria [13], participants were required to meet the criteria for an FEP disorder or mood disorder with psychotic features according to the Diagnostic and Statistical Manual of Mental Disorders 4th Edition [23], to have not been treated with antipsychotic medication for more than 6 months before attending EPPIC, and to have demonstrated remission of positive symptoms of psychosis for 4 weeks or more at the time of enrolment in the Horyzons study, as measured by the Positive and Negative Syndrome Scale [24].

System Usage Metrics

System usage metrics were extracted from the Horyzons online platform for each user for each day of their trial involvement (range: 282-528 days per user), resulting in a total of 47,060 cases of user interaction. See Table 1 for an overview of the metrics representing aspects of daily usage of the intervention's therapeutic and social components.

Table 1. System usage variables extracted from the Horyzons platform.

Variable type	Variables
Therapy-related variables (number of)	• Steps[a] started • Pathways[b] started • Actions[c] done • Visits to suggested content[d] • Visits to messages[e] • Visits to therapy[f]
Social networking–related variables (number of)	• News feed[g] posts • News feed comments • Likes made • Reactions[h] made

[a]Steps refer to the intervention modules.

[b]Pathways refer to a collection of intervention modules related to a topic (eg, anxiety).

[c]Actions refer to behavioral activities aiming to translate learning into behavior.

[d]Suggested content refers to therapeutic content recommended by clinical moderators.

[e]Messages refer to a private message section, where moderators could contact participants directly.

[f]Visiting therapy refers to visiting the home page of the therapy component of the intervention.

[g]The news feed refers to the social network.

[h]Reactions refer to short support messages in response to a post (eg, thinking of you).

We categorized *pathways* based on their therapeutic targets, which included understanding psychosis; identifying early warning signs to prevent relapse; identifying and exercising personal strengths; promoting social connections and positive emotions; and managing stress, anxiety, and depression. As a means to increase engagement, *Pathways* were distilled into

shorter, interactive *Steps* (eg, illustrating how to respond empathically to others [to foster positive connections]). See Multimedia Appendix 1 for an example of a *Step* on Horyzons. Each *Step* was partnered with an *Action* or "*Do It*," which was designed to support the translation from learning into behavioral change (eg, suggestions on how to exercise empathy in specific contexts). Expert clinical moderators could also recommend *Pathways*,*Steps*, *Actions*, and *Talk it Outs* (described in the following paragraphs) that were personally appropriate to the young person via a private message, which would appear as a notification in the user's dashboard. Furthermore, users could visit the therapeutic components of Horyzons without completing any therapeutic content (eg, viewing what *Pathway* and *Step* were currently allocated to them).

The social network was moderated and led by peer workers. MOST peer workers were young people who identified as having a lived experience of mental ill-health and who had been employed and trained to offer support and guidance to others on the MOST platform. The social network was designed for participants to communicate and foster a sense of social support. Participants were able to post comments, or "like," "respond," or "react," to comments posted by other young people. A set of limited *reactions* (or *emojis*) was available to communicate social support in response to posts (eg, "I get you," "thinking of you"). Finally, the *Talk It Out* function allowed young people to suggest relevant discussion topics to take place in a separate forum moderated by the peer workers. The function was informed by an evidence-based problem-solving framework [25]. Participants received notifications when other users communicated on the social network. Participants also received private messages when a moderator contacted them directly via the platform. See Multimedia Appendix 2 for an example of a newsfeed post with *likes* and *reactions* on the Horyzons social network.

Statistical Analyses

MCCM

MCCM was used to determine the causal and potentially bidirectional relationship between the aforementioned system usage metrics. For example, we tested whether usage of certain aspects of the social network leads to usage of certain therapy components or vice versa. This is a new methodology that extends beyond existing CCM methods and was adapted from that of Van Berkel and colleagues [17].

CCM is a core component of the EDM approach, which is a set of methods designed to characterize and test causality in complex dynamic systems, such as users interacting with technology over time [17]. CCM was developed for use with time series data to distinguish causality from correlation [26]. CCM considers time series data from a complex and dynamic systems perspective and investigates the relationships between variables in a system that is not completely random [17]. CCM is considered to be a novel and suitable method for studying human behavior when it is modeled on a complex system [19,27,28]. Analyses were conducted using the "rEDM" R package [29,30]. The 4 steps of the method are described in the following paragraphs.

Identify the Optimum Value for E (Embedding Dimension)

The optimum embedding dimension (E) was identified between each pair of system usage variables tested between the social and therapy components of the intervention. This was done using simplex projection, which is the most direct projection technique and recommended for evaluating embedding dimensions for EDM [29]. The method uses time delay embedding on a single variable (y) to generate a complex system reconstruction, by using history information of another variable (x), and determines how much information about x has been encoded into y [31]. It is justified by the embedding theorem by Takens [31-33], where, if x influences y, the historical values of x can be recovered from y. It then applies the simplex projection algorithm to make forecasts, and the highest E value is selected [19].

Test for Nonlinearity

As CCM is a nonlinear approach, it was also important to test whether system usage evolves in a nonlinear way. The rEMD package uses S-maps to characterize the degree of nonlinearity in the time series, by using the E chosen from the previous step of simplex projection and then estimating a linear map that uses the E-dimensional points on a manifold's surface (ie, dimension of Euclidean space) to predict the future [17,26]. The data are nonlinear if the maximum forecasts skill is greater than 0.

Convergent Cross Mapping

CCM was then used to identify potential causal links between pairs of system usage variables for each user. Using this approach, pairs of variables are mapped to each other using the nearest neighbors of each point on the E-dimensional manifolds (ie, the causal effects of x on y are determined by how well y cross maps x). When the number of points on the manifold increases, the nearest neighbors tend to become nearer, which improves predictions if the variables are causally linked (ie, convergence) [17].

Multiple Convergent Cross Mapping

The previous step analyzed the causal relationship between variables for each individual user, which works well when trying to understand the dynamics of an individual ecosystem (ie, 1 user). This step extends upon that approach using a geometric approach and summarizes the results from multiple CCM analyses. This is called MCCM and enables us to obtain insights into the behavior of the entire study population [17]. Effect sizes were calculated using Pearson r correlation (along with standard deviations), with scores ranging from 0 to 1.

Results

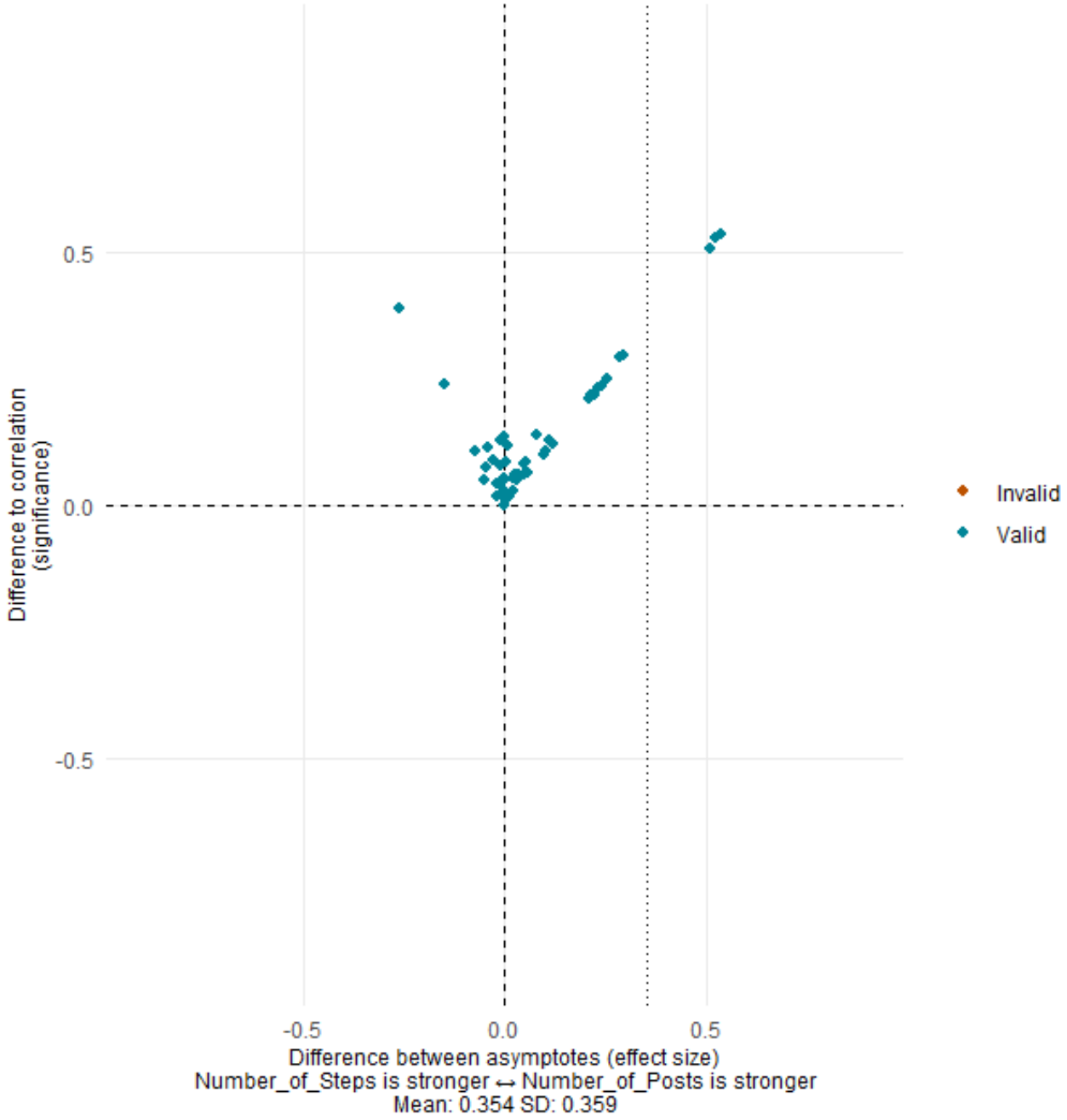
Overview

The causal and potentially bidirectional relationship between usage of social networking and therapeutic aspects of Horyzons were investigated, to determine whether social networking usage leads to engagement with therapy content or vice versa.

A visual representation of results for each pair of variables for all combinations of variables described in Table 1 (eg, steps

started [therapy] versus posts made [social networking]) is available in Multimedia Appendices 3-6, and an example of a graph for 1 pair of variables can be seen in Figure 1. Each participant in the graph is represented by a dot. For each participant, the CCM algorithm determines whether variable 1 (completing therapy steps) is driven by variable 2 (posting on the social network) or vice versa.

Figure 1. Visual representation of the direction of the causal relationship between completing therapeutic steps and posting on the social network.



The direction and magnitude of effects are represented by the position of the dots on the x axis. If the dot is to the left of the y axis, it indicates that the effect is in a certain direction (ie, completing therapy steps drives posting on the social network). If the dot is to the right, it indicates that the effect is in the other direction (ie, posting on the social network drives completing therapy steps; see Figure 1).

The distance from the x axis indicates how strong the forecasting ability is compared with a plain correlation, and the dashed vertical line indicates the mean effect size of causality across all participants. Any red dots that are present on graphs should be ignored as they represent invalid participant data, due to CCM not being able to provide power beyond that of a simple correlation [19].

Directional Relationship Between Posting on the Social Network and Therapy Engagement

MCCM was used to determine the causal relationship between social networking and therapy engagement. Table 2 shows the mean effect size of causality (Pearson *r*) among the analyzed variable pairs. These results indicate that posting on the social network drives all aspects of therapy engagement, including

completing steps (r=0.354), completing pathways (r=0.154), completing actions (r=0.055), visiting suggested content (r=0.295), visiting messages (r=0.284), and visiting therapy (r=0.360). A visual inspection of these results can be found in Multimedia Appendix 3, where the prevalence of points to the right of most of the graphs indicates that posts made on the social network lead to usage of therapeutic content for most participants. However, there are also some participants with dots to the left of the graphs, indicating that their usage of therapeutic content leads to posting on the social network. Overall, however, we conclude that posts drive therapy engagement as indicated by the vertical dotted line to the right of all graphs in Multimedia Appendix 3.

Table 2. Effect sizes (and standard deviations) for causality between therapy usage and social network usage.

Therapy usage	Social network usage			
	Posts made, *r* (SD)	Comments made, *r* (SD)	Likes made, *r* (SD)	Reactions made, *r* (SD)
Steps started	0.354 (0.359)[a]	0.161 (0.326)[a]	0.166 (0.315)[a]	0.653 (0.230)[a]
Pathways started	0.154 (0.467)[a]	–0.047 (0.418)[b]	–0.058 (0.397)[b]	0.484 (0.446)[a]
Actions completed	0.055 (0.437)[a]	–0.142 (0.393)[b]	–0.148 (0.397)[b]	0.390 (0.422)[a]
Suggested content visited	0.295 (0.373)[a]	0.120 (0.315)[a]	0.110 (0.327)[a]	0.582 (0.316)[a]
Messages visited	0.284 (0.359)[a]	0.112 (0.306)[a]	0.087 (0.338)[a]	0.589 (0.288)[a]
Therapy visited	0.360 (0.350)[a]	0.177 (0.307)[a]	0.167 (0.296)[a]	0.649 (0.230)[a]

[a]Social network usage drives therapy usage.

[b]Therapy usage drives social network usage.

Directional Relationship Between Commenting on the Social Network and Therapy Engagement

On the other hand, comments made on the social network and therapy engagement demonstrated a bidirectional relationship, whereby some aspects of therapy engagement such as completing pathways (r=–0.047) and completing actions (r=–0.142) drove the commenting on the social network, while commenting on the social network drove other aspects of therapy engagement such as completing steps (r=0.161), visiting suggested content (r=0.120), visiting messages (r=0.112), and visiting therapy (r=0.177; see Table 2). A visual inspection of these results can be found in Multimedia Appendix 4.

Directional Relationship Between Liking Posts on the Social Network and Therapy Engagement

Similarly, likes made on the social network and therapy engagement have a bidirectional relationship, whereby some aspects of therapy engagement such as completing pathways (r=–0.058) and completing actions (r=–0.148) drove liking a post on the social network, while liking a post on the social network drove other aspects of therapy engagement such as completing steps (r=0.166), visiting suggested content (r=0.110), visiting messages (r=0.087), and visiting therapy (r=0.167; see Table 2). A visual inspection of these results can be found in Multimedia Appendix 5.

Directional Relationship Between Reacting to Posts on the Social Network and Therapy Engagement

Finally, our MCCM analysis showed that reacting to posts made on the social network drove all aspects of therapy engagement, including completing steps (r=0.653), completing pathways (r=0.484), completing actions (r=0.390), visiting suggested content (r=0.582), visiting messages (r=0.589), and visiting therapy (r=0.649; see Table 2). A visual inspection of these results can be found in Multimedia Appendix 6.

Discussion

Principal Findings

This was the first study to use a novel modeling technique like MCCM to determine the causal relationship between different aspects of use of a long-term multicomponent digital intervention (Horyzons) to improve social functioning, improve vocational recovery, and prevent relapse in FEP. We found that posting on the social network and reacting to posts on the social network led to engaging with all aspects of therapy (including completing steps, pathways, and actions and visiting suggested content, messages, and therapy). We also found that commenting on the social network and liking posts made on the social network led to most aspects of therapy engagement (including completing steps and visiting suggested content, messages, and therapy). In other words, young people's use of the social network was found to increase therapy use on Horyzons. Conversely, we found a bidirectional relationship between therapeutic and social components, whereby completing actions and pathways led to commenting on posts and liking posts on the social network, even though these aspects of social network usage led to engagement with all other measured aspects of therapy.

Overall, we found that usage of social networking aspects of Horyzons drove engagement with the therapeutic aspects in most cases. Therefore, we need to further understand what drives engagement with the social network [34], but it is also important to understand how the social network drives engagement with digital interventions in general (eg, self-determination theory [SDT] can inform engagement) and what aspects of the social network can drive engagement with therapy specifically (eg,

social validation and social comparison). Previous research has also indicated that the use of prompts, texts, and emails may enhance engagement with digital interventions in general [6,35]. Horyzons was underpinned by SDT to inform engagement, which predicts that support for the following 3 basic psychological needs promotes motivation for behavior change: (1) autonomy (having a choice about how to behave), (2) competence (being able to make changes to achieve desired outcomes), and (3) relatedness (feeling accepted in one's social environment) [36]. According to SDT, participants' relatedness needs have to be supported by the system for engagement to occur, which was incorporated into supportive accountability. However, future qualitative research could further explore why the use of social networking aspects of interventions lead people to then engage with therapy components and explore whether positive feedback or experiences from the social network encouraged people to further engage with therapy content.

It is then essential to engineer social networks in digital mental health interventions in ways that promote engagement with key therapeutic ingredients. This is important, as recent research has found that those who maintain use of both social networking and therapy components of an intervention display improved outcomes, but use of social networks alone may not lead to improvements [11]. Interestingly, this study found that the social network was useful to promote overall engagement as well as engagement with key therapeutic ingredients of the platform. This is needed, as sustained engagement with both social and therapy content has been associated with improved social functioning, negative symptoms, and overall symptoms [11]. Therefore, although the social network may not be therapeutic in isolation, it works as part of a multicomponent digital mental health intervention and has significant value in promoting engagement with key therapeutic ingredients of the platform, especially with the use of reactions leading to use of all therapy components (r=0.39-0.65). Sustained engagement is needed in digital mental health interventions, especially in long-term relapse prevention interventions like Horyzons targeting long-term outcomes. Horyzons showcased elements of therapy on the social network, which possibly contributed to the findings in this study. However, this could be improved by adding additional engaging social networking components, and social networks more generally could aim to improve outcomes in this way in the future, by showcasing therapy on their social networks and providing more personalized suggestions for therapy via the social network.

In particular, the largest effect sizes were observed for reacting to posts on the social network, which led to usage of all assessed aspects of therapy (r=0.39-0.65). This is an interesting finding, as reacting to posts on the social network was deemed passive use of the social network [11]. However, passive use of social media has been associated with increased anxiety in adolescents, but this was not linked to therapeutic sites specially [37]. Therefore, although passive elements of the social network may be the most engaging, it is important that they lead young people to therapeutic content to bring about clinical benefits. Furthermore, the reactions used in Horyzons were more active than reactions on traditional social networks, as they were based on social support and validation (eg, thinking of you, I get you). It may be the case that these types of reactions, which were engineered to promote safe and meaningful connections, may be a better driver of engagement with therapy than more passive reactions, providing insight into types of novel design solutions that can enhance engagement with key therapeutic ingredients.

However, it should be noted that this study also indicated that some aspects of therapy led to social network use, potentially generating synergistic effects between social networking and therapy usage once the transition to therapy is made. This, in turn, could promote engagement with both aspects of the intervention where usage of the social network and key therapeutic ingredients reinforce each other, leading to sustained engagement. For example, once the transition is made to engage with therapy, there may be a synergistic effect whereby social networking drives therapy engagement but also some aspects of therapy usage that drive social networking (eg, in this study, we found that the number of actions completed drives commenting on the social network). Learning from others and sharing experiences may lead to increased motivation to engage with the platform and with therapy more specifically and, in turn, lead to people sharing their therapy journeys with each other [38]. In fact, the MOST platform (which Horyzons was based on) was created to do this by including *talking points* in the interactive therapy content, where a user could comment on the therapy content that was then shared to the social network, thus promoting back and forth cycles between both elements. The findings from this study partly support this hypothesis but indicate that more research is needed to optimize the synergistic use of both elements, which are likely needed for long-term engagement with digital interventions and is a challenge in the field [38].

This study also suggests we need to establish how different components of a platform interact to promote engagement and drive positive outcomes, and the MCCM method has enormous potential to optimize platforms to be more engaging and more therapeutic. Future research should focus on how to optimize the platform to make the social network more therapeutic, promote more engagement with therapy, and generate back and forth virtuous cycles between both intervention components. Predictors of engagement could also be explored (eg, using prompts) and applied to aspects of the intervention that drive engagement with therapeutic content or key active ingredients (eg, a notification asking a participant if they would like to react to a new post on the social network) [39]. Establishing engagement with therapeutic social networks can promote engagement with other therapeutic aspects of digital mental health interventions, which have displayed high rates of attrition to date. This, in turn, could lead to optimization of the synergistic effects of intervention components in multicomponent digital interventions to improve outcomes.

Limitations

However, a number of limitations must also be noted. The sample size for this study was small, comprising 82 young people, so findings should be interpreted with caution. Although this was the first causal analysis exploring how the use of certain intervention components leads to subsequent use of other components, this study only tested the direction of the

relationship between therapy and social networking components of a digital mental health intervention. It is possible that other multicomponent digital mental health interventions may include additional components, and some may be more engaging than a social network. It is also possible that multicomponent interventions not focused on mental health may have different engaging components, and future research should consider this.

Conclusions

To date, digital interventions have shown significant limitations with long-term participant engagement, a necessary ingredient toward bringing about long-term improvements in mental health. Multicomponent digital interventions that incorporate social networks have been proposed as a means to promote long-term engagement and tackle elusive long-term targets such as relapse prevention, which are rarely targeted by digital interventions. In a previous study, we identified that sustained use of both therapy and social networking components of a digital mental health intervention led to improved outcomes for young people recovering from FEP. This study adds to this knowledge by showing that the social network is possibly a necessary ingredient for long-term engagement and positive outcomes but not sufficient when used alone. Our study highlights the possible value of the social network in promoting engagement with key therapeutic ingredients of the intervention. Future research should aim to confirm these findings with a larger sample size and, if confirmed, could focus on how to optimize the social network to be more therapeutic, to more widely promote engagement with key therapeutic ingredients of the intervention, and to create virtuous cycles between use of the social network and therapeutic content.

Acknowledgments

The authors would like to thank the young people who participated in the Horyzons study and agreed to share their data for research purposes. The study was funded by the Mental Illness Research Fund (Victorian Government, Australia), the Australian National Health and Medical Research Council (NHMRC), and the Telstra Foundation. MAJ was supported by an Investigator Grant (APP1177235) from the NHMRC and a Dame Kate Campbell Fellowship from The University of Melbourne. LS was supported by an NHMRC Investigator Grant (2017962). SB was supported by the University of Melbourne Ronald Philip Griffiths Fellowship. BN was supported by an NHMRC senior research fellowship (1137687).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Horyzons step: how to flourish.
[PNG File , 242 KB - mental_v10i1e44812_app1.png]

Multimedia Appendix 2
Horyzons social network.
[PNG File , 221 KB - mental_v10i1e44812_app2.png]

Multimedia Appendix 3
Social network posts drive therapy engagement.
[PNG File , 107 KB - mental_v10i1e44812_app3.png]

Multimedia Appendix 4
Comments made and therapy engagement have a bidirectional relationship.
[PNG File , 115 KB - mental_v10i1e44812_app4.png]

Multimedia Appendix 5
Likes made and therapy engagement have a bidirectional relationship.
[PNG File , 124 KB - mental_v10i1e44812_app5.png]

Multimedia Appendix 6
Reactions drive therapy engagement.
[PNG File , 118 KB - mental_v10i1e44812_app6.png]

References

1. de Beurs D, van Bruinessen I, Noordman J, Friele R, van Dulmen S. Active involvement of end users when developing web-based mental health interventions. Front Psychiatry 2017 May 03;8:72 [FREE Full text] [doi: 10.3389/fpsyt.2017.00072] [Medline: 28515699]
2. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
3. Gönül S, Namlı T, Coşar A, Toroslu. A reinforcement learning based algorithm for personalization of digital, just-in-time, adaptive interventions. Artif Intell Med 2021 May;115:102062. [doi: 10.1016/j.artmed.2021.102062] [Medline: 34001322]
4. Lewis R, Ferguson C, Wilks C, Jones N, Picard RW. Can a Recommender System Support Treatment Personalisation in Digital Mental Health Therapy? A Quantitative Feasibility Assessment Using Data from a Behavioural Activation Therapy App. 2022 Presented at: CHI Conference on Human Factors in Computing Systems; April 29-May 5, 2022; New Orleans, LA URL: https://dl.acm.org/doi/abs/10.1145/3491101.3519840 [doi: 10.1145/3491101.3519840]
5. Gan DZQ, McGillivray L, Larsen ME, Christensen H, Torok M. Technology-supported strategies for promoting user engagement with digital mental health interventions: A systematic review. Digit Health 2022 Jun 01;8:20552076221098268 [FREE Full text] [doi: 10.1177/20552076221098268] [Medline: 35677785]
6. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6. [doi: 10.2196/jmir.4790] [Medline: 26747176]
7. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
8. Alvarez-Jimenez M, Bendall S, Lederman R, Wadley G, Chinnery G, Vargas S, et al. On the HORYZON: moderated online social therapy for long-term recovery in first episode psychosis. Schizophr Res 2013 Jan;143(1):143-149. [doi: 10.1016/j.schres.2012.10.009] [Medline: 23146146]
9. Alvarez-Jimenez M, Gleeson J, Bendall S, Penn D, Yung A, Ryan R, et al. Enhancing social functioning in young people at Ultra High Risk (UHR) for psychosis: A pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]
10. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2018 Aug 17;12(4):613-625. [doi: 10.1111/eip.12354] [Medline: 27311581]
11. O'Sullivan S, Schmaal L, D'Alfonso S, Toenders YJ, Valentine L, McEnery C, et al. Characterizing use of a multicomponent digital intervention to predict treatment outcomes in first-episode psychosis: cluster analysis. JMIR Ment Health 2022 Apr 07;9(4):e29211 [FREE Full text] [doi: 10.2196/29211] [Medline: 35389351]
12. Alvarez-Jimenez M, Bendall S, Koval P, Rice S, Cagliarini D, Valentine L, et al. HORYZONS trial: protocol for a randomised controlled trial of a moderated online social therapy to maintain treatment effects from first-episode psychosis services. BMJ Open 2019 Feb 19;9(2):e024104 [FREE Full text] [doi: 10.1136/bmjopen-2018-024104] [Medline: 30782893]
13. Alvarez-Jimenez M, Koval P, Schmaal L, Bendall S, O'Sullivan S, Cagliarini D, et al. The Horyzons project: a randomized controlled trial of a novel online social therapy to maintain treatment effects from specialist first-episode psychosis services. World Psychiatry 2021 Jun 18;20(2):233-243 [FREE Full text] [doi: 10.1002/wps.20858] [Medline: 34002511]
14. Alvarez-Jimenez M, Alcazar-Corcoles M, González-Blanch C, Bendall S, McGorry P, Gleeson J. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
15. Sugihara G, May R, Ye H, Hsieh C, Deyle E, Fogarty M, et al. Detecting causality in complex ecosystems. Science 2012 Oct 26;338(6106):496-500. [doi: 10.1126/science.1227079] [Medline: 22997134]
16. Tsonis AA, Deyle ER, Ye H, Sugihara G. Convergent Cross Mapping: Theory and an Example. In: Tsonis AA, editor. Advances in Nonlinear Geosciences. Cham, Switzerland: Springer; 2018:587-600.
17. van Berkel N, Dennis S, Zyphur M, Li J, Heathcote A, Kostakos V. Modeling interaction as a complex system. Human–Computer Interaction 2020 Jan 30;36(4):279-305. [doi: 10.1080/07370024.2020.1715221]
18. Mehrotra A, Tsapeli F, Hendley R, Musolesi M. MyTraces: Investigating correlation and causation between users' emotional states and mobile phone interaction. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-21. [doi: 10.1145/3130948]
19. Sarsenbayeva Z, Marini G, van Berkel N, Luo C, Jiang W, Yang K, et al. Does smartphone use drive our emotions or vice versa? A causal analysis. 2020 Presented at: CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI URL: https://dl.acm.org/doi/10.1145/3313831.3376163
20. Alvarez-Jimenez M, Rice S, D'Alfonso S, Leicester S, Bendall S, Pryor I, et al. A novel multimodal digital service (Moderated Online Social Therapy+) for help-seeking young people experiencing mental ill-health: pilot evaluation within a national youth e-Mental health service. J Med Internet Res 2020 Aug 13;22(8):e17155 [FREE Full text] [doi: 10.2196/17155] [Medline: 32788151]

21. McGorry PD, Edwards J, Mihalopoulos C, Harrigan SM, Jackson HJ. EPPIC: an evolving system of early detection and optimal management. Schizophr Bull 1996 Jan 01;22(2):305-326. [doi: 10.1093/schbul/22.2.305] [Medline: 8782288]
22. Cotton SM, Filia KM, Ratheesh A, Pennell K, Goldstone S, McGorry PD. Early psychosis research at Orygen, The National Centre of Excellence in Youth Mental Health. Soc Psychiatry Psychiatr Epidemiol 2016 Jan 23;51(1):1-13. [doi: 10.1007/s00127-015-1140-0] [Medline: 26498752]
23. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV). Washington, DC: American Psychiatric Association; 1994.
24. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987 Jan 01;13(2):261-276. [doi: 10.1093/schbul/13.2.261] [Medline: 3616518]
25. D'Zurilla TJ, Goldfried MR. Problem solving and behavior modification. J Abnorm Psychol 1971 Aug;78(1):107-126. [doi: 10.1037/h0031360] [Medline: 4938262]
26. Sugihara G. Nonlinear forecasting for the classification of natural time series. Phil. Trans. R. Soc. Lond. A 1997 Jan 15;348(1688):477-495. [doi: 10.1098/rsta.1994.0106]
27. Carroll JM. Human-computer interaction: psychology as a science of design. Annu Rev Psychol 1997 Feb;48(1):61-83. [doi: 10.1146/annurev.psych.48.1.61] [Medline: 15012476]
28. Gaver WW. Technology affordances. 1991 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 2, 1991; New Orleans, LA. [doi: 10.1145/108844.108856]
29. Sugihara G, Park J, Deyle E, Saberski E, Smith C, Ye H. Empirical Dynamic Modeling. The Comprehensive R Archive Network. 2022. URL: https://cran.r-project.org/web/packages/rEDM/vignettes/rEDM-tutorial.pdf [accessed 2023-05-01]
30. Park J, Smith C, Sugihara G, Deyle E, Saberski E, Ye H. Package 'rEDM'. The Comprehensive R Archive Network. 2023. URL: https://cran.r-project.org/web/packages/rEDM/rEDM.pdf [accessed 2023-05-01]
31. Ye H, Beamish RJ, Glaser SM, Grant SCH, Hsieh C, Richards LJ, et al. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proc Natl Acad Sci U S A 2015 Mar 31;112(13):E1569-E1576 [FREE Full text] [doi: 10.1073/pnas.1417063112] [Medline: 25733874]
32. Pan S, Duraisamy K. Physics-informed probabilistic learning of linear embeddings of nonlinear dynamics with guaranteed stability. SIAM J. Appl. Dyn. Syst 2020 Feb 12;19(1):480-509. [doi: 10.1137/19m1267246]
33. Takens F. Detecting strange attractors in turbulence. In: Rand D, Young LS, editors. Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics, vol 898. Berlin, Heidelberg: Springer; 1981:366-381.
34. Valentine L, McEnery C, O'Sullivan S, Gleeson J, Bendall S, Alvarez-Jimenez M. Young people's experience of a long-term social media-based intervention for first-episode psychosis: qualitative analysis. J Med Internet Res 2020 Jun 26;22(6):e17570 [FREE Full text] [doi: 10.2196/17570] [Medline: 32384056]
35. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries NK, Brug J. An exploration of factors related to dissemination of and exposure to internet-delivered behavior change interventions aimed at adults: a Delphi study approach. J Med Internet Res 2008 Apr 16;10(2):e10 [FREE Full text] [doi: 10.2196/jmir.956] [Medline: 18417443]
36. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]
37. Thorisdottir IE, Sigurvinsdottir R, Asgeirsdottir BB, Allegrante JP, Sigfusdottir ID. Active and passive social media use and symptoms of anxiety and depressed mood among Icelandic adolescents. Cyberpsychol Behav Soc Netw 2019 Aug;22(8):535-542. [doi: 10.1089/cyber.2019.0079] [Medline: 31361508]
38. Valentine L, McEnery C, O'Sullivan S, D'Alfonso S, Gleeson J, Bendall S, et al. Young people's experience of online therapy for first-episode psychosis: A qualitative study. Psychol Psychother 2022 Mar 12;95(1):155-172. [doi: 10.1111/papt.12356] [Medline: 34252267]
39. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial intelligence-assisted online social therapy for youth mental health. Front Psychol 2017 Jun 02;8:796 [FREE Full text] [doi: 10.3389/fpsyg.2017.00796] [Medline: 28626431]

Abbreviations

CCM: convergent cross mapping
EDM: empirical dynamic modeling
EPPIC: Early Psychosis Prevention and Intervention Centre
FEP: first-episode psychosis
HCI: human computer interaction
MCCM: multiple convergent cross mapping
MOST: moderated online social therapy
RCT: randomized controlled trial
SDT: self-determination theory
TAU: treatment as usual

Edited by J Torous; submitted 06.12.22; peer-reviewed by A Hudon, E Yom-Tov; comments to author 05.01.23; revised version received 02.03.23; accepted 01.04.23; published 22.05.23.

Please cite as:
O'Sullivan S, van Berkel N, Kostakos V, Schmaal L, D'Alfonso S, Valentine L, Bendall S, Nelson B, Gleeson JF, Alvarez-Jimenez M Understanding What Drives Long-term Engagement in Digital Mental Health Interventions: Secondary Causal Analysis of the Relationship Between Social Networking and Therapy Engagement
JMIR Ment Health 2023;10:e44812
URL: https://mental.jmir.org/2023/1/e44812
doi:10.2196/44812
PMID:37213197

Original Paper

Moderated Online Social Therapy for Carers of Early Psychosis Clients in Real-World Settings: Cluster Randomized Controlled Trial

John Gleeson[1], BA (Hons), MPsych, PhD; Ashleigh Lin[2], BSc (Hons), MPsych, PhD; Peter Koval[3], BA (Hons), MPhil, PhD; Liza Hopkins[4], BA (Hons), MA, PhD; Paul Denborough[4], MBBS, MMed; Reeva Lederman[5], PhD; Helen Herrman[6,7], BMedSc, MB BS, MD; Sarah Bendall[6,7], BA, PG Dip (Clin Psych), MA, PhD; Dina Eleftheriadis[6,7], BA; Sue Cotton[6,7], Grad Dip (App Sc), MAppSc, PhD; Yael Perry[2], BPsych (Hons), MPsych, PhD; Michael Kaess[8], MD; Mario Alvarez-Jimenez[6,7], BSc (Hons), MA, DClinPsy, PhD

[1]Healthy Brain and Mind Research Centre, School of Behavioural and Health Sciences, Australian Catholic University, Fitzroy, Australia

[2]University of Western Australia, Nedlands, Australia

[3]Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Australia

[4]Alfred Health, Melbourne, Australia

[5]School of Computing and Information Systems, Faculty of Engineering and Information Technology, The University of Melbourne, Parkville, Australia

[6]Orygen, Parkville, Australia

[7]Centre for Youth Mental Health, The University of Melbourne, Parkville, Australia

[8]University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland

Corresponding Author:
John Gleeson, BA (Hons), MPsych, PhD
Healthy Brain and Mind Research Centre
School of Behavioural and Health Sciences
Australian Catholic University
Level 5, Daniel Mannix Building
Young Street
Fitzroy, 3065
Australia
Phone: 61 39953 3108
Email: john.gleeson@acu.edu.au

Abstract

Background: Family carers of youth recovering from early psychosis experience significant stress; however, access to effective family interventions is poor. Digital interventions provide a promising solution.

Objective: Our objective was to evaluate across multiple Australian early psychosis services the effectiveness of a novel, web-based early psychosis intervention for carers.

Methods: In this cluster randomized controlled trial conducted across multiple Australian early psychosis services, our digital moderated online social therapy for carers (Altitudes) plus enhanced family treatment as usual (TAU) was compared with TAU alone on the primary outcome of perceived stress and secondary outcomes including mental health symptoms and family variables at the 6-month follow-up.

Results: Eighty-six caregivers were randomized and data were available for 74 young people in their care. Our primary hypothesis that carers randomized to Altitudes+TAU would report greater improvements in perceived stress at follow-up compared with carers randomized to TAU alone was not supported, with the TAU alone group showing more improvement. For secondary outcomes, the TAU alone group showed improved mindfulness over time. Regardless of group assignment, we observed improvements in satisfaction with life, quality of life, emotional overinvolvement, and burden of care. In contrast, hair cortisol concentration increased. Post hoc analyses revealed more contact with early psychosis services in the intervention group compared to TAU alone and that improvements in perceived stress and social support were associated with use of the intervention in the Altitudes+TAU group. In this study, 80% (12/15) reported a positive experience with Altitudes and 93% (14/15) would recommend it to others.

Conclusions: Our trial did not show a treatment effect for Altitudes in perceived stress. However, our post hoc analysis indicated that the amount of use of Altitudes related to improvements in stress and social support. Additional design work is indicated to continue users' engagement and to significantly improve outcomes in problem-solving, communication, and self-care.

Trial Registration: Australian and New Zealand Clinical Trials Registry ACTRN12617000942358; https://trialsearch.who.int/Trial2.aspx?TrialID=ACTRN12617000942358

(JMIR Ment Health 2023;10:e47722) doi:10.2196/47722

KEYWORDS

first-episode psychosis; carers; eHealth; families; stress; psychosis; digital mental health intervention; web-based therapy; social therapy

Introduction

Psychotic disorders most often develop during late adolescence. In most cases acute positive psychotic symptoms, namely, hallucinations, delusions, and thought disorder, are responsive to treatment [1]. However, the recovery phase is often accompanied by deficits in psychosocial functioning (ie, lack of engagement with social relationships and vocational or educational pathways) [2] and secondary mental health symptoms [3-5]. In more than 50% of cases, psychotic relapse occurs during the first 3 years after onset of a first episode of a psychotic disorder [6].

Formal treatments for early psychosis are provided within treatment systems that prioritize early detection and comprehensive biopsychosocial interventions, with antipsychotic medication as the cornerstone [7]. Psychosocial interventions focus on recovery of psychosocial functioning via cognitive behavior therapy, vocational rehabilitation, and cognitive remediation [7]. Given the age of onset, families are the mainstay of informal care in early psychosis. A warm family environment is an important protective factor during recovery [8]; however, the burden of caring for a family member with early psychosis is associated with elevated distress [9].

As highlighted by the World Health Organization (WHO), the impacts upon caregivers warrant the dissemination of these effective family interventions for the benefit of both carers and the young person experiencing psychosis [10]. Meta-analytic evidence supports the effectiveness of family interventions in reducing relapse rates in early psychosis as compared to treatment as usual (TAU) [11]. However, access to family interventions is poor [10], resulting in missed opportunities for prevention [12]. This study is aimed at addressing poor accessibility via the provision of a family intervention using digital technology.

Digital mental health interventions provide a promising solution for poor accessibility [13]. Developed in partnership with carers and based on our Moderated Online Social Therapy (MOST) framework [14-17], our Altitudes intervention integrates evidence-based psychoeducation, peer-to-peer social networking, and web-based moderation in a single digital application [18].

We recently evaluated Altitudes via a cluster randomized controlled trial (ACTRN12616000968471). We compared Altitudes plus specialist first-episode family TAU with specialist first-episode family TAU alone within a single flagship early psychosis program [19,20]. At the 6-month follow-up, carers in both groups significantly improved on the primary outcome of stress in addition to a range of secondary outcomes, including mental health symptoms, carer self-efficacy, and expressed emotion [20]. In addition, there were significantly fewer visits to emergency departments by patients with early psychosis from the Altitudes group [20]. However, the trial was conducted at the Early Psychosis Prevention and Intervention Centre—a long-established, flagship early psychosis program in Melbourne, Australia, embedded in an academic research program that may have resulted in enriched background treatments [21]. This left open the question of whether Altitudes could add significant benefit for carers attending real-world early psychosis programs. Testing the feasibility of providing Altitudes across multiple early psychosis sites based across different cities was also a critical additional step in assessing its scalability.

Our aim in this trial was to evaluate whether a digital intervention (Altitudes), which uses MOST, improved perceived stress at the 6-month follow-up in carers with a relative receiving treatment for early psychosis, when added to real-world early psychosis services (ACTRN12617000942358).

The primary hypothesis was that carers randomized to the Altitudes web-based application+enhanced family TAU would report significantly greater improvements in perceived stress at the 6-month follow-up compared with carers randomized to enhanced TAU alone. The secondary hypothesis was that carers randomized to Altitudes+TAU would experience reduced activity of the hypothalamic-pituitary adrenal (HPA) axis—one of the major stress response systems of the human body—as well as improved positive coping, self-efficacy, depression, and perceived social support compared with carers randomized to TAU alone at 6 months follow-up.

Methods

Design

The East-West Altitudes trial was a single-blinded, cluster randomized controlled trial with clusters comprising individual families. A cluster design, with family as the unit of randomization, was selected to ensure that members of the same family were randomized to the same condition. It would have been infeasible to have family members participating concurrently in different treatment groups. The 2 treatment conditions included Altitudes plus enhanced treatment as usual (TAU) and enhanced TAU alone. The assessment time points

were baseline (prior to randomization), 3 months (for the primary outcome only), and 6 months. The clinical sites were in Melbourne and Perth on Australia's southeast and west coasts, respectively.

Recruitment

Recruitment of the trial participants was undertaken at 13 sites across Melbourne and Perth between October 2018 and October 2019 including 1 clinical service in Melbourne (headspace early psychosis; hEP) and 2 in Perth (hEP and Western Australia Department of Health Early Intervention in Psychosis Services). Follow-up was completed in July 2020. The Melbourne hEP was managed by Alfred Health and consisted of a hub and spokes model across 5 sites (Bentleigh [hub], Elsternwick, Frankston, Dandenong, and Narre Warren). The Perth hEP consisted of the hub and spoke model of Joondalup, Osbourne Park, and Midland. The Western Australia Department of Health Early Intervention in Psychosis Services included coordinated sites, namely, Bentley, Fremantle, Peel, Rockingham, and Kwinana.

The research assistant (RA) met with potential participants face-to-face at the early psychosis service or in a location convenient to the carer. The RA presented the information regarding the study in writing and orally and invited prospective participants to ask any questions about any aspect of the study before obtaining signed informed consent.

Ethics Approval

Ethics approval for Victorian sites was granted by the Alfred Hospital Human Research Ethics Committee (project No 298/17) and by Western Australia Department of Health (PRN: RGS0000000416) and University of Western Australia (RA/4/20/4245) for Western Australian sites.

Participants and Setting

Eligible participants included carers who were aged ≥18 years (namely parents, grandparents, siblings, and partners) of young people who were currently receiving treatment at an early psychosis service (n=86). Up to 4 family members were eligible to participate from each family. For hEP services, eligibility criteria for the clients were (1) age 12-25 years; (2) a diagnosis of a first episode of psychotic disorder or at ultrahigh risk for developing psychosis [22]; and (3) no more than 12 months of continuous care upon presentation. For Western Australia Health, early psychosis services eligibility criteria for clients were (1) age between 16 and 40 years; (2) the client presented with early psychosis (no diagnosis required); (3) no more than 12 months of continual care upon presentation; and (4) duration of untreated psychosis ≤3 years.

Carers who did not have sufficient English to provide informed consent were excluded. We determined our target sample size based on an a priori power analysis, for which we assumed a moderate treatment effect (Cohen d=0.5) for our primary outcome of perceived stress at the 6-month follow-up [23]. Setting α at .05 (2-tailed), we determined that a sample size of 64 per group was required to achieve 80% power (G*Power Release; version 3.1.9.2; Heinrich Heine Univeritat Dusseldorf). Adjusting for the design effect (equal to 1.05), this equated to 68 per group or a total of 136 participants.

Interventions

Overview

The Altitudes+TAU condition involved participation in our digital application [18]. TAU comprised the usual array of services for carers at each service in addition to a psychoeducation booklet. Altitudes, powered by our MOST software framework [14], integrated within 1 web-based application: evidence-based psychoeducation, peer-to-peer social networking, and expert and peer web-based moderation. Each user could log on 24 hours per day for the duration of the trial.

Altitudes Interactive Psychoeducation

The web-based psychoeducation was developed to target carer stress. New users were given an introductory welcome to Altitudes which highlighted ways to optimize their use of the system and how to access system help. Users were invited to complete, at their own convenience, a series of 8 web-based modules (known as "pathways") which addressed themes of self-care, understanding psychosis, early warning signs and prevention of relapse, understanding their personal strengths as a carer, communicating with their relative, dealing with unhelpful thinking, self-compassion, and mindfulness. These pathways were divided into thematically-related psychoeducation "steps" to maximize the usability of the material. Each step was designed to be completed within 5-20 minutes (see Multimedia Appendix 1 for details for each step). The content of these steps comprised text, illustrations, and audio tracks and were designed to improve carer stress, for example, by encouraging self-care and by targeting problematic appraisals known to increase carer stress. In addition, the content of steps was influenced by social cognition concepts of "agency" and self-efficacy in family life [24]. The steps and pathways entailed regular prompts to users to share their reactions to material with other users through a series of "talking points." Users' responses populated the content of the social networking newsfeed. To facilitate the process of reflection on the content by participants, moderators could also contribute to the talking points. In addition, users could indicate their preference for material through "like" buttons, share content with other users, and keep track of which users had completed specific pathways and which users shared their specific personal strengths.

Altitudes Social Networking Features

The Altitudes social network enabled users to develop a web-based profile, communicate via posts with other users and web-based moderators, and comment on the web-based psychoeducation material. The application was hosted on a secure University of Melbourne web server. In addition, the web application included measures to secure the application and database against unauthorized access. Privacy and web-based safety were managed in accordance with the Online Social Networking Guidelines published by Cybersmart, a national cybersafety and cybersecurity education program managed by the Australian Communications and Media Authority.

Role of Moderators

In contrast to web-based self-help without human support, the guidance provided by moderators was integral to Altitudes. The model of web-based moderation was informed by the Supportive Accountability Framework of eHealth interventions, which emphasizes the importance of human support for engagement in eHealth systems [25].

Expert moderators were clinical psychologists with specialist family work experience. Their role was to optimize engagement, suggest specific content, facilitate joint problem-solving, and monitor safety daily. If the engagement was low, moderators would prompt participants via follow-up phone calls.

Peer moderators, with lived experience of caring for a relative with psychosis, modeled the use of the system and facilitated web-based interactions. At weekly supervision sessions with author JFMG, moderators reviewed progress and system engagement.

Enhanced Family TAU

All participating carers received a psychoeducation booklet that included information about psychosis and treatment, helpful contacts, as well as coping skills and suggestions for ways to assist and communicate with their relative. Additional carer services potentially included meetings with their relative's case manager, psychiatrist, family peer worker, or carer consultant, as well as access to carer support groups. These differed by service.

Measures

Primary Outcome

Perceived stress in carers over the preceding month was measured by the Perceived Stress Scale (PSS) [26]—a valid and reliable 10-item measure rated on a Likert scale ranging from 0 (never) to 4 (very often).

Secondary Outcomes

Hair cortisol is a biomarker of basal HPA axis activity [27]. The advantages of this measure are the validity as an index of long-term systemic cortisol levels, its reliability across repeated assessments, and its relative robustness to a range of potential confounding influences [28]. To assess potential changes of this biological stress response system, mean baseline HPA system activity during the last month was measured by a validated procedure for measuring hair cortisol [27]. Carer depressive symptoms were measured via the Centre for Epidemiological Studies Depression Scale—Revised [29]and substance use via the Alcohol, Smoking and Substance Involvement Screening Test [30]. Worry was measured via the Penn State Worry Questionnaire [31], loneliness via the UCLA (University of California, Los Angeles) Loneliness Scale [32], and social support via the Medical Outcomes Study Social Support Survey (MOS-SSS) [33]. The Me as a Parent Questionnaire was used to measure parental self-efficacy [34], coping was assessed via the Ways of Coping Scale [35,36], and personal strengths use via the Strengths Use Scale [37]. Self-compassion was measured via the Self-Compassion Scale Short Form [38] and mindfulness using the Mindful Attention Awareness Scale [39]. Satisfaction with life was assessed by the Satisfaction With Life Scale (SWLS) [40] and emotional, psychological, and social well-being via the Mental Health Continuum Short Form (MHC-SF) [41]. Quality of life (QoL) was measured by the Assessment of Quality of Life-8 dimensions [42].

Patient and Family Characteristics

Carer demographic variables included age, living situation, years of education completed, employment and marital status, country of birth, and source of income. Relevant family-level variables were measured including expressed emotion measured via the Family Questionnaire [43], and the degree of openness and extent of problems in family communication using the Parent-Adolescent Communication (PAC) scale [44]. Carer burden was assessed via the Experience of Care-giving Inventory [45]. The demographic characteristics of the young person were collected from carers.

A Resource Use Questionnaire was used to determine resource and treatment use by carers and young people. The use of services provided by the early psychosis service was documented via a self-report survey.

Altitude-Specific Measures

The use of Altitudes was continuously monitored across the study intervention period via frequency of log-ons. At 6 months, Altitudes users completed a self-report measure of their perception of Altitudes moderation [46] and a self-report usability measure [47,48].

Procedure

RAs attended clinical team meetings across sites to promote the study. Participating carers were asked if their young relative could be approached to seek their consent to access data from their medical record. The young person was contacted by the study RA. The RAs undertaking the follow-up assessments were kept blind to treatment allocation. The participants were not blinded to their treatment allocation, and they were aware that Altitudes was the intervention of interest.

Randomization occurred after each baseline assessment. An independent statistician created the randomization sequence, which included permutated blocks. The block sizes and randomization sequence were concealed from the study coordinator, RAs, and investigators. The study coordinator randomized the family via a secure web-based clinical trials management system with stratification by state (Victoria or Western Australia). The system generated an email to the RA who telephoned the participant to let the family know which group they had been allocated to.

The primary and secondary outcomes were measured prior to randomization and repeated at the 6-month follow-up. The PSS was completed at 3 months using a telephone-administered version.

To enable analysis of hair cortisol, a single hair sample (at least 3 cm long, approximately 0.5 cm in diameter) was taken from a posterior vertex region on the head and stored at an ambient temperature. Screening prior to collection of the hair sample was used to determine factors that may affect the analysis of

cortisol, such as hair products used. The hair cortisol analysis procedures included repeated washing of hair samples with isopropanol, drying, weighing (weight: 7.5, SD 0.5 mg), steroid extraction with methanol, evaporation of methanol (at 50 °C), and cortisol determination [27]. This analysis provided a total pg/mg value of cortisol in the hair corresponding to approximately one month prior to sampling.

Statistical Analysis

Statistical analyses were conducted using Stata (version 16.1; StataCorp). Statistical analyses are presented in terms of the CONSORT (Consolidated Standards of Reporting Trials) extension pertaining to cluster analyses [49] and the International Conference on Harmonization Topic E9 Statistical Principals for Clinical Trials. Descriptive statistics were used to describe the total cohort as well as separately for the Altitudes+TAU and TAU alone groups. Baseline differences between the groups were not analyzed using inferential statistics as recommended in CONSORT guidelines [50]. Because of the randomization process, any baseline difference between 2 treatment groups is the result of chance and not due to external factors impacting treatment allocation [51]. Therefore, it is considered to be absurd to test for such differences and there have been calls for such practices to be eliminated [52]. Inferential statistics were used to compare baseline differences between caregivers who did and did not have follow-up data. For intent-to-treat analysis (ITT), all cases were included in the analyses, regardless of whether they had follow-up data.

For the primary outcome variable, the PSS, between group differences were examined using mixed effects repeated measures (MMRM) models [53]. In MMRM models, all observed information is used to derive the models (including estimation but not imputation of missing data). They are considered the preferred method for ITT analyses in clinical trials [54]. For these models, individual timepoint measures are considered nested within individual carers, who may be considered nested within families. For the main analysis, the parameters included group, time (includes baseline, 3 and 6 months), and the group x time interaction. All cases with at least 1 observation were included in the ITT. The default independence covariance structure was modeled in these models. As per protocol, analyses were also conducted with participants who had completed the intervention induction and had follow-up data. For resource use data, group differences were examined using either chi-square or Fisher exact test. Simple Pearson correlations were calculated for the association between parameters of intervention usage and outcomes in the Altitudes group.

Results

Participant Characteristics

There were 86 caregivers that participated in the study, with most being female and a mother of the young person (see Table 1). Caregivers ranged in age from 18 to 76 years (mean 51.2, SD 9.9 years). Most of the cohort were in married or de facto relationships and lived in their own house or flat. Just over 50% (n=46) of caregivers were born in Australia. Of those born overseas, 37% (n=15) were from European countries. Just over three-quarters of caregivers had completed their secondary school education and most had full-time employment. There were 46 caregivers from Victoria and 40 from Western Australia.

In the study, there were 74 young people. For 85% (n=63) of the young people, there was 1 caregiver in the study, 13% (n=10) had 2 caregivers and 1 young person had 3 caregivers. The mean cluster size was therefore 1.16 (SD 0.41). The characteristics of the young people are detailed in Table 2.

Table 1. Baseline demographic characteristics of the total cohort of caregivers (N=86) as well as separately for the Altitudes plus treatment as usual and only treatment as usual groups.

Variable	Total cohort (N=8)	Altitudes (n=43)	TAU[a] (n=43)
Female (gender), n (%)	71 (83)	35 (81)	36 (84)
Age (years), mean (SD)	51.2 (9.9)	51.6 (10.8)	50.8 (9.2)
Relationship to the young person, n (%)			
Mother	65 (76)	31 (72)	34 (79)
Father	13 (15)	8 (19)	5 (12)
Other	8 (9)	4 (9)	4 (9)
Married or de facto married, n (%)	59 (69)	35 (81)	24 (57)
Children (n), mean (SD)	2.7 (1.6)	2.7 (1.5)	2.8 (1.6)
Accommodation, n (%)			
Rented flat or room	19 (22)	8 (19)	11 (26)
Own flat or house	61 (71)	31 (72)	30 (70)
Other	6 (7)	4 (9)	2 (5)
Born in Australia, n (%)	46 (53)	26 (60)	20 (46)
Born outside Australia, n (%)	40 (46)	17 (39)	23 (53)
Oceania and Antarctica	7 (17)	3 (18)	4 (17)
Europe	15 (37)	6 (35)	9 (39)
North Africa and Middle East	2 (5)	2 (12)	0 (0)
Asia	5 (12)	2 (12)	3 (13)
Sub-Saharan Africa	11 (27)	4 (23)	7 (30)
English main language spoken, n (%)	81 (94)	40 (93)	41 (95)
Command of English, n (%)			
Good	10 (12)	6 (14)	4 (9)
Native speaker	76 (88)	37 (86)	39 (91)
Highest level of education year 12, n (%)	65 (76)	31 (72)	34 (81)
Additional qualifications, n (%)			
No further education	10 (12)	6 (14)	4 (9)
Trade or technical training	31 (36)	16 (37)	15 (36)
Tertiary degree	26 (31)	12 (28)	14 (33)
Postgraduate degree	18 (21)	9 (21)	9 (21)
Currently in paid work, n (%)	69 (80)	31 (72)	38 (88)
Full-time work	37 (54)	12 (39)	25 (66)
Annual income, n (%)			
Aus $20,799[b] or less	16 (19)	12 (29)	4 (10)
Aus $28,000-51,999	23 (28)	14 (33)	9 (22)
Aus $52,000-77,999	20 (24)	7 (17)	13 (32)
Aus $78,000-103,999	11 (13)	2 (5)	9 (22)
Aus $104,000 or more	13 (16)	7 (17)	6 (15)

[a]TAU: treatment as usual.

[b]At time of study commencement on October 1, 2018, the conversation rate was Aus $1=US $0.7200.

Table 2. Demographic characteristics of the young people receiving treatment at an early psychosis service (N=86) with Altitudes plus treatment as usual or only treatment as usual.

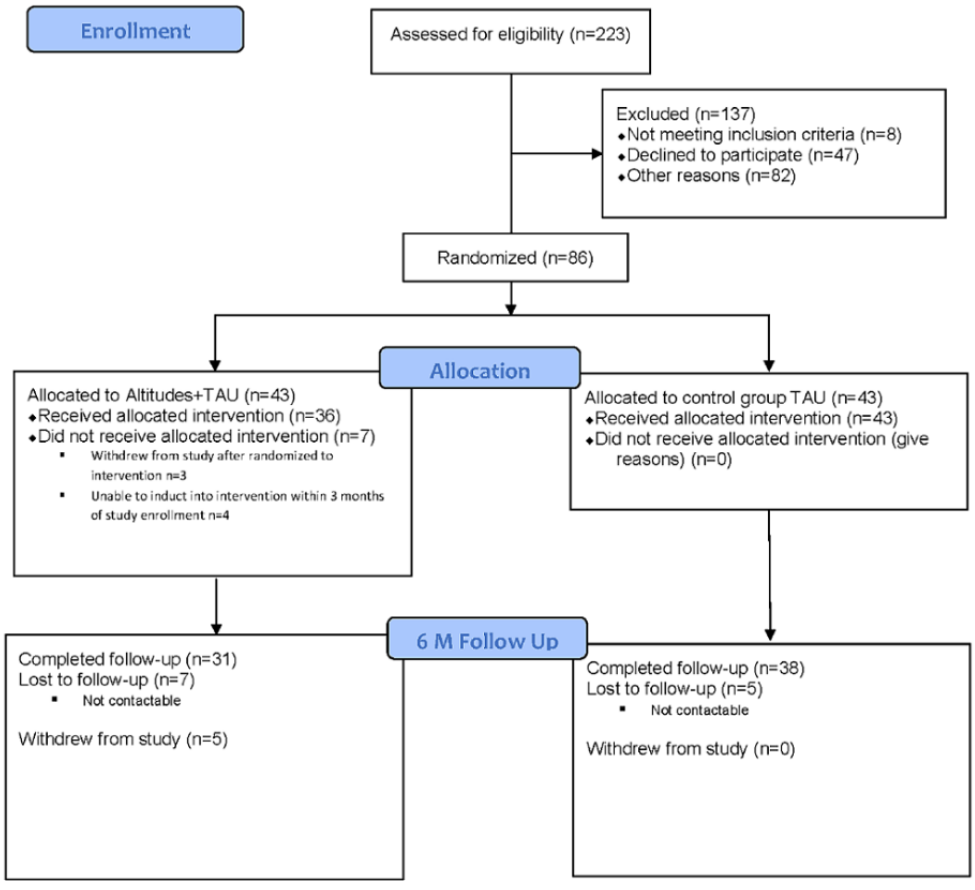
Variable	Total cohort (N=74)	Altitudes (n=36)	TAU[a] (n=38)
Female (gender), n (%)	28 (38)	10 (28)	18 (47)
Age (years), mean (SD)	21.1 (3.5)	20.9 (3.8)	21.3 (3.3)
Accommodation, n (%)			
House with family of origin	60 (81)	29 (81)	31 (82)
Born in Australia, n (%)	56 (76)	29 (81)	27 (71)
English main language spoken, n (%)	71 (96)	35 (97)	36 (95)
Studying status, n (%)			
Not studying	45 (61)	23 (64)	22 (58)
Studying part-time	13 (18)	8 (22)	5 (13)
Studying full-time	16 (22)	5 (14)	11 (29)
Highest level of education year 12, n (%)	39 (53)	16 (44)	23 (62)
Currently in paid work, n (%)	22 (30)	13 (36)	9 (24)
Full-time work	6 (27)	3 (23)	3 (33)

[a]TAU: treatment as usual.

Participant Flow

Figure 1 shows a CONSORT diagram of the participant flow through the study. Sixty-nine caregivers had data at follow-up, indicating that missing data at follow-up were 20% (n=17). The percentage of caregivers in TAU alone that had follow-up data (n=38, 88%) did not significantly differ from caregivers in Altitudes+TAU (n=31, 72%; χ^2_1=3.6; *P*=.06). Those who did not have follow-up data were more likely to have been born in Australia (χ^2_1=4.5; *P*=.03). No other differences were found between the groups with respect to caregiver variables.

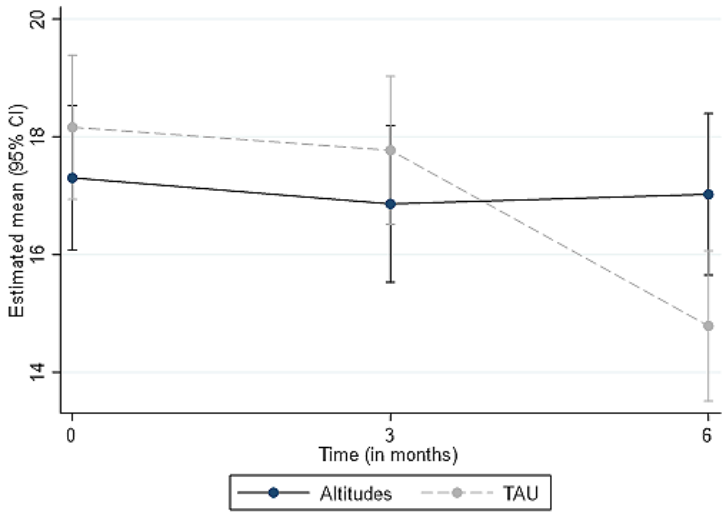
Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. TAU: treatment as usual.



Primary Outcome—Perceived Stress

Figure 2 displays the differences between the groups with respect to PSS across baseline, at 3 and 6 months. The variance associated with family cluster was close to 0, therefore, the models were rerun with the caregiver as the unit of analysis. For the PSS, there was a significant interaction between group and time (z=−2.14; P=.03), with TAU alone showing more change (ie, improvement in PSS) from baseline to follow-up, particularly between 3 and 6 months.

Figure 2. Estimated mean (95% CI) from the MMRMs for the 2 carer groups across the 3 points on the PSS. MMRMs: mixed effects repeated measures; PSS: Perceived Stress Scale; TAU: treatment as usual.



Secondary Outcomes

Table 3 comprises the adjusted means and SEs for the 2 carer groups on the secondary outcome measures. There was a significant interaction between group and time for mindfulness (z=2.04, P=.04), with the TAU alone group showing an increase over time (see Table 3). There was also a significant interaction between group and time for carer burden associated with problems with services, with the TAU alone group experiencing significantly less burden in this domain, whereas Altitudes+TAU group remained relatively stable over time (z=−2.10; P=.04; see Table 4).

With respect to changes over time regardless of group, there were significant improvements in satisfaction with life (z=2.02; P=.04; higher scores were indicative of greater satisfaction), and QoL (z=−2.08; P=.04; lower scores indicated better QoL). Both groups also improved over time with respect to emotional overinvolvement (z=−2.52; P=.01) and dependency (z=−2.79; P=.01).

Table 3. Estimated marginal mean (SE) for baseline and follow-up from MMRMS for secondary outcome measures for Altitudes plus treatment as usual or only treatment as usual.

Variable	Pre, mean (SE)		Post (6 months), mean (SE)		Interaction	
	Altitudes (n=43)	TAU[a] (n=43)	Altitudes (n=31)	TAU (n=38)	Estimate, β (SE)	*P* value
Secondary outcomes						
UCLA[b] Loneliness Scale-3	25.4 (0.8)	26.9 (0.8)	26.6 (0.9)	26.3 (0.9)	−1.66 (−4.27 to 0.95)	.21
CESD-R[c]	12.8 (1.8)	12.4 (1.8)	14.8 (2.0)	12.5 (1.9)	−1.83 (−6.32 to 2.64)	.42
SUS[d]	59.1 (1.8)	63.2 (1.8)	61.1 (2.0)	64.1 (1.9)	−1.04 (−5.75 to 3.66)	.66
SCS-SF[e]	3.3 (0.1)	3.3 (0.1)	3.4 (0.1)	3.6 (0.1)	.22 (−0.05 to 0.50)	.11
SWLS[f]	19.0 (1.1)	22.1 (1.1)	21.3 (1.2)	22.6 (1.1)	−1.66 (−4.59 to 1.28)	.27
PSWQ[g]	50.9 (2.0)	50.2 (2.0)	48.9 (2.2)	46.9 (2.0)	−1.35 (−5.85 to 3.15)	.56
MaaP[h]	50.1 (1.1)	49.8 (1.1)	49.7 (1.3)	52.2 (1.2)	2.79 (−0.89 to 6.47)	.14
Mindful Attention Awareness Scale	4.0 (0.1)	3.9 (0.1)	3.9 (0.2)	4.2 (0.1)	.36 (0.01 to 0.71)	.04
MOS-SSS[i]	70.8 (2.9)	71.0 (2.9)	72.6 (3.2)	74.8 (3.0)	2.02 (−4.26 to 8.31)	.53
MHC-SF[j]	46.2 (2.1)	45.7 (2.1)	45.6 (2.4)	46.0 (2.2)	.89 (−4.37 to 6.16)	.74
Ways of Coping Scale						
Confrontive coping	0.9 (0.1)	0.9 (0.1)	1.0 (0.1)	1.0 (0.1)	−.03 (−0.30 to 0.23)	.80
Self-controlling (sc)	1.2 (0.1)	1.4 (0.1)	1.2 (0.1)	1.3 (0.1)	−.07 (−0.34 to 0.21)	.64
Self-controlling (ss)	1.2 (0.1)	1.3 (0.1)	1.3 (0.1)	1.2 (0.1)	−.23 (−0.61 to 0.14)	.22
Distancing	0.8 (0.1)	1.1 (0.1)	0.8 (0.1)	0.9 (0.1)	−.13 (−0.37 to 0.10)	.26
Escape avoidance	0.7 (0.1)	0.8 (0.1)	0.8 (0.1)	0.7 (0.1)	−.23 (−0.49 to 0.02)	.07
Planful problem-solving	1.3 (0.1)	1.5 (0.1)	1.3 (0.1)	1.6 (0.1)	0 (−0.33 to 0.34)	.98
Alcohol, Smoking and Substance Involvement Screening Test						
Tobacco	3.1 (1.1)	4.1 (1.1)	2.6 (1.2)	4.9 (1.2)	1.37 (−0.24 to 2.98)	.10
Alcohol	7.5 (1.0)	6.5 (1.0)	7.9 (1.0)	6.4 (1.0)	−.56 (−2.68 to 1.56)	.60
Cannabis	0.6 (0.3)	0.4 (0.3)	0.6 (0.3)	0.1 (0.3)	−.24 (−0.72 to 0.23)	.32
AQoL[k]	70.4 (1.6)	73.4 (1.6)	73.3 (1.7)	75.5 (1.6)	−.77 (−4.41 to 25.86)	.68

[a]TAU: treatment as usual.

[b]UCLA: University of California, Los Angeles, Loneliness Scale (total 20-80).

[c]CESD-R: Centre for Epidemiological Studies Depression Scale—Revised (range 0-60).

[d]SUS: Strengths Use Scale.

[e]SCS-SF: Self-Compassion Short Form (1-5).

[f]SWLS: Satisfaction With Life Scale (5-35).

[g]PSWQ: Penn State Worry Questionnaire (16-80).

[h]MaaP: Me as a Parent Questionnaire (subscales 4-20; total 16-80).

[i]MOS-SSS: Medical Outcomes Study Social Support Survey (0-100).

[j]MHC-SF: Mental Health Continuum Short Form (0-70).

[k]AQoL: Assessment of Quality of Life (0-100).

Table 4. Estimated marginal means (SE) for baseline and follow-up from MMRM for the Family Questionnaire and Experience of Caregiving Scale for Altitudes plus treatment as usual or only treatment as usual.

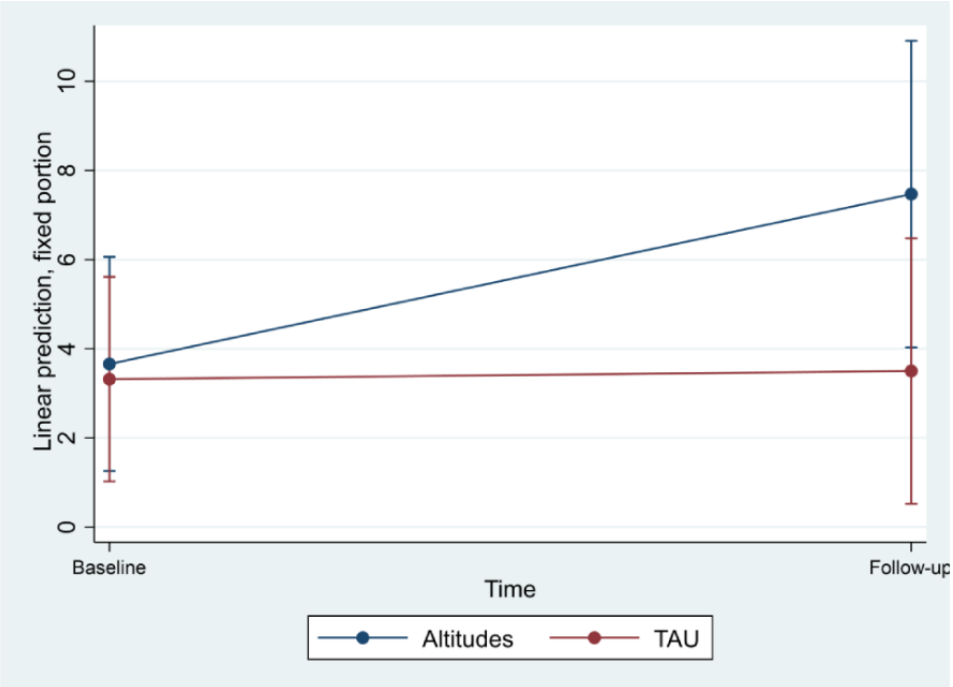
Variable	Pre, mean (SE)		Post (6 months), mean (SE)		Interaction	
	Altitudes (n=43)	TAU[a] (n=43)	Altitudes (n=31)	TAU (n=38)	Estimate, β (SE)	*P* value
Family questionnaire						
Critical comments	13.8 (1.0)	12.3 (1.0)	12.0 (1.1)	11.2 (1.1)	.68 (−1.80 to 3.16)	.59
Emotional overinvolvement	17.5 (0.9)	17.7 (0.9)	15.5 (1.0)	15.3 (1.0)	−.45 (−2.45 to 1.54)	.66
Experience of caregiving						
Difficult behaviors	14.5 (1.2)	13.7 (1.2)	13.0 (1.4)	11.1 (1.3)	−1.21 (−4.06 to 2.28)	.50
Negative symptoms	13.5 (1.0)	14.3 (1.0)	12.4 (1.1)	12.2 (1.1)	−.93 (−3.76 to 1.91)	.52
Stigma	7.8 (0.7)	7.8 (0.7)	6.7 (0.8)	5.4 (0.7)	−1.27 (−3.16 to 0.62)	.19
Problem with services	10.1 (1.0)	11.6 (1.0)	9.4 (1.1)	8.2 (1.0)	−2.76 (−5.33 to −0.19)	.04
Effects on family	11.7 (0.9)	11.7 (0.9)	11.5 (1.0)	10.6 (1.0)	−.96 (−3.25 to 1.34)	.42
Need to back up	11.8 (0.8)	10.2 (0.8)	10.9 (0.8)	9.2 (0.8)	−.05 (−1.91 to 1.81)	.96
Dependency	11.1 (0.7)	11.3 (0.7)	9.6 (0.8)	9.6 (0.7)	−.08 (−1.55 to 1.39)	.92
Loss	12.0 (0.9)	13.1 (0.9)	11.2 (1.0)	11.2 (0.9)	−1.21 (−3.46 to 1.04)	.29
Positive personal experiences	16.5 (1.0)	17.6 (1.0)	17.1 (1.1)	17.5 (1.0)	−.74 (−3.14 to 1.67)	.55
Good aspect of the relationship	13.2 (0.7)	13.8 (0.7)	12.7 (0.7)	13.4 (0.7)	.09 (−1.70 to 1.88)	.92

[a]TAU: treatment as usual.

Hair Cortisol

At baseline 66 caregivers had hair cortisol data and 34 had data at follow-up. One case had very high cortisol levels with values >145 pg/mg of hair cortisol both at baseline and at follow-up. This case was an extreme outlier and distorted the estimated means, therefore was excluded from the analyses. Figure 3 details the estimated means (with 95% CI) for the hair cortisol data. The interaction between group and time was not significant (z=−1.46; P=.14); however, the time main effect was significant (z=2.04; P=.04), demonstrating a moderate increase in cortisol concentration overall and regardless of intervention. The analyses were rerun with age as a covariate, and these findings were upheld. There was no relationship between change scores on hair cortisol and the PSS (r=0.07; P=.70).

Figure 3. Estimated marginal mean (95% CI) for baseline and follow-up from MMRM for hair cortisol. MMRM: mixed effects repeated measure; TAU: treatment as usual.



Resource Use

As shown in Table 4 there was a significant difference between the caregiver groups with respect to early psychosis services used by the caregiver with the Altitudes group having significantly higher use of services (Fisher exact test; *P*=.01). There were no differences between the caregiver groups with respect to medication use, including use of psychotropic medications over the previous 6 months.

Data were available on resource usage for 57 young people at follow-up (see Table 5). There were no differences between groups with respect to young people's use of emergency services and hospitalizations in the previous 6 months (see Table 5).

Table 5. Resource usage data in caregivers and young people.

	Altitudes, n (%)	Controls, n (%)	*P* value
Caregivers[a]			
Use of early psychosis service over past 6 months	28 (93)	25 (66)	.01
Services from early psychosis service contact			
Case managers	23 (77)	22 (58)	.08
Psychiatrists	13 (43)	13 (34)	.30
Family peer worker (lived experience)	5 (17)	10 (26)	.26
Family worker or carer consultant	2 (7)	5 (13)	.32
Written psychoeducation	11 (37)	11 (29)	.34
Community or private health services	30 (100)	36 (95)	.31
Any medications	28 (93)	32 (84)	.22
Psychotropic medications	10 (33)	12 (32)	.54
Young person[b]			
Emergency presentation over past 6 months	6 (26)	5 (15)	.23
Hospitalization over past 6 months	4 (17)	6 (18)	.63

[a]For Altitudes and control groups: n=30 and n=38, respectively.

[b]For Altitudes and control groups: n=23 and n=34, respectively.

Per Protocol Analyses

There were 7 caregivers in the Altitudes+TAU treatment arm who did not receive the Altitudes intervention. Three of these cases never received the intervention or had follow-up data. One case was lost to follow-up and did not complete the intervention induction and had no follow-up data. Three cases could not be contacted, and the intervention induction was not started but they did have follow-up data. These 7 cases were excluded from per protocol analyses. There were some differences between ITT and per protocol analyses. With per protocol analyses, the interaction for PSS over the 3 time points was no longer significant (z=−1.73; P=.08); however, parental self-efficacy was significant, that is, there was a greater increase in the TAU alone group (z=2.01; P=.04). The interactions between group and time were no longer significant for mindfulness (z=1.86; P=.06) and carer burden related to problems with services (z=−1.63; P=.10).

The per protocol analyses indicated that the Altitudes+TAU group (n=22, 81%) were more likely to be in contact with the young person's case manager than the TAU alone group (n=22, 58%; χ^2_1=4; P=.04).

Altitudes Specific Measures and Outcomes

The median duration of engagement with Altitudes was 7.5 (range 1-36) weeks with median number of logins 13.5 (range 1-96); 39% (n=14) remained active for more than 11 weeks. Two participants logged on only once. The number of newsfeed posts ranged from 0 to 31 with a mean of 3.56 (SD 6.19). In relation to the carers' perceived support for autonomy needs from web-based moderators (n=16), the mean was 5.5 (SD 1.38) on the 7-point scale. Our post hoc analysis of the relationship between activity in Altitudes and outcomes revealed that the total number of logins (r=0.49; P=.01) and the weeks of logins (r=0.47; P=.01) were both moderately and significantly correlated with improvement in PSS scores from baseline to 6 months. The number of logins (r=−0.46; P=.02) and the weeks of logins (r=−0.42; P=.03) were both moderately and negatively correlated with change from baseline to 6 months on the MOS-SSS. In relation to usability data (n=15), 80% (n=12) reported a positive and constructive experience with Altitudes and 93% (n=14) would recommend it to others.

There were no known privacy breaches or outages of Altitudes during the trial and there were no serious adverse events.

Discussion

Principal Results

The primary hypothesis that carers randomized to Altitudes plus family TAU would report significantly greater reductions in stress at the 6-month follow-up compared to TAU alone was not supported, with no significant change over time in

self-reported severity of carer stress in the Altitudes+TAU group. The secondary hypothesis, that carers randomized to Altitudes+TAU would experience reduced hair cortisol, improved positive coping, self-efficacy, depression, and perceived social support compared with carers randomized to TAU alone at the 6-month follow-up, was also not supported. There was no improvement over 6 months on hair cortisol, mental health symptoms, loneliness, social support, self-efficacy, coping, personal strengths use, and self-compassion in the Altitudes+TAU group compared to TAU alone. There was an effect for time on stress and mindfulness with group interactions favoring the TAU alone group which was not upheld in the per protocol analysis excluding participants that did not use Altitudes. In addition, there was a group by time interaction for problems with services with TAU alone carers experiencing less burden over time which was not upheld in the per protocol analysis. There was an improvement over time, regardless of group, in satisfaction with life, QoL, emotional overinvolvement, and burden of care associated with dependency. The per-protocol analysis indicated a significant group by time interaction for parental self-efficacy with an increase for the TAU alone group.

There was no suggestion that the findings could be accounted for by the Altitudes+TAU group receiving less support outside of the intervention compared with the TAU alone group. Conversely, the Altitudes+TAU group had a higher use of services and findings from the per protocol analysis showed that the Altitudes+TAU group was more likely to be in contact with the young person's case manager.

Comparison With Prior Work

We have previously shown that stress appraisal improves in first-episode psychosis family caregivers through a face-to-face first-episode psychosis family intervention [55] and via bibliotherapy [56]. The current findings suggest that when added to an enhanced TAU in standard early psychosis services Altitudes did not confer benefits in relation to reducing stress, other mental health symptoms, or improving family related outcomes such as expressed emotion and perceived burden of caregiving over a the 6-month period. However, our post hoc analysis revealed significant correlations between log-ins and improvement in perceived stress and social support within the Altitudes+TAU group suggesting that with adequate engagement Altitudes may confer benefits to carers.

The stage of recovery in young patients with early psychosis may be important in the interpretation of the current findings. The greater level of contact with services and a higher level of perceived problems with services in the Altitudes+TAU group suggests that the young people in their care were experiencing more complex recoveries from the acute phase of early psychosis treatments compared to the TAU alone group. It is also possible that Altitudes led to contact with mental health services for young people. Consistent with the former interpretation, more use of Altitudes was associated with decreased stress in carers and there were no outcomes favoring TAU in our initial trial in a flagship service [20].

The median duration of engagement in Altitudes was 11 weeks (approximately 77 days) compared with 119 days in our initial trial. Given the moderate correlation between log-ins and reduced stress, it is possible that this lower level of engagement may account for the overall lack of treatment effect.

The current findings leave open the question of how the effective components of carer interventions can be translated into an accessible and effective digital version. It is probable that the mechanisms of change in effective carer interventions were not adequately targeted through Altitudes. We developed Altitudes with text- and audio-based psychoeducation; however, the format may not have adequately engaged the target skills. Educational instructional design frameworks for a digital mode may be required to ensure that continued learning outcomes are achieved [57,58]. Compared with typical behavioral family interventions [59], the duration of engagement with Altitudes was relatively brief. Continued skill development may require longer and more intensive engagement with monitoring and structured feedback on skill acquisition. Recently, bibliotherapy for early psychosis caregivers based on problem-solving, adapted for delivery via smartphone, has shown superior outcomes on carer burden, caregiving experiences, and problem-solving at the 6-month follow-up compared with either a psychoeducation family group or usual care [60]. Importantly, this intervention included repeated structured practice of problem-solving skills. Other recent findings support the incorporation of acceptance and commitment therapy principles and practice for the amelioration of burnout symptoms for early psychosis caregivers which may also have benefits for perceived stress [60].

Carers in this study overall showed improvements over time in satisfaction with life, QoL, and emotional overinvolvement. However, the trends in the proportion reporting high levels of stress and being prescribed psychotropics were notable along with increased cortisol levels potentially highlighting continued activation of the stress response system in these carers' lives.

Strengths and Limitations

We demonstrated that it is feasible to conduct Altitudes across sites in real-world early psychosis settings and maintain a supportive web-based environment providing highly specialized support. There were multiple measures of stress with a robust study design. The major limitation of the current study was the smaller sample size than planned, which was a consequence of a lower rate of recruitment than projected due to the unexpected complexity of recruiting carers into a research study from hub and spoke service structures. This resulted in an underpowered study which makes findings more difficult to interpret [61]. Additional limitations were the limited data regarding the young people's mental health, which reduced the confidence in the equivalency of participant characteristics at baseline, and the exclusion of non-English speaking carer participants. Young people may have differed across groups and it was difficult to fully characterize the use of informal supports across services.

Conclusions

This study comprised the first randomized controlled trial to investigate the effectiveness of a MOST intervention for early psychosis carers recruited from multiple real-world early psychosis programs. Altitudes was based upon our MOST which was associated with reduced emergency department visits for

young people recovering from early psychosis [15]. While we did not find specific treatment effects, there were significant improvements across both treatment conditions on emotional overinvolvement, QoL, satisfaction with life, and aspects of caregiver burden and greater usage of Altitudes among carers randomized to the treatment group was associated with improvements in stress and social support. It was evident that early psychosis carers outside of flagship programs experienced levels of stress that warrant further innovations. We are currently modifying the Altitudes content to better engage mechanisms of change.

As early psychosis programs expand globally, the goal of delivering effective and accessible family interventions for carers remains an important research and clinical priority. The solid empirical foundations of effective family interventions for psychosis will continue to provide a source of inspiration for this important unmet global need.

Acknowledgments

This trial was funded by the Australian Catholic University Research Fund. We acknowledge Black Swan Health and Youth Focus who are lead agencies of the Western Australia headspace early psychosis sites. MAJ was supported by an investigator grant (APP1177235) from the National Health and Medical Research Council and a Dame Kate Campbell Fellowship from the University of Melbourne.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Altitudes Steps.
[XLSX File (Microsoft Excel File), 21 KB - mental_v10i1e47722_app1.xlsx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2859 KB - mental_v10i1e47722_app2.pdf]

References

1. Emsley R, Oosthuizen PP, Kidd M, Koen L, Niehaus DJH, Turner HJ. Remission in first-episode psychosis: predictor variables and symptom improvement patterns. J Clin Psychiatry 2006 Nov;67(11):1707-1712. [doi: 10.4088/jcp.v67n1106] [Medline: 17196049]
2. Santesteban-Echarri O, Paino M, Rice S, González-Blanch C, McGorry P, Gleeson J, et al. Predictors of functional recovery in first-episode psychosis: A systematic review and meta-analysis of longitudinal studies. Clin Psychol Rev 2017 Dec;58:59-75. [doi: 10.1016/j.cpr.2017.09.007] [Medline: 29042139]
3. Wilson RS, Yung AR, Morrison AP. Comorbidity rates of depression and anxiety in first episode psychosis: A systematic review and meta-analysis. Schizophr Res 2020 Feb;216:322-329. [doi: 10.1016/j.schres.2019.11.035] [Medline: 31791816]
4. Archie S, Rush BR, Akhtar-Danesh N, Norman R, Malla A, Roy P, et al. Substance use and abuse in first-episode psychosis: prevalence before and after early intervention. Schizophr Bull 2007 Nov;33(6):1354-1363 [FREE Full text] [doi: 10.1093/schbul/sbm011] [Medline: 17337748]
5. Sin GL, Abdin E, Lee J, Poon LY, Verma S, Chong SA. Prevalence of post-traumatic stress disorder in first-episode psychosis. Early Interv Psychiatry 2010 Nov;4(4):299-304. [doi: 10.1111/j.1751-7893.2010.00199.x] [Medline: 20977686]
6. Alvarez-Jimenez M, Priede A, Hetrick SE, Bendall S, Killackey E, Parker AG, et al. Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies. Schizophr Res 2012 Aug;139(1-3):116-128. [doi: 10.1016/j.schres.2012.05.007] [Medline: 22658527]
7. Royal AustralianNew Zealand College of Psychiatrists Clinical Practice Guidelines Team for the Treatment of SchizophreniaRelated Disorders. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the treatment of schizophrenia and related disorders. Aust N Z J Psychiatry 2005;39(1-2):1-30. [doi: 10.1080/j.1440-1614.2005.01516.x] [Medline: 15660702]
8. Lee G, Barrowclough C, Lobban F. Positive affect in the family environment protects against relapse in first-episode psychosis. Soc Psychiatry Psychiatr Epidemiol 2014 Mar;49(3):367-376. [doi: 10.1007/s00127-013-0768-x] [Medline: 24081324]
9. Jansen JE, Gleeson J, Cotton S. Towards a better understanding of caregiver distress in early psychosis: a systematic review of the psychological factors involved. Clin Psychol Rev 2015 Feb;35:56-66. [doi: 10.1016/j.cpr.2014.12.002] [Medline: 25531423]
10. Eckardt JP. Barriers to WHO Mental Health Action Plan updates to expand family and caregiver involvement in mental healthcare. Gen Psychiatr 2022;35(2):e100784 [FREE Full text] [doi: 10.1136/gpsych-2022-100784] [Medline: 35372788]

11. Camacho-Gomez M, Castellvi P. Effectiveness of family intervention for preventing relapse in first-episode psychosis until 24 months of follow-up: A systematic review with meta-analysis of randomized controlled trials. Schizophr Bull 2020 Jan 04;46(1):98-109 [FREE Full text] [doi: 10.1093/schbul/sbz038] [Medline: 31050757]
12. Eassom E, Giacco D, Dirik A, Priebe S. Implementing family involvement in the treatment of patients with psychosis: a systematic review of facilitating and hindering factors. BMJ Open 2014 Oct 03;4(10):e006108 [FREE Full text] [doi: 10.1136/bmjopen-2014-006108] [Medline: 25280809]
13. Onwumere J, Kuipers E. Does the internet have a role in helping families of people with psychosis? Psychiatr Serv 2017 Apr 01;68(4):419-420. [doi: 10.1176/appi.ps.68401] [Medline: 28366125]
14. Lederman R, Wadley G, Gleeson J, Bendall S, Álvarez-Jiménez M. Moderated online social therapy: designing and evaluating technology for mental health. ACM Trans Comput Hum Interact 2014;21(1):1-26. [doi: 10.1145/2513179]
15. Alvarez-Jimenez M, Koval P, Schmaal L, Bendall S, O'Sullivan S, Cagliarini D, et al. The Horyzons project: a randomized controlled trial of a novel online social therapy to maintain treatment effects from specialist first-episode psychosis services. World Psychiatry 2021 Jun;20(2):233-243 [FREE Full text] [doi: 10.1002/wps.20858] [Medline: 34002511]
16. Alvarez-Jimenez M, Gleeson JF, Bendall S, Penn DL, Yung AR, Ryan RM, et al. Enhancing social functioning in young people at Ultra High Risk (UHR) for psychosis: A pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]
17. Rice S, Gleeson J, Davey C, Hetrick S, Parker A, Lederman R, et al. Moderated online social therapy for depression relapse prevention in young people: pilot study of a 'next generation' online intervention. Early Interv Psychiatry 2018 Aug;12(4):613-625. [doi: 10.1111/eip.12354] [Medline: 27311581]
18. Gleeson J, Lederman R, Koval P, Wadley G, Bendall S, Cotton S, et al. Moderated online social therapy: A model for reducing stress in carers of young people diagnosed with mental health disorders. Front Psychol 2017;8:485 [FREE Full text] [doi: 10.3389/fpsyg.2017.00485] [Medline: 28421012]
19. Gleeson J, Lederman R, Herrman H, Koval P, Eleftheriadis D, Bendall S, et al. Moderated online social therapy for carers of young people recovering from first-episode psychosis: study protocol for a randomised controlled trial. Trials 2017 Jan 17;18(1):27 [FREE Full text] [doi: 10.1186/s13063-016-1775-5] [Medline: 28095883]
20. Gleeson JFM, Koval P, Zyphur M, Lederman R, Herrman H, Eleftheriadis D, et al. A randomized controlled trial of moderated online social therapy for family carers of first-episode psychosis patients in a specialist treatment setting. Schizophr Res 2023 May;255:203-212. [doi: 10.1016/j.schres.2023.03.019] [Medline: 37004332]
21. McGorry PD, Edwards J, Mihalopoulos C, Harrigan SM, Jackson HJ. EPPIC: an evolving system of early detection and optimal management. Schizophr Bull 1996;22(2):305-326. [doi: 10.1093/schbul/22.2.305] [Medline: 8782288]
22. Yung AR, Nelson B, Stanford C, Simmons MB, Cosgrave EM, Killackey E, et al. Validation of "prodromal" criteria to detect individuals at ultra high risk of psychosis: 2 year follow-up. Schizophr Res 2008 Oct;105(1-3):10-17. [doi: 10.1016/j.schres.2008.07.012] [Medline: 18765167]
23. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
24. Bandura A, Caprara GV, Barbaranelli C, Regalia C, Scabini E. Impact of family efficacy beliefs on quality of family functioning and satisfaction with family life. Appl Psychol 2011;60(3):421-448. [doi: 10.1111/j.1464-0597.2010.00442.x]
25. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
26. Cohen S, Williamson GM. Perceived stress in a probability sample of the United-States. In: Spacapan S, Oskamp S, editors. Social Psychology of Health. Norderstedt, Germany: Books on Demand; 1988:31-67.
27. Davenport MD, Tiefenbacher S, Lutz CK, Novak MA, Meyer JS. Analysis of endogenous cortisol concentrations in the hair of rhesus macaques. Gen Comp Endocrinol 2006 Jul;147(3):255-261. [doi: 10.1016/j.ygcen.2006.01.005] [Medline: 16483573]
28. Stalder T, Kirschbaum C. Analysis of cortisol in hair--state of the art and future directions. Brain Behav Immun 2012 Oct;26(7):1019-1029. [doi: 10.1016/j.bbi.2012.02.002] [Medline: 22366690]
29. Eaton WW. Center for Epidemiologic Studies Depression Scale-Revised (CESD-R). Sarasota, FL: Professional Resource Press/Professional Resource Exchange; 2001.
30. WHO ASSIST Working Group. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. Addiction 2002 Sep;97(9):1183-1194 [FREE Full text] [doi: 10.1046/j.1360-0443.2002.00185.x] [Medline: 12199834]
31. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]
32. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
33. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med 1991;32(6):705-714. [doi: 10.1016/0277-9536(91)90150-b] [Medline: 2035047]
34. Hamilton VE, Matthews JM, Crawford SB. Development and preliminary validation of a Parenting Self-Regulation Scale: “Me as a Parent”. J Child Fam Stud 2014 Dec 13;24(10):2853-2864. [doi: 10.1007/s10826-014-0089-z]

35. Folkman S, Lazarus RS, Dunkel-Schetter C, DeLongis A, Gruen RJ. Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. J Pers Soc Psychol 1986 May;50(5):992-1003. [doi: 10.1037//0022-3514.50.5.992] [Medline: 3712234]
36. MacCarthy B, Brown R. Psychosocial factors in Parkinson's disease. Br J Clin Psychol 1989 Feb;28(1):41-52. [doi: 10.1111/j.2044-8260.1989.tb00810.x] [Medline: 2924026]
37. Wood AM, Linley PA, Maltby J, Kashdan TB, Hurling R. Using personal and psychological strengths leads to increases in well-being over time: A longitudinal study and the development of the strengths use questionnaire. Pers Individ Dif 2011 Jan;50(1):15-19. [doi: 10.1016/j.paid.2010.08.004]
38. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the Self-Compassion Scale. Clin Psychol Psychother 2011;18(3):250-255. [doi: 10.1002/cpp.702] [Medline: 21584907]
39. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003 Apr;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
40. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
41. Keyes CLM, Wissing M, Potgieter JP, Temane M, Kruger A, van Rooy S. Evaluation of the mental health continuum-short form (MHC-SF) in Setswana-speaking South Africans. Clin Psychol Psychother 2008;15(3):181-192. [doi: 10.1002/cpp.572] [Medline: 19115439]
42. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. Patient 2014;7(1):85-96 [FREE Full text] [doi: 10.1007/s40271-013-0036-x] [Medline: 24271592]
43. Wiedemann G, Rayki O, Feinstein E, Hahlweg K. The Family Questionnaire: development and validation of a new self-report scale for assessing expressed emotion. Psychiatry Res 2002 Apr 15;109(3):265-279. [doi: 10.1016/s0165-1781(02)00023-9] [Medline: 11959363]
44. Barnes HL, Olson DH. Parent-adolescent communication and the circumplex model. Child Dev 1985 Apr;56(2):438. [doi: 10.2307/1129732]
45. Szmukler GI, Burgess P, Herrman H, Benson A, Colusa S, Bloch S. Caring for relatives with serious mental illness: the development of the Experience of Caregiving Inventory. Soc Psychiatry Psychiatr Epidemiol 1996 Jun;31(3-4):137-148. [doi: 10.1007/BF00785760] [Medline: 8766459]
46. Williams GC, Grow VM, Freedman ZR, Ryan RM, Deci EL. Motivational predictors of weight loss and weight-loss maintenance. J Pers Soc Psychol 1996 Jan;70(1):115-126. [doi: 10.1037//0022-3514.70.1.115] [Medline: 8558405]
47. Alvarez-Jimenez M, Bendall S, Lederman R, Wadley G, Chinnery G, Vargas S, et al. On the HORYZON: moderated online social therapy for long-term recovery in first episode psychosis. Schizophr Res 2013 Jan;143(1):143-149. [doi: 10.1016/j.schres.2012.10.009] [Medline: 23146146]
48. Bargas-Avila JA, Hornbæk K. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. 2011 Presented at: CHI '11: SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC p. 2689-2698. [doi: 10.1145/1978942.1979336]
49. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 04;345:e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
50. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Consolidated Standards of Reporting Trials Group. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010 Aug;63(8):e1-37 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.03.004] [Medline: 20346624]
51. Altman DG. Comparability of randomised groups. Statistician 1985;34(1):125-136. [doi: 10.2307/2987510]
52. de Boer MR, Waterlander WE, Kuijper LDJ, Steenhuis IHM, Twisk JWR. Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. Int J Behav Nutr Phys Act 2015 Jan 24;12:4 [FREE Full text] [doi: 10.1186/s12966-015-0162-z] [Medline: 25616598]
53. Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry 2004 Mar;61(3):310-317. [doi: 10.1001/archpsyc.61.3.310] [Medline: 14993119]
54. Mallinckrodt CH, Clark SWS, Carroll RJ, Molenbergh G. Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations. J Biopharm Stat 2003 May;13(2):179-190. [doi: 10.1081/BIP-120019265] [Medline: 12729388]
55. Gleeson JFM, Cotton SM, Alvarez-Jimenez M, Wade D, Crisp K, Newman B, et al. Family outcomes from a randomized control trial of relapse prevention therapy in first-episode psychosis. J Clin Psychiatry 2010 Apr;71(4):475-483. [doi: 10.4088/JCP.08m04672yel] [Medline: 20021994]
56. McCann TV, Lubman DI, Cotton SM, Murphy B, Crisp K, Catania L, et al. A randomized controlled trial of bibliotherapy for carers of young people with first-episode psychosis. Schizophr Bull 2013 Nov;39(6):1307-1317 [FREE Full text] [doi: 10.1093/schbul/sbs121] [Medline: 23172001]
57. Keller JM. Development and use of the ARCS model of instructional design. J Instr Dev 1987 Sep;10(3):2-10. [doi: 10.1007/bf02905780]

58. Noetel M, Griffith S, Delaney O, Harris NR, Sanders T, Parker P, et al. Multimedia design for learning: an overview of reviews with meta-meta-analysis. Rev Educ Res 2021;92(3):413-454. [doi: 10.3102/00346543211052329]
59. Pilling S, Bebbington P, Kuipers E, Garety P, Geddes J, Orbach G, et al. Psychological treatments in schizophrenia: I. Meta-analysis of family intervention and cognitive behaviour therapy. Psychol Med 2002 Jul;32(5):763-782. [doi: 10.1017/s0033291702005895] [Medline: 12171372]
60. Chien WT, Bressington D, Lubman DI, Karatzias T. A randomised controlled trial of a caregiver-facilitated problem-solving based self-learning program for family carers of people with early psychosis. Int J Environ Res Public Health 2020;17(24):9343 [FREE Full text] [doi: 10.3390/ijerph17249343] [Medline: 33327452]
61. Kapur S, Munafò M. Small sample sizes and a false economy for psychiatric clinical trials. JAMA Psychiatry 2019 Jul 01;76(7):676-677. [doi: 10.1001/jamapsychiatry.2019.0095] [Medline: 30916722]

Abbreviations

hEP: headspace early psychosis
HPA: hypothalamic-pituitary adrenal
ITT: intent-to-treat
MHC-SF: Mental Health Continuum Short Form
MMRM: mixed effects repeated measures
MOS-SSS: Medical Outcomes Study Social Support Survey
MOST: moderated online social therapy
PAC: Parent-Adolescent Communication
PSS: Perceived Stress Scale
PSWQ: Penn State Worry Questionnaire
QoL: quality of life
RA: research assistant
SCS-SF: Self-Compassion Short Form
SUS: Strengths Use Scale
SWLS: Satisfaction With Life Scale
TAU: treatment as usual
UCLA: University of California, Los Angeles
WHO: World Health Organization

Edited by J Torous; submitted 03.04.23; peer-reviewed by A Hudon, A AL-Asadi; comments to author 10.05.23; revised version received 31.05.23; accepted 19.06.23; published 06.09.23.

Please cite as:
Gleeson J, Lin A, Koval P, Hopkins L, Denborough P, Lederman R, Herrman H, Bendall S, Eleftheriadis D, Cotton S, Perry Y, Kaess M, Alvarez-Jimenez M
Moderated Online Social Therapy for Carers of Early Psychosis Clients in Real-World Settings: Cluster Randomized Controlled Trial
JMIR Ment Health 2023;10:e47722
URL: https://mental.jmir.org/2023/1/e47722
doi:10.2196/47722
PMID:37672335

Original Paper

User-Centered Development of Bolster, an mHealth Intervention for Early Psychosis Caregivers: Needs Assessment, Prototyping, and Field Trial

Benjamin Buck[1], PhD; Mary Wingerson[1], BA; Erica Whiting[1], BS; Jaime Snyder[2], MFA, PhD; Maria Monroe-DeVita[3], PhD; Dror Ben-Zeev[1], PhD

[1]Behavioral Research in Technology and Engineering (BRiTE) Center, Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

[2]Information School, University of Washington, Seattle, WA, United States

[3]Supporting Psychosis Innovation through Research Implementation and Training (SPIRIT) Lab, Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

Corresponding Author:
Benjamin Buck, PhD
Behavioral Research in Technology and Engineering (BRiTE) Center
Department of Psychiatry and Behavioral Sciences
University of Washington
Box 356560
1959 NE Pacific Street
Seattle, WA, 98105
United States
Phone: 1 206 221 8518
Fax: 1 206 543 9520
Email: buckbe@uw.edu

Abstract

Background: Caregivers play a critical role in the treatment and recovery of youth and young adults at risk for psychosis. Caregivers often report feeling isolated, overwhelmed, and lacking in resources. Mobile health (mHealth) has the potential to provide scalable, accessible, and in-the-moment support to caregivers. To date, few if any mHealth resources have been developed specifically for this population.

Objective: The aim of this study was to conduct user-centered design and testing of an mHealth intervention to support early psychosis caregivers.

Methods: We conducted a multiphase user-centered development process to develop the Bolster mobile app. In phase 1, a total of 21 caregivers were recruited to participate in a qualitative needs assessment and respond to an initial prototype of the Bolster platform. Content analysis was used to identify key needs and design objectives, which guided the development of the Bolster mobile app. In phase 2, a total of 11 caregivers were recruited to participate in a 1-week field trial wherein they provided qualitative and quantitative feedback regarding the usability and acceptability of Bolster; in addition, they provided baseline and posttest assessments of the measures of distress, illness appraisals, and family communication.

Results: In phase 1, participants identified psychoeducation, communication coaching, a guide to seeking services, and support for coping as areas to address. Live prototype interaction sessions led to multiple design objectives, including ensuring that messages from the platform were actionable and tailored to the caregiver experience, delivering messages in multiple modalities (eg, video and text), and eliminating a messaging-style interface. These conclusions were used to develop the final version of Bolster tested in the field trial. In phase 2, of the 11 caregivers, 10 (91%) reported that they would use Bolster if they had access to it and would recommend it to another caregiver. They also reported marked changes in their appraisals of illness (Cohen *d*=0.55-0.68), distress (Cohen *d*=1.77), and expressed emotion (Cohen *d*=0.52).

Conclusions: To our knowledge, this study is the first to design an mHealth intervention specifically for early psychosis caregivers. Preliminary data suggest that Bolster is usable, acceptable, and promising to improve key targets and outcomes. A future fully powered clinical trial will help determine whether mHealth can reduce caregiver burdens and increase engagement in services among individuals affected by psychosis.

(JMIR Ment Health 2023;10:e50522) doi:10.2196/50522

KEYWORDS

caregiving; psychosis; mobile health; mHealth; user-centered design; mobile phone; artificial intelligence; AI

Introduction

Background

Caregivers, or individuals who provide ongoing support or help to an individual (usually a family member, eg, a child, spouse, or other relative) with a mental illness [1], play a critical role in the recovery of young adults at risk for psychosis [2]. Most youth with early psychosis live with a caregiver [3], and these caregivers are often the first to detect and respond to the signs of psychosis in their loved ones [4]. They also often seek resources and treatment [5], establish contact with providers [6], or work to persuade the individual to make such contacts on their own [7]. Furthermore, once their loved ones are connected to care, caregivers provide logistical and emotional support [8]. Having a caregiver engaged in supporting care is associated with increased engagement in services [9,10] and improved outcomes [11,12].

Caregivers face significant barriers and challenges in their efforts to facilitate help seeking. Many lack accurate information about psychosis and are prone to misattribute symptoms to substances or developmental transitions [13]. Stigmatizing attitudes related to mental illness and help seeking can manifest in unwillingness to reach out to providers or disbelief in the seriousness of symptoms [14]. Caregivers are also prone to severe levels of distress related to confusion, frustration, and worry induced by responding to a loved one's psychotic episodes [15] and the experience of mental illnesses of their own [16,17]. Better caregiver knowledge of psychosis symptoms is linked with an increased likelihood of recommending professional help seeking for the affected person [18]. However, caregivers commonly report a skill deficit when faced with the challenges of caregiving in the context of psychosis. They often report feeling isolated [19], distressed, and unable to help [20]. Without effective caregiving skills, they are vulnerable to a communication style that is overly emotionally involved, alienating, or hostile (ie, interactions high in *expressed emotion*) [21,22]. This interaction style can worsen family communication [23], which impedes the process of treatment facilitation [24] and is linked with elevated symptoms [25-27] and increased risk for future episodes [28-32]. Family psychoeducation interventions are designed specifically to address these needs; however, specialty psychosis clinics that offer them are out of reach for families who—as a result of limited access or reluctance—have not yet accessed them or are unable to engage in services (eg, because of the distance from clinics and other responsibilities). To address these needs, caregivers often search the internet for resources, information, and support early in the help-seeking process [33]. A majority of caregivers report dissatisfaction with the illness-related information available to them, and they report particular difficulties accessing necessary and actionable information when it is most needed [34]. Caregivers report that although information can be accessed easily on the internet, much of it lacks direct and actionable steps [4].

Mobile health (mHealth) may provide unique advantages in addressing the needs of early psychosis caregivers. First, mHealth tools can serve as *just-in-time* interventions [35], wherein users can use content that responds to particular in-the-moment needs. Caregiving challenges can unfold in an unpredictable manner; optimal resources for caregivers may be those that are available in the environments and moments when these needs emerge. mHealth could also enable more frequent ongoing engagement with intervention content than is possible through psychoeducational websites, many of which provide information in a single large bolus. Second, mobile apps have greater penetration than other web-based platforms. In a day's engagement with media, the average adult spends much more time engaged with mobile devices than internet-connected computers [36]. Third, mHealth interventions show promise in their scalability. A new generation of self-guided interventions have demonstrated efficacy in a number of psychiatric concerns, including depression [37], anxiety [38], and psychosis [39]. Such interventions can be provided rapidly, at low cost, to individuals who face significant barriers to access the traditional modes of mental health services.

mHealth Interventions for Caregivers

mHealth interventions designed for caregivers lag behind those designed for individuals with psychosis or other mental health conditions. A recent systematic review of digital technologies for early psychosis caregivers examined 8538 studies and identified no mHealth intervention designed for this high-need population [40]. Given the fact that many mental health care settings focus exclusively on the needs of the identified patient and lack specific services for caregivers [1], self-guided digital tools might fill a particular glaring need in this population. There is a need for literature describing the specific applications of mHealth to address the needs of caregivers of individuals at risk for psychosis. Our team has conducted a multiphase user-centered design and development process to develop an mHealth intervention for caregivers of individuals with early psychosis to support treatment facilitation, including a qualitative needs assessment, live prototyping sessions, and a field trial. The finished product—Bolster—is one of the first ever mHealth interventions designed specifically to support caregivers of young adults with early psychosis. In this paper, we describe the user-centered design process through which Bolster evolved from a set of design objectives to a prototype to a fully functional mHealth support tool.

Methods

Overview

Following examples from our team [41,42] and others [43,44], we conducted a multiphase user-centered design and development process that aimed to optimize the Bolster

intervention to meet the needs of caregivers of young adults at risk for psychosis. This user-centered development process involved 2 phases. In study phase 1, a total of 21 caregivers participated in a qualitative needs assessment to identify key intervention goals and engaged in a live interaction with a preliminary prototype. Our team analyzed the results of this first phase to develop and optimize the beta version of the Bolster intervention. In study phase 2, a second sample of 11 caregivers participated in a 1-week field trial of the Bolster mobile app.

Recruitment

Several channels were used to recruit participants. First, advertisements were purchased on Google by the research team to target individuals using particular search terms (eg, "schizophrenia symptoms," "psychosis," and "bipolar symptoms") and iteratively optimized using the Google Ads *broad match* algorithm. Second, the study team placed advertisements in the newsletters of multiple advocacy organizations (eg, the National Alliance on Mental Illness [NAMI] and Mental Health America) and sent email postings to national networks focused on early psychosis (eg, the Psychosis-Risk and Early Psychosis Program Network [PEPPNET] and Washington state's New Journeys network) encouraging distribution to caregivers served by members of these networks. In each of these postings or flyers, participants were linked to a study landing page that summarized the study and contained links to the consent form and eligibility survey. After opting to take the eligibility survey, participants were required to read the study consent form and confirm understanding in comprehension questions covering study key points. All participants were contacted by a member of the study team to confirm and inform participants of their eligibility status.

Participants

Across study phases, participants were 32 caregivers of youth and young adults who had experienced symptoms consistent with psychosis. All met the following inclusion criteria: (1) responses of *somewhat agree* or *definitely agree* to ≥2 items on the Caregiver Prime Screen–Revised (CGPS-R) [45]; (2) loved one met the criteria of psychosis onset (per caregiver report) following definitions derived from previous work [46,47], including (2A) experiencing at least 1 positive symptom of psychosis and (2B) at least 2 of the following: serious deterioration of functioning, marked social withdrawal, persistent self-neglect, or episodic marked anxiety; and (3) within the past 5 years, caregivers became aware of their loved one's symptoms, their loved one first experienced a psychotic episode, or their loved one first engaged in treatment for psychosis. To represent varying caregiver experiences concerning treatment engagement, the first sample (ie, those completing the needs assessment and prototyping) was stratified such that only half provided caregiving to an individual who was already established in specialty services for psychosis. In the phase 2 field trial, all participants were caregivers of young adults who were engaged in specialty psychosis programs.

Phase 1: Intervention Development

Procedures

In the study's first phase, we aimed to develop Bolster through identifying user needs and gathering concrete feedback in response to a preliminary prototype. This study builds on our team's previous quantitative work [33] by gathering detailed qualitative data on needs and preferences as well as more direct and actionable information in response to a concrete prototype. Participants completed semistructured needs assessment interviews followed by live prototype interaction sessions through videoconferencing links provided by the research team. The interviewer (MW) was a BA-level research coordinator following a semistructured interview guide developed by senior team members with experience in user-centered design; these senior team members also supervised the interviewer. The interviewer had no established previous relationship with participants, and she provided them with the goals of the research project during the informed consent process. Participants used their own devices for videoconferencing and thus could choose their own environments to complete data collection. No other individuals were present for these interviews aside from the interviewer and the participant. Qualitative interview topics included barriers and facilitators of help seeking, currently available resources that support caregiving, unmet needs in supporting the affected relative's help seeking, and proposed digital solutions to better meet these needs (ie, what ideal digital supports the participant would create to meet their needs if they had a *magic wand*). Live prototyping sessions involved engaging with a preliminary but fully responsive prototype. This prototype included the following features: an interactive messaging-style (ie, preprogrammed rule-based *chatbot*) interface where users interacted with the system as it provided psychoeducational content. Example modules focused on caregiver resilience (ie, the importance of taking care of oneself to more effectively provide caregiving support) and the communication skill of using *I-statements*. Additional modules involved a video-based deep breathing exercise and a text-based psychoeducation page introducing the stress-vulnerability model of psychosis. This prototype allowed the demonstration and testing of the structures and functions of the platform as well as sample content to allow test participants to provide feedback. The Bolster prototype was connected to the videoconferencing interview session, and cursor control was provided to the participants such that they could click through and respond to prompts in the app remotely. Participants were asked to complete *tasks* (eg, complete a module) while *thinking aloud*. In this approach, participants are asked to express reactions to the functions and content of the app as they come to mind and after completing each task. Interview sessions lasted approximately 90 minutes and were audio recorded and later transcribed for research team analysis. Participants were compensated with US $75 gift cards for their time. There were no additional contacts with participants for repeat interviews or feedback on findings.

Analysis

Participants' qualitative responses were analyzed for key themes using conventional content analysis [48]. Interviews were

segmented such that themes could only be coded in relevant sections of the interview. Two coders independently assigned first-level codes (concise summaries no longer than a phrase, eg, "Ability to persuade loved one to take her medication") to segments of participants' responses. The 2 coders grouped first-level codes and collaboratively developed second-level interpretive codes (eg, "Communication with loved one") that also serve as themes reported here. Second-level codes were developed iteratively following discussion of first-level coding and compiled into a codebook with thorough definitions and rules for application in text. The coders then independently returned to the text to apply second-level codes to interview segments. A high degree of agreement was established between the coders (κ>0.72), and all disagreements were resolved through consensus discussion. Prototype reactions (ie, problematic or positive interactions and issues noted in the *think aloud* sessions) were listed exhaustively to provide a comprehensive review of areas to adjust in the final version of Bolster. We also report on our team's design objectives and overall features of the version of Bolster tested in the phase 2 field trial.

Phase 2: Field Trial

Procedures

This field trial deployed the updated version of Bolster based on phase 1 feedback in a sample of caregivers for 1 week. Eligible participants were texted links to a battery of baseline questionnaires (refer to the next subsection). Once these questionnaires were completed, a study team member scheduled and completed a remote *installation session* wherein the study team member ensured that Bolster was installed on the participants' own device and provided orientation to the app, the 1-week testing period, and the remaining assessment schedule. Participants were encouraged to use the app as they would normally but also to do so in a manner that allowed them to provide honest and detailed feedback. If participants did not engage in the app for multiple consecutive days, a member of the study team contacted the participant on the third day to ensure that the platform was functioning properly or to provide a reminder of the upcoming end of the testing period. This type of outreach occurred for 4 (36%) of the 11 participants. Participants were sent links to complete the full baseline assessment battery a second time on day 8. Following completion of the postintervention questionnaires, participants were invited to postintervention qualitative interviews. In these interview sessions, participants connected to the videoconferencing platform with their own smartphones such that the interviewer could observe their interactions with the platform in real time. During these interviews, participants were asked (1) to provide ratings of (1A) how easy to use and (1B) useful they found each feature, as well as (2) to show on their devices something memorable and representative (positive or negative) of their experience with Bolster. Participants were compensated with a US $50 gift card per completed assessment (i.e. baseline and post-test) as well as a US $40 gift card for completing a qualitative interview.

Measures

Bolster was designed to have an impact on the primary targets of illness knowledge, illness appraisals (caregiving-related appraisals and psychosis-related appraisals), and coping in hopes that changes in these primary targets would result in improvements in the primary outcomes of caregiver distress and expressed emotion. Illness knowledge was assessed with the Knowledge About Schizophrenia Test (KAST) [49], an 18-item multiple-choice assessment that scored participants' knowledge of the etiology, symptoms, and prognosis of schizophrenia on the number of items answered correctly. The Illness Perception Questionnaire for Schizophrenia: Relatives' version (IPQSR) [50] is a self-report scale of caregivers' beliefs about the severity, prognosis, and responsiveness to treatment of mental illnesses and used to assess caregiver knowledge of illness appraisals. We administered subscales focused on emotional representation (high scores reflecting high levels of distress related to the condition), incoherence (high scores reflecting lacking an understanding of the illness), and the measures of control (high scores reflecting belief that the actions of the individual and the caregiver can affect the course of the illness) and consequences (high scores reflecting a perception that psychosis results in greater negative consequences for the individual and the caregiver; given evidence from Lobban et al [50] suggesting relationships between the caregiver version and patient version of each *consequences* and *control* subscale, these were combined for simplicity). Caregiving-related appraisals were assessed with the Experiences of Caregiving Inventory (ECI) [51], a 66-item assessment of the perceived impact of caregiving on the individual's life, with subscale scores for both negative and positive experiences. Coping was assessed with 2 measures: the Brief Coping Orientation to Problems Experienced Inventory (Brief-COPE) [52], a 28-item self-report scale of coping skills in response to stressors, with 16 items assessing coping skills proposed a priori as positive [51]; and the Coping Self-Efficacy Scale (CSES) [53], a 26-item self-report questionnaire measuring the perceived ability to cope with various life challenges. Expressed emotion was assessed using the Family Questionnaire (FQ) [54], a 20-item self-report assessment of emotional expression in family members toward patients with mental illness, and caregiver distress was assessed a revised version of the General Health Questionnaire, 12-item version (GHQ-12) [55,56], a self-administered questionnaire that measures general psychological distress. The scale was revised consistent with recent psychometric studies examining the confounds of wording effects on this measure [57] in that scale items were kept consistent (ie, ranging from *not at all* to *much more than usual*). Positively worded items are reverse scored such that higher scores indicate greater distress.

Analysis

Analyses were conducted using SPSS software (version 28.0; IBM Corp). Acceptability and usability were assessed through a review of individual items on the modified System Usability Scale. With regard to use statistics, we also reviewed the following metrics: (1) the percentage of days on which participants opened Bolster of those days on which they had access, (2) the number of minutes per day that participants used Bolster, and (3) the number and rate of completion of Bolster

lessons and practices. We conducted exploratory analyses of clinical outcomes using paired sample 2-tailed *t* tests, assessing significant within-individual change during the testing period and reviewing effect sizes.

Ethical Considerations

The institutional review board of the University of Washington approved all study procedures (STUDY00013334). All participants provided informed consent to participate. Any identifiable information was kept on secure and password-protected servers; data were deidentified for analysis.

Results

Participant Characteristics

Participant demographics across study phases are provided in Table 1. The sample consisted almost exclusively of women (31/32, 97%) and parents (31/32, 97%) aged approximately 50 to 55 (mean 50.19, SD 10.74 in phase 1; mean 55.27, SD 7.85 in phase 2) years. The affected relative was on average aged in their early to mid-20s and had been experiencing symptoms of psychosis for approximately 3 years. The majority of participants (23/32, 72%) were non-Hispanic White. Most of the caregivers (22/32, 69%) endorsed a diagnosis of schizophrenia for their loved one and endorsed symptoms and scores on the CGPS-R that reflected a high severity of illness. As stipulated by our a priori stratified recruitment strategy in phase 1, a total of 10 (48%) of the 21 participants were caring for individuals who were not engaged in treatment or were in the process of supporting their loved one seeking a higher level of care. The remainder (11/21, 52%) were caring for individuals who were already established in specialty mental health services. Across phases, of participants whose loved ones were engaged in specialty mental health services, the majority (17/22, 77%) were engaged in coordinated specialty care programs. No participant dropped out or refused to participate in any study procedures.

Table 1. Demographic characteristics of study participants.

Characteristics	Sample 1: needs assessment and prototyping (n=21)	Sample 2: field trial (n=11)
Age (y), mean (SD)	50.19 (10.74)	55.27 (7.85)
Age (y) of loved one, mean (SD)	23.33 (5.18)	24.55 (3.36)
Sex, n (%)		
Female	20 (95)	11 (100)
Male	1 (5)	0 (0)
Gender, n (%)		
Woman	20 (95)	11 (100)
Man	1 (5)	0 (0)
Relationship to loved one, n (%)		
Parent	20 (95)	11 (100)
Sibling	1 (5)	0 (0)
Race and ethnicity, n (%)		
Asian	3 (14)	1 (9)
Black or African American	1 (5)	2 (18)
Hispanic White	2 (10)	0 (0)
Non-Hispanic White	15 (71)	8 (73)
Loved one's diagnosis[a], n (%)		
Schizophrenia or schizoaffective disorder	14 (67)	8 (73)
Bipolar disorder	8 (38)	6 (55)
PTSD[b]	5 (24)	3 (27)
Major depressive disorder	6 (29)	2 (18)
Other psychotic disorder	4 (19)	1 (9)
Not sure or prefer not to say	1 (5)	1 (9)
CGPS-R[c], mean (SD)	51.33 (9.88)	54.88 (10.10)
CGPS-R (number of items rated *somewhat agree* or higher), mean (SD)	6.33 (2.33)	7.09 (2.98)
Symptoms endorsed, n (%)		
Hallucinations	17 (81)	10 (91)
Thought disorder	19 (91)	11 (100)
Delusions	20 (95)	10 (91)
Bizarre or erratic behavior	20 (95)	10 (91)
Strange psychomotor behavior	11 (52)	7 (64)
Functional impacts endorsed, n (%)		
Deterioration in daily functioning	19 (91)	11 (100)
Social withdrawal	18 (86)	11 (100)
Persistent self-neglect	13 (62)	7 (64)
Severe anxiety or agitation	20 (95)	10 (91)
Years since first aware of symptoms, mean (SD)	2.93 (2.15)	3.63 (2.69)
Years since first episode (if applicable), mean (SD)	2.98 (2.07)	3.01 (2.34)

[a]Coded nonexclusively.

[b]PTSD: posttraumatic stress disorder.

[c]CGPS-R: Caregiver Prime Screen–Revised.

Phase 1: Intervention Development

Qualitative Needs Assessment

During the course of the analysis, the study team observed that the structure of 2 qualitative interview prompts—unmet needs and ideal supports (the *magic wand* question)—often generated responses that aligned with one another or reiterated key themes (eg, unmet need of *actionable information related to help seeking* and ideal feature of *an actionable guide to help seeking*). In light of this, our team analyzed and interpreted these items together (refer to Table 2 for examples and theme frequencies). Thus, these themes reflected salient areas to address in a digital intervention based on the ways in which they represented current unmet needs. Themes fell into four broad categories: (1) communication coaching, (2) general psychoeducation, (3) guide to treatment seeking, and (4) support for caregiver coping.

Table 2. Combination of themes from unmet needs and magic wand items linked to Bolster elements (n=21).

Category and theme identified in response to *unmet needs* prompt	Theme frequency, n (%)	Theme identified in response to *magic wand* prompt	Theme frequency, n (%)	Bolster elements (ie, ways to address themes using scalable self-guided mHealth[a])
Communication coaching				
Communication with loved one	18 (86)	• Tool supporting communication with loved one	• 11 (52)	Interactive communication coaching such that the caregiver can practice various communication skills in the app via text or be coached through how to complete an in vivo practice session
Communication with others about psychosis	7 (33)	• No theme identified	—[b]	Lessons and practices that target stigmatizing attitudes alongside information on strategies for how to talk with others about your loved one's experience
Psychoeducation				
General information and knowledge of psychosis	11 (52)	• Psychosis-related information and guide to recognizing symptoms • Expert Q&A[c]	• 11 (52) • 5 (24)	*Psychosis 101* lessons providing introduction to key points about psychosis
General information and knowledge of treatments and medications	7 (33)	• Treatment-related information	• 9 (43)	*Psychosis 101* lessons providing introduction to key points about treatment and recovery
No theme identified	—	• Tracking loved one's symptom presentation	• 6 (29)	Structured instrument to aid caregivers in tracking symptom changes over time
Guide to seeking services				
Actionable guide to help seeking	11 (52)	• Actionable guide to help seeking	• 11 (52)	Actionable step-by-step guide on how to encourage your loved one to seek services, as well as how to go about finding appropriate treatment for your loved one
Actionable guide to other (ie, nonmedical) resources	5 (24)	• Actionable guide to nonmedical or nonpsychiatric resources	• 7 (33)	Links to information about resources related to housing, employment, or food security
Planning for, and responding to, emergencies and crisis situations	7 (33)	• Crisis-related information and support	• 9 (43)	Step-by-step guide to evaluating and responding to a crisis, as well as links to emergency resources if caregiver is responding to crisis in the moment
Support for coping				
Support for caregiver's coping	11 (52)	• Peer connections • Self-guided support for caregiver well-being	• 9 (43) • 8 (38)	Video and interactive text-based practices supporting cognitive restructuring (eg, self-compassion), behavioral exercises (eg, heathy habits), and simple mindfulness practices (eg, deep breathing)

[a]mHealth: mobile health.

[b]Not applicable.

[c]Q&A: question and answer.

Responses related to communication were common and salient in many interviews, at times regarding communication with others (eg, family and friends) about the loved one's symptoms but primarily in relation to communicating directly with the loved one and especially in response to difficult topics (eg, in response to delusions, disorganization, or opposition to engagement in mental health treatment):

> *How can you give her support when it doesn't make sense because the support she wants doesn't make any sense because it's not real?...Learning how to*

> *support them without trying to make them think that you believe what they're saying is true.* [Unmet needs, participant 67]
>
> *So, if there was a nice app that said "Okay, people in psychosis may interact with you like this and here are some techniques to respond and here's when you need to disengage," because it was clear to me at some points that the best thing I could do was get out, my presence was not helpful.* [Magic wand, participant 5]

On the basis of insights from caregivers related to this topic, the project team ensured that Bolster was primarily oriented around communication skills, including both foundational listening skills (eg, active listening) and those related to particular challenging topics (eg, responding to delusions). This content would provide both clear directives and opportunities to practice new skills.

Caregivers also commonly reported feeling overwhelmed and in need of support in coping with this demanding role. This manifested in desires for reminders to engage in self-care and supportive connections with peer caregivers who had shared experiences:

> *How do you take care of yourself, remembering to take care of yourself. Because you're so much into the problem, that then you forget you. So people forget themselves to the extent that they start failing in their jobs, and that's important.* [Unmet needs, participant 126]
>
> *[T]o have a trained person who can work with you on developing boundaries and helping you create a strategy, helping to hold you accountable to the things you said you were going to do. That would be really helpful.* [Magic wand, participant 66]

Caregivers noted the common experience of neglecting their own needs and well-being, often acknowledging the dissonance between an awareness of the importance of one's own needs while lamenting the frequency with which they neglected them.

Caregivers emphasized 2 broad categories related to informational needs as well, including general psychoeducation (eg, information about psychosis as well as its treatment and prognosis) and a guide to engaging in the mental health services system, both for ongoing outpatient treatment and in response to psychiatric crises. Many participants noted that although informational resources exist on the web (and in concert with treatment programs), it is not always easy to tell which sources are trustworthy, and even among those that seem to be trustworthy, many provide general overviews more effectively targeting a general or academic audience rather than a caregiver needing direct guidance in response to specific challenging situations:

> *When you research something like that, you have to read a lot of different sources, so having one place that has as much information as it can pack in, that explains in a way that anyone can understand.* [Unmet needs, participant 73]
>
> *I didn't know how to find support, I didn't have a list of potential therapists outside of the [treatment] program...He came home untreated in a lot of ways and I didn't have any direction on how to find those resources.* [Unmet needs, participant 82]

In response to the magic wand item, a number of caregivers (6/21, 29%) also reported a desire to have the ability to assess, track, and monitor changes in symptoms over time, either to be able to determine whether symptoms were improving or to more easily report changes to providers at future clinical visits. Two additional prompts in the qualitative interview focused specifically on the positive features of existing resources that caregivers had used for support as well as attributes that were negative or lacking. First, digital or web-based resources that caregivers had found useful were most commonly described as detailed (12/21, 57%), hopeful (11/21, 52%), and providing clear actionable steps for help seeking (4/21, 19%) or in communicating with their loved one (3/21, 14%). Frustrations that were related to web-based resources often stemmed from such resources simply being absent (7/21, 33%) or the fact that existing resources were provided in the context of clinical services that caregivers found to be inadequate or unhelpful (10/21, 48%); were lacking in detail (3/21, 14%); discouraging (5/21, 24%); or, again, lacking clear actionable guidance (7/21, 33%):

> *I mean there's a lot of general information about psychosis, Google is there...When you see somebody who's in psychosis, you don't need to read an article about what is psychosis. You're like, "I know what psychosis is, thank you." But what do you do? General information isn't helpful when you are in a crisis situation.* [Disappointing resources, participant 11]

Prototyping

Open-ended feedback on the Bolster prototype was overall positive, but it also identified areas for improvement. Participants reported finding the information useful and new and the tone of the content to be encouraging and comforting. Several participants were particularly enthusiastic about the opportunities for interactive practices and reacted positively to opportunities to write their responses into the app, which made the platform feel more personalized.

Most notable among areas for improvement was the messaging-style interface. Participants expressed some frustration with the fact that this method of delivery was not self-paced (eg, "It's choppy"), and others were confused by whether and from whom app content was being delivered conversationally (eg, "Who am I talking to?"). In addition, many participants expressed a desire for more detailed information in response to the sample psychoeducational prompt, but they also expressed some reticence about providing too much text. One comment that balanced these concerns was a suggestion to provide information in diverse presentation modes (eg, video and audio in addition to text) as well as highlighting key or important points within larger chunks of text. Participants also stressed the importance of setting expectations in advance of each Bolster module (eg, providing a summary statement of the

purpose of a module and a time estimate). Several participants reported a desire for a symptom tracker that they could use to track and follow changes in their loved one's symptoms over time, and, if necessary, share these updates with a provider at a future visit. Most other comments pertained to suggestions for additional content (eg, crisis line telephone numbers and information on communicating with others about psychosis) and minor changes to navigation (eg, adding home, back, and next buttons) or display (eg, greater color contrast and removing particular icons).

Insights and Design Objectives

Design objectives were identified based on insights collected from qualitative interviews and prototyping sessions. First, our team identified content areas to develop based on the qualitative needs assessment, including communication coaching, general psychoeducation, information on seeking services, and support for coping. Given the fact that many caregivers expressed concerns about their difficulty *fitting in* additional activities or practices because of the current demands on their time, we resolved to create modules that were brief and simple and that normalized difficulties in making time to engage in the system. In addition, it was important for Bolster content to communicate information with an appropriate tone that balanced accurate information about the seriousness of psychosis with appropriate levity and positive messages about recovery.

Second, participants expressed a desire for clear actionable guidance rather than more general psychoeducational information. On the basis of these comments, a key guiding design principle for Bolster was to provide information and support that was targeted and actionable, as opposed to generic. The Bolster app could provide ongoing scaffolding and support for caregivers that differs from informational websites that are designed to provide overviews of key topics in a few visits (or a single visit). These comments resulted in the proposal of new features, including an index of clear directive guides in responding to common challenging situations (refer to *Action Plans* in the next subsection); curated links to resources, additional information, and treatment listings (refer to *Resources* in the next subsection); and the ability to track the loved one's symptoms over time (refer to *Tracking* in the next subsection).

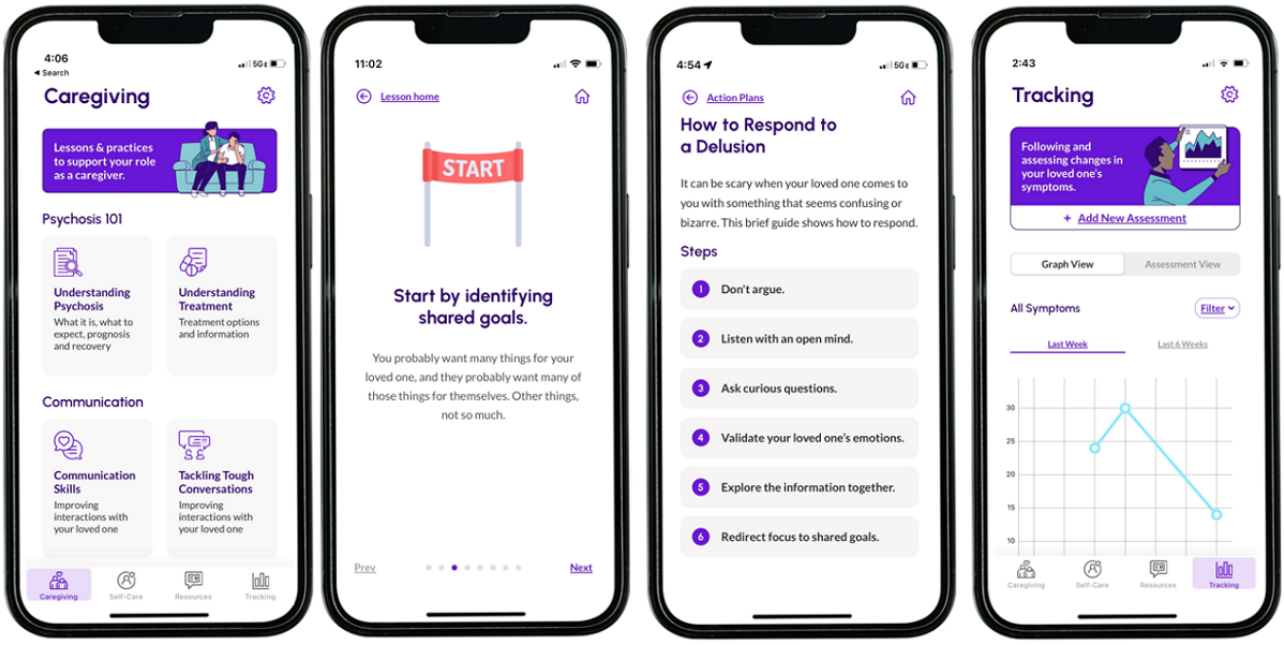
Third, participants had difficulties with content delivered in a messaging-style interface. Our team responded to these concerns by altering the method of presentation of psychoeducational content. All message-style interactions were removed. In their place, all psychoeducation lessons were delivered through 2 modalities: videos of a clinician explaining a key idea and a text-based *carousel*. In these carousels, participants could swipe through a series of screens that each introduced 1 key idea in 1 or 2 sentences at a time with a related illustration or icon. This method responds to several key comments made by participants: it is self-paced, highlights key information, and reduces the overall amount of text on each page.

Fourth, given participants' positive reaction to interactive features (eg, those where they could enter information specific to their situation) and their expressed need for communication skills coaching, more interactive text-based modules were added such that each lesson introducing a skill included an accompanying practice; for example, a video and carousel would introduce the communication skill of using open-ended questions (ie, the *lesson*), and an interactive text-based module (ie, the *practice*) would prompt the user to brainstorm ways to adjust their communication style to include more open-ended questions. The user could choose to engage in either the lesson or the practice on its own or one after the other. Finally, to address concerns about user privacy, our team designed Bolster to only save written text on the user's device (and not in a database); thus, no one—neither the app developers nor the study team members—could access what the user wrote in free-response items. On the basis of these conclusions, new mock-ups were developed by the research team mapping out design objectives, and these were incorporated by the software developers who built the final source code for Bolster.

The Bolster mHealth Intervention

The final Bolster mobile app is a web-based mobile app for iOS, with the test version available on the Apple App Store. Intervention content was drafted by the study team, 3 of whom are clinical psychologists with specialized expertise in cognitive behavioral therapy for psychosis and/or family psychoeducation. The Bolster app's components are based on the cognitive model of caregiving [16], according to which caregiver appraisals lead to emotional and behavioral changes that affect interactions with the affected individual and with service providers. Bolster aims to improve illness and caregiving appraisals and support coping through 4 primary sections (Figure 1). *Caregiving* provides users with psychoeducation and communication coaching through paired lessons and practices. *Self-care* similarly offers paired lessons and practices; however, these focus specifically on skills related to managing one's own stress and well-being (eg, mindfulness, behavioral activation, and self-compassion). *Resources* offer links to external web pages, treatment listings, and videos of young adults and family members with lived experience describing their experiences. Within the *Resources* tab, the *Action Plans* feature provides users with specific targeted guidance in responding to challenging situations related to caregiving (eg, responding to delusions and encouraging help seeking). Finally, *Tracking* provides users with the ability to enter their perception of their loved ones' symptoms (based on *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* [DSM-5] symptoms of psychosis [58,59]), and then graphs and indexes these ratings so that changes can be tracked over time. Importantly, Bolster was designed such that no identifying information is collected by the app, and nor are written responses stored anywhere in the app or in a database, thus providing users with privacy. Even so, data are encrypted in transit and at rest using secure HTTP and transport layer security.

Figure 1. Screenshots of the Bolster final version: (A) the caregiving main menu screen, (B) an example of a lesson page delivered through a swiping carousel, (C) the action plans menu, and (D) the tracking graph view.



Phase 2: Field Trial

Use of Bolster

All participants who installed Bolster completed the study, and none were lost to follow-up. Participants opened Bolster on 66% (58/88) of the days that it was available and on average had Bolster open for 26.64 (SD 13.61) minutes per use day. This equated to 14.87 (SD 7.21) minutes per day on average during the testing period overall. Participants opened 22.91 (SD 17.69; 2.86/participant/d) lessons and 10.91 (SD 8.80; 1.32/participant/d) practices on average. Completion rates for practices were quite high: participants fully completed 86.5% (218/252) of the lessons and 80% (96/120) of the practices that they initiated. Among the lessons opened, 47.6% (120/252) pertained to psychoeducation about psychosis and treatment, 29.8% (75/252) to communication skills, and 22.6% (57/252) to self-care skills. Practices were somewhat evenly split between their 2 categories: communication skills (70/120, 58.3%) and self-care (50/120, 41.7%).

Usability and Acceptability of Bolster

Participants also reported that Bolster was highly usable. All items on the modified System Usability Scale can be found in Table 3. Of the 11 caregivers, 10 (91%) reported that they would recommend Bolster to another caregiver and that they would use Bolster if they had access to it. Most of the caregivers (9/11, 82%) also reported that they were satisfied with Bolster and would like to use it often. Negatively worded items also consistently reflected positive experiences with Bolster because no participant reported finding Bolster to be inconsistent, awkward, or needing lots of training to use. Nearly all participants (10/11, 91%) disagreed with the item stating that Bolster was very complicated. In postintervention qualitative interviews, participants rated (on scales of usefulness and ease of use ranging from 1 to 10) overall Bolster usefulness on average as 8.95 (SD 0.98) out of 10 and gave an average easy-to-use score of 9.36 (SD 0.50) out of 10. Notably, no participant gave an overall easy-to-use rating of <9 or an overall usefulness score of <7. Qualitative comments typically emphasized how participants enjoyed that Bolster was comprehensive and detailed, that it was simple, had a *clean design*, and was easy to use.

Table 3. Participant usability and acceptability ratings (n=11).

Items	Disagree, n (%)	Neutral, n (%)	Agree, n (%)
Acceptability			
"If I have access to Bolster, I will use it"	0 (0)	1 (9)	10 (91)
"I would recommend Bolster to another caregiver"	0 (0)	1 (9)	10 (91)
"I think that I would like to use Bolster often"	0 (0)	2 (18)	9 (82)
"I am satisfied with Bolster"	0 (0)	2 (18)	9 (82)
"I feel I need to have Bolster" (n=10)	1 (10)	2 (20)	7 (70)
"Bolster is fun to use"	1 (9)	4 (36)	6 (55)
Usability			
"Overall, I am satisfied with how easy it is to use Bolster"	0 (0)	0 (0)	11 (100)
"I felt comfortable using Bolster"	0 (0)	0 (0)	11 (100)
"It was easy to learn to use Bolster"	0 (0)	0 (0)	11 (100)
"I found that the different parts of Bolster work well together"	0 (0)	1 (9)	10 (91)
"I would imagine that most people would learn to use Bolster very quickly"	0 (0)	1 (9)	10 (91)
"I felt very confident using Bolster"	0 (0)	1 (9)	10 (91)
"I was able to complete the lessons and practices quickly in Bolster"	0 (0)	1 (9)	10 (91)
"The information provided for Bolster was easy to understand"	0 (0)	1 (9)	10 (91)
"How things appeared on the screen was clear"	0 (0)	1 (9)	10 (91)
"I thought Bolster was easy to use"	0 (0)	2 (18)	9 (82)
"Bolster helped me with caregiving"	0 (0)	2 (18)	9 (82)
"Bolster was interactive enough"	0 (0)	2 (18)	9 (82)
"Whenever I made a mistake using Bolster, I could recover easily and quickly"	0 (0)	3 (27)	8 (73)
"It was easy to find the information I needed" (n=10)	0 (0)	3 (30)	7 (70)
"Bolster works the way I want it to work"	0 (0)	5 (46)	6 (56)
"I found Bolster to be very complicated"[a]	10 (91)	0 (0)	1 (6)
"I think that I would need the support of a technical person to be able to use Bolster"[a]	10 (91)	1 (9)	0 (0)
"I thought there was too much inconsistency in Bolster"[a]	11 (100)	0 (0)	0 (0)
"I found Bolster very awkward to use"[a]	11 (100)	0 (0)	0 (0)
"I needed to learn a lot of things before I could get going with Bolster"[a]	11 (100)	0 (0)	0 (0)

[a]Reverse coded such that disagreement denotes higher perceived usability or acceptability.

Effects on Targets and Outcomes

An exploratory examination of changes from baseline to posttest assessment (Tables 4 and 5) showed improvements in clinical targets. Participants experienced improvements consistent with medium effects in 3 kinds of illness appraisal or knowledge variables—emotional representation (Cohen *d*=0.63), coherence (Cohen *d*=0.55), and consequences (Cohen *d*=0.68)—as well as in coping self-efficacy (Cohen *d*=0.54). They experienced small improvements in coping skills practiced (Cohen *d*=0.27) and appraisals of caregiving experiences as positive (Cohen *d*=0.26). With regard to primary outcomes, participants experienced large improvements in overall distress (Cohen *d*=1.77) and medium-level improvements in expressed emotion (Cohen *d*=0.52). All assessed outcomes moved in the direction associated with improvement, with the exception of illness knowledge and appraisals related to the *controllability* of psychosis, each of which did not seem to change during the study period. Notably, illness knowledge scores seemed to be affected by ceiling effects because many participants received a high score on the KAST at baseline (mean 16.72 correct out of 20 compared with psychometric work on the initial development of the KAST [49] suggesting typical average scores among family members of 10.9 out of 20 or among lay community members of 9.3 out of 20).

Table 4. Baseline and posttest scores of intervention targets.

Variable[a]	Measure	Baseline score, mean (SD)	Posttest score, mean (SD)	2-tailed		
				t test (*df*)	*P* value	Cohen *d*[b]
Illness knowledge	KAST[c]	16.72 (1.19)	16.63 (2.01)	−0.28 (10)	.78	−0.09
Illness knowledge	IPQSR[d]: coherence	10.18 (2.64)	8.82 (2.56)	−1.81 (10)	.10	0.55
Illness appraisals	IPQSR: emotional representation	31.90 (5.45)	29.82 (5.91)	−2.10 (10)	.06	0.63
Illness appraisals	IPQSR: consequences	71.72 (8.54)	68.27 (10.26)	−2.03 (10)	.07	0.68
Illness appraisals	IPQSR: control	33.00 (2.76)	32.72 (2.97)	−0.29 (10)	.78	−0.09
Caregiving appraisals	ECI[e]: negative	77.55 (33.72)	75.58 (33.81)	−0.62 (10)	.55	0.19
Caregiving appraisals	ECI: positive	34.90 (9.57)	37.18 (7.47)	0.85 (10)	.42	0.26
Coping	Brief COPE[f]: positive coping	23.97 (9.08)	25.86 (7.20)	0.91 (10)	.39	0.27
Coping	CSES[g]: total	168.18 (59.10)	199.57 (27.77)	1.79 (10)	.10	0.54

[a]Mean imputation (by factor-analytically defined subscale) was used to replace missing values.

[b]Cohen *d* values are scaled such that a positive value denotes movement in the hypothesized direction.

[c]KAST: Knowledge About Schizophrenia Test.

[d]IPQSR: Illness Perception Questionnaire for Schizophrenia: Relatives' version.

[e]ECI: Experiences of Caregiving Inventory.

[f]Brief COPE: Brief Coping Orientation to Problems Experienced Inventory.

[g]CSES: Coping Self-Efficacy Scale.

Table 5. Baseline and posttest scores of intervention outcomes.

Variable[a]	Measure	Baseline score, mean (SD)	Posttest score, mean (SD)	2-tailed		
				t test (*df*)	*P* value	Cohen *d*[b]
Distress	GHQ-12[c]	17.48 (3.49)	11.35 (2.34)	−5.88 (*10*)	<.001	1.77
Expressed emotion	FQ[d]	50.22 (11.14)	46.64 (10.92)	−1.72 (*10*)	.12	0.52

[a]Mean imputation (by factor-analytically defined subscale) was used to replace missing values.

[b]Cohen *d* values are scaled such that a positive value denotes movement in the hypothesized direction.

[c]GHQ-12: General Health Questionnaire, 12-item version.

[d]FQ: Family Questionnaire.

Discussion

Principal Findings

mHealth could play a useful role in supporting caregivers of young adults at risk for psychosis. mHealth interventions are scalable, acceptable, and can provide in-the-moment support well suited to the challenges of caregiving. Our team developed, optimized, and tested Bolster, an mHealth intervention designed in response to the needs and preferences identified by members of this population. In a qualitative needs assessment, caregivers expressed interest in a tool that provided actionable guidance, opportunities for personalization, and the ability to track changes over time. Prototyping revealed preferences for static text and video (rather than messaging-style interactions), personalization, and interactive features. The intervention that was optimized based on this feedback—Bolster—provides written and video-based lessons and practices focused on psychoeducation, communication skills, and self-care, as well as assessments to track changes in the relative's symptoms and links to curated resources related to psychosis, treatment, and treatment listings. In a 1-week usability field trial, Bolster seemed promising. Participants found it to be acceptable, usable, and reported marked changes in their appraisals of illness, distress, and expressed emotion. Given the nature of this study as a usability field trial, the lack of a control condition, and the absence of hypotheses about clinical targets and outcomes, the results related to intervention effects are highly preliminary and should be interpreted with caution. However, they do seem to reflect that providing mHealth specific to caregiving support is promising and that additional, larger, controlled studies are warranted.

Our systematic Bolster development process led to several clear conclusions about the needs and preferences of caregivers. First, participant responses suggested needs in a few key areas: psychoeducation, information about seeking services, coping, and communication skills. Second, many caregivers expressed

frustration with the lack of *actionable* information available on the internet today and described finding general information about psychosis that did not provide clear guidance in how to respond to caregiving-related challenges. This reflects previous qualitative work in which caregivers expressed the desire for step-by-step guides or instructional resources [4]. Others were disappointed when they found information that they found demoralizing or discouraging about psychosis and its prognosis. Previous literature suggests that caregivers can receive stigmatizing reactions from general practitioners and hospital staff, and the ongoing stigma of psychotic illness persists as a barrier to information and treatment seeking [60,61].

Initial prototype usability testing led to several key principles for the optimization of Bolster. Most notably, the initial messaging-style interface of the prototype seemed to have several limitations, including difficulties in pacing, confusion about the messaging-style interface, and the lack of a mechanism for highlighting key information. It is unclear based on this feedback whether these identify generalizable limitations with messaging-style interventions or chatbots in this population or whether these comments were specific to our prototype. Many other examples of rule-based mental health chatbots have demonstrated promising usability [62,63], and recent and ongoing developments in language-based artificial intelligence technologies [64-66] will likely lead to concurrent innovations that facilitate positive user experiences with these tools. By contrast, many participants reacted positively to interactive text modules wherein they could enter particulars related to their caregiving situation. Such interactive features seem to be linked with improvements in help-seeking behaviors in studies of interventions of this construct [67]. Additional usability changes responded to user comments related to clear expectation setting (eg, time estimates and a summary statement describing modules) and changes to navigation (eg, consistent back, home, and next buttons). The subsequent field trial suggested that the user-centered design process was effective. All participants who completed baseline assessments engaged in the intervention and were retained at 1-week follow-up. Participants provided highly positive usability feedback: nearly all participants (10/11, 91%) reported that they would use Bolster if they had continued access and that they would recommend it to a fellow caregiver. Although exploratory, preliminary analyses of clinical effects unveiled a promising result and suggest that a sufficiently powered trial of Bolster is warranted. Participants provided scores indicating changes in illness knowledge, appraisals, coping self-efficacy, distress, and expressed emotion. Other caregiver-focused interventions, including those delivered in person by providers [68,69] or remotely in asynchronous web-based interactions [70,71], have demonstrated similar promising effects; however, as a fully self-guided mobile intervention, Bolster may have several advantages: it is scalable, can be used repeatedly over long periods of time, and is well suited for in-the-moment needs.

Limitations

This study has several limitations. First, our user-centered development process was meant to solicit detailed guidance from caregivers with lived experience; study conclusions are drawn from small samples and may not generalize to all members of this population. Furthermore, given the fact that recruitment occurred primarily on the internet, participants may be drawn from a population that is already predisposed to digital tools. Second, design objectives for Bolster were constrained to some extent by predefined project goals. Bolster was a priori proposed as a self-guided mobile intervention. Thus, caregiver feedback related to intervention features that could require clinical support (eg, an expert question-and-answer feature) or social features (eg, peer forums) was incorporated into Bolster in a manner consistent with these a priori constraints (eg, providing clinician videos rather than synchronous communication and links to support websites rather than the development of new tools). Third, field trial results may speak primarily to caregiver experiences during a test period rather than the real-world deployment of Bolster. Participants were likely motivated to provide helpful feedback, and this likely affected their frequency of use of the intervention, particularly given the fact that the field trial lasted just 1 week. Interactions with the study team also likely encouraged use. This is particularly the case given the fact that participants received outreach to ensure that the app was functioning. The conditions of a usability field trial may not fully reflect how such an intervention would be deployed under real-world conditions. Finally, analysis of clinical effects was exploratory. The trial was brief and conducted in a small sample of caregivers already connected to treatment.

Conclusions

The development of Bolster responds to the identified lack of mobile interventions designed for caregivers [40]. Our team's development of Bolster was grounded in user-centered design and thus seems well suited for future testing to determine whether mHealth can assist caregivers in their critical roles of early identification, treatment facilitation, and ongoing support. These results add to the growing body of literature supporting the use of digital technologies to reduce barriers to treatment and recovery for families affected by psychosis. Bolster seems to respond to specific user-identified interest in more structured, directive, and concrete guidance, as well as support that attends to the emotional experience of caregiving. Future work will examine whether Bolster is an effective tool to support early psychosis caregivers, and our findings here suggest that similar digital approaches to support caregivers of individuals experiencing mental health conditions may have promise. These results demonstrate promise regarding the use of mHealth to support caregivers who—in many settings—lack structured supports to meet their own needs.

Acknowledgments

This study was supported by an R34 exploratory research grant (R34MH124878; principal investigator: BB) from the National Institute of Mental Health. BB is supported by a K23 Mentored Patient-Oriented Research Career Development Award from the National Institute of Mental Health (K23MH122504; principal investigator: BB).

Conflicts of Interest

DB-Z has financial interests in Merlin LLC and FOCUS Technology. He has provided consultation services to Trusst Health, K Health, Boehringer Ingelheim, eQuility, Deep Valley Labs, and Otsuka Pharmaceutical. All other authors declare no other conflicts of interest.

References

1. Kuipers E. Time for a separate psychosis caregiver service? J Ment Health 2010 Oct 13;19(5):401-404. [doi: 10.3109/09638237.2010.510155] [Medline: 20836686]
2. Anderson KK, Fuhrer R, Malla AK. "There are too many steps before you get to where you need to be": help-seeking by patients with first-episode psychosis. J Ment Health 2013 Aug 07;22(4):384-395. [doi: 10.3109/09638237.2012.705922] [Medline: 22958140]
3. Cotton SM, McCann TV, Gleeson JF, Crisp K, Murphy BP, Lubman DI. Coping strategies in carers of young people with a first episode of psychosis. Schizophr Res 2013 May;146(1-3):118-124. [doi: 10.1016/j.schres.2013.02.008] [Medline: 23490761]
4. Cheng SC, Backonja U, Buck B, Monroe-DeVita M, Walsh E. Facilitating pathways to care: a qualitative study of the self-reported needs and coping skills of caregivers of young adults diagnosed with early psychosis. J Psychiatr Ment Health Nurs 2020 Aug 11;27(4):368-379. [doi: 10.1111/jpm.12591] [Medline: 31930633]
5. Hansen H, Stige SH, Davidson L, Moltu C, Veseth M. How do people experience early intervention services for psychosis? a meta-synthesis. Qual Health Res 2018 Jan 17;28(2):259-272. [doi: 10.1177/1049732317735080] [Medline: 29039239]
6. Addington J, van Mastrigt S, Hutchinson J, Addington D. Pathways to care: help seeking behaviour in first episode psychosis. Acta Psychiatr Scand 2002 Nov;106(5):358-364. [doi: 10.1034/j.1600-0447.2002.02004.x] [Medline: 12366470]
7. Anderson KK, Fuhrer R, Malla AK. The pathways to mental health care of first-episode psychosis patients: a systematic review. Psychol Med 2010 Oct;40(10):1585-1597. [doi: 10.1017/S0033291710000371] [Medline: 20236571]
8. Lucksted A, Stevenson J, Nossel I, Drapalski A, Piscitelli S, Dixon LB. Family member engagement with early psychosis specialty care. Early Interv Psychiatry 2018 Oct;12(5):922-927 [FREE Full text] [doi: 10.1111/eip.12403] [Medline: 27863039]
9. Stowkowy J, Addington D, Liu L, Hollowell B, Addington J. Predictors of disengagement from treatment in an early psychosis program. Schizophr Res 2012 Apr;136(1-3):7-12. [doi: 10.1016/j.schres.2012.01.027] [Medline: 22336955]
10. Morgan C, Mallett R, Hutchinson G, Bagalkote H, Morgan K, Fearon P, et al. Pathways to care and ethnicity. 2: source of referral and help-seeking. report from the AESOP study. Br J Psychiatry 2005 Apr 02;186(4):290-296. [doi: 10.1192/bjp.186.4.290] [Medline: 15802684]
11. Revier CJ, Reininghaus U, Dutta R, Fearon P, Murray RM, Doody GA, et al. Ten-year outcomes of first-episode psychoses in the MRC ÆSOP-10 study. J Nerv Ment Dis 2015 May;203(5):379-386 [FREE Full text] [doi: 10.1097/NMD.0000000000000295] [Medline: 25900547]
12. Lee G, Barrowclough C, Lobban F. Positive affect in the family environment protects against relapse in first-episode psychosis. Soc Psychiatry Psychiatr Epidemiol 2014 Mar 1;49(3):367-376. [doi: 10.1007/s00127-013-0768-x] [Medline: 24081324]
13. Wong DT, Tong SF, Daud TI, Aziz SA, Midin M. Factors influencing professional help-seeking behavior during first episode psychosis in schizophrenia: an exploratory study on caregivers' perspective. Front Psychiatry 2019 Feb 14;10:962 [FREE Full text] [doi: 10.3389/fpsyt.2019.00962] [Medline: 32116809]
14. Skubby D, Bonfine N, Tracy H, Knepp K, Munetz MR. The help-seeking experiences of parents of children with a first-episode of psychosis. Community Ment Health J 2015 Nov 17;51(8):888-896. [doi: 10.1007/s10597-015-9877-1] [Medline: 25876766]
15. Onwumere J, Lotey G, Schulz J, James G, Afsharzadegan R, Harvey R, et al. Burnout in early course psychosis caregivers: the role of illness beliefs and coping styles. Early Interv Psychiatry 2017 Jun 26;11(3):237-243. [doi: 10.1111/eip.12227] [Medline: 25721376]
16. Kuipers E, Onwumere J, Bebbington P. Cognitive model of caregiving in psychosis. Br J Psychiatry 2010 Apr;196(4):259-265. [doi: 10.1192/bjp.bp.109.070466] [Medline: 20357299]
17. Pirkis J, Burgess P, Hardy J, Harris M, Slade T, Johnston A. Who cares? a profile of people who care for relatives with a mental disorder. Aust N Z J Psychiatry 2010 Oct 01;44(10):929-937. [doi: 10.3109/00048674.2010.493858] [Medline: 20932207]

18. López SR, Gamez D, Mejia Y, Calderon V, Lopez D, Ullman JB, et al. Psychosis literacy among Latinos with first-episode psychosis and their caregivers. Psychiatr Serv 2018 Nov 01;69(11):1153-1159 [FREE Full text] [doi: 10.1176/appi.ps.201700400] [Medline: 30220245]
19. McCann TV, Lubman DI, Clark E. Responding to stigma: first-time caregivers of young people with first-episode psychosis. Psychiatr Serv 2011 May;62(5):548-550. [doi: 10.1176/ps.62.5.pss6205_0548] [Medline: 21532084]
20. Fortune DG, Smith JV, Garvey K. Perceptions of psychosis, coping, appraisals, and psychological distress in the relatives of patients with schizophrenia: an exploration using self-regulation theory. Br J Clin Psychol 2005 Sep;44(Pt 3):319-331. [doi: 10.1348/014466505X29198] [Medline: 16238880]
21. Izon E, Berry K, Law H, French P. Expressed emotion (EE) in families of individuals at-risk of developing psychosis: a systematic review. Psychiatry Res 2018 Dec;270:661-672. [doi: 10.1016/j.psychres.2018.10.065] [Medline: 30384287]
22. Leff J, Vaughn C. The role of maintenance therapy and relatives' expressed emotion in relapse of schizophrenia: a two-year follow-up. Br J Psychiatry 1981 Aug 29;139(2):102-104. [doi: 10.1192/bjp.139.2.102] [Medline: 7306748]
23. Sadath A, Muralidhar D, Varambally S, Gangadhar BN, Jose JP. Do stress and support matter for caring? the role of perceived stress and social support on expressed emotion of carers of persons with first episode psychosis. Asian J Psychiatr 2017 Feb;25:163-168. [doi: 10.1016/j.ajp.2016.10.023] [Medline: 28262142]
24. Peterson EC, Docherty NM. Expressed emotion, attribution, and control in parents of schizophrenic patients. Psychiatry 2004;67(2):197-207. [doi: 10.1521/psyc.67.2.197.35959] [Medline: 15262588]
25. Kuipers E, Bebbington P, Dunn G, Fowler D, Freeman D, Watson P, et al. Influence of carer expressed emotion and affect on relapse in non-affective psychosis. Br J Psychiatry 2006 Feb;188(2):173-179. [doi: 10.1192/bjp.bp.104.007294] [Medline: 16449706]
26. Finnegan D, Onwumere J, Green C, Freeman D, Garety P, Kuipers E. Negative communication in psychosis: understanding pathways to poorer patient outcomes. J Nerv Ment Dis 2014 Nov;202(11):829-832. [doi: 10.1097/NMD.0000000000000204] [Medline: 25357253]
27. Vasconcelos E Sa D, Wearden A, Hartley S, Emsley R, Barrowclough C. Expressed Emotion and behaviourally controlling interactions in the daily life of dyads experiencing psychosis. Psychiatry Res 2016 Nov 30;245:406-413. [doi: 10.1016/j.psychres.2016.08.060] [Medline: 27611070]
28. Haidl T, Rosen M, Schultze-Lutter F, Nieman D, Eggers S, Heinimaa M, et al. Expressed emotion as a predictor of the first psychotic episode - results of the European Prediction of Psychosis Study. Schizophr Res 2018 Sep;199:346-352. [doi: 10.1016/j.schres.2018.03.019] [Medline: 29661524]
29. Weintraub MJ, Hall DL, Carbonella JY, Weisman de Mamani A, Hooley JM. Integrity of literature on expressed emotion and relapse in patients with schizophrenia verified by a p-curve analysis. Fam Process 2017 Jun;56(2):436-444 [FREE Full text] [doi: 10.1111/famp.12208] [Medline: 26875506]
30. Rutter M, Brown GW. The reliability and validity of measures of family life and relationships in families containing a psychiatric patient. Soc Psychiatry 1966 Aug;1(1):38-53. [doi: 10.1007/bf00583828]
31. Bebbington P, Kuipers L. The predictive utility of expressed emotion in schizophrenia: an aggregate analysis. Psychol Med 1994 Aug 09;24(3):707-718. [doi: 10.1017/s0033291700027860] [Medline: 7991753]
32. Alvarez-Jimenez M, Priede A, Hetrick SE, Bendall S, Killackey E, Parker AG, et al. Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies. Schizophr Res 2012 Aug;139(1-3):116-128. [doi: 10.1016/j.schres.2012.05.007] [Medline: 22658527]
33. Buck B, Chander A, Monroe-DeVita M, Cheng SC, Stiles B, Ben-Zeev D. Mobile health for caregivers of young adults with early psychosis: a survey study examining user preferences. Psychiatr Serv 2021 Aug 01;72(8):955-959 [FREE Full text] [doi: 10.1176/appi.ps.202000193] [Medline: 34235943]
34. Vermeulen B, Lauwers H, Spruytte N, Van Audenhove C. P.1.k.038 Experiences of family caregivers for persons with severe mental illness: an international exploration. European Neuropsychopharmacology 2015 Sep;25:S374 [FREE Full text] [doi: 10.1016/s0924-977x(15)30484-3]
35. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs): an organizing framework for ongoing health behavior support. Methodol Center Tech Rep 2014;073975(14):1-37 [FREE Full text]
36. The Nielsen total audience report: Q1 2018. Nielsen. 2018 Jul. URL: https://www.nielsen.com/insights/2018/q1-2018-total-audience-report/ [accessed 2023-09-24]
37. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
38. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? a meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
39. Schlosser DA, Campellone TR, Truong B, Etter K, Vergani S, Komaiko K, et al. Efficacy of PRIME, a mobile app intervention designed to improve motivation in young people with schizophrenia. Schizophr Bull 2018 Aug 20;44(5):1010-1020 [FREE Full text] [doi: 10.1093/schbul/sby078] [Medline: 29939367]

40. Onwumere J, Amaral F, Valmaggia LR. Digital technology for caregivers of people with psychosis: systematic review. JMIR Ment Health 2018 Sep 05;5(3):e55 [FREE Full text] [doi: 10.2196/mental.9857] [Medline: 30185402]
41. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013 Dec;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]
42. Ben-Zeev D, Meller S, Snyder J, Attah D, Albright L, Le H, et al. A digital toolkit (M-Healer) to improve care and reduce human rights abuses against people with mental illness in West Africa: user-centered design, development, and usability study. JMIR Ment Health 2021 Jul 02;8(7):e28526 [FREE Full text] [doi: 10.2196/28526] [Medline: 34255712]
43. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
44. Schlosser D, Campellone T, Kim D, Truong B, Vergani S, Ward C, et al. Feasibility of PRIME: a cognitive neuroscience-informed mobile app intervention to enhance motivated behavior and improve quality of life in recent onset schizophrenia. JMIR Res Protoc 2016 Apr 28;5(2):e77 [FREE Full text] [doi: 10.2196/resprot.5450] [Medline: 27125771]
45. Kline E, Thompson E, Schimunek C, Reeves G, Bussell K, Pitts SC, et al. Parent-adolescent agreement on psychosis risk symptoms. Schizophr Res 2013 Jun;147(1):147-152. [doi: 10.1016/j.schres.2013.03.007] [Medline: 23570897]
46. Craig TJ, Bromet EJ, Fennig S, Tanenberg-Karant M, Lavelle J, Galambos N. Is there an association between duration of untreated psychosis and 24-month clinical outcome in a first-admission series? Am J Psychiatry 2000 Jan;157(1):60-66. [doi: 10.1176/ajp.157.1.60] [Medline: 10618014]
47. Lappin JM, Morgan KD, Morgan C, Dazzan P, Reichenberg A, Zanelli JW, et al. Duration of untreated psychosis and neuropsychological function in first episode psychosis. Schizophr Res 2007 Sep;95(1-3):103-110. [doi: 10.1016/j.schres.2007.05.037] [Medline: 17644343]
48. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
49. Compton MT, Quintero L, Esterberg ML. Assessing knowledge of schizophrenia: development and psychometric properties of a brief, multiple-choice knowledge test for use across various samples. Psychiatry Res 2007 May 30;151(1-2):87-95. [doi: 10.1016/j.psychres.2006.05.019] [Medline: 17379321]
50. Lobban F, Barrowclough C, Jones S. Assessing cognitive representations of mental health problems. II. The illness perception questionnaire for schizophrenia: relatives' version. Br J Clin Psychol 2005 Jun;44(Pt 2):163-179. [doi: 10.1348/014466504X19785] [Medline: 16004652]
51. Szmukler GI, Burgess P, Herrman H, Benson A, Colusa S, Bloch S. Caring for relatives with serious mental illness: the development of the Experience of Caregiving Inventory. Soc Psychiatry Psychiatr Epidemiol 1996 Jun;31(3-4):137-148. [doi: 10.1007/BF00785760] [Medline: 8766459]
52. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
53. Chesney MA, Neilands TB, Chambers DB, Taylor JM, Folkman S. A validity and reliability study of the coping self-efficacy scale. Br J Health Psychol 2006 Sep;11(Pt 3):421-437 [FREE Full text] [doi: 10.1348/135910705X53155] [Medline: 16870053]
54. Wiedemann G, Rayki O, Feinstein E, Hahlweg K. The Family Questionnaire: development and validation of a new self-report scale for assessing expressed emotion. Psychiatry Res 2002 Apr 15;109(3):265-279. [doi: 10.1016/s0165-1781(02)00023-9] [Medline: 11959363]
55. Goldberg DP, Blackwell B. Psychiatric illness in general practice. a detailed study using a new method of case identification. Br Med J 1970 May 23;1(5707):439-443 [FREE Full text] [doi: 10.1136/bmj.2.5707.439] [Medline: 5420206]
56. Hardy GE, Shapiro DA, Haynes CE, Rick JE. Validation of the General Health Questionnaire-12: using a sample of employees from England's health care services. Psychological Assessment 1999 Jun;11(2):159-165. [doi: 10.1037/1040-3590.11.2.159]
57. Wang L, Lin W. Wording effects and the dimensionality of the General Health Questionnaire (GHQ-12). Pers Individ Dif 2011 May;50(7):1056-1061 [FREE Full text] [doi: 10.1016/j.paid.2011.01.024]
58. Ritsner MS, Mar M, Arbitman M, Grinshpoon A. Symptom severity scale of the DSM5 for schizophrenia, and other psychotic disorders: diagnostic validity and clinical feasibility. Psychiatry Res 2013 Jun 30;208(1):1-8. [doi: 10.1016/j.psychres.2013.02.029] [Medline: 23582209]
59. American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, 5th edition. New York City, NY: American Psychiatric Association; 2013.
60. Gerson R, Davidson L, Booty A, McGlashan T, Malespina D, Pincus HA, et al. Families' experience with seeking treatment for recent-onset psychosis. Psychiatr Serv 2009 Jun;60(6):812-816 [FREE Full text] [doi: 10.1176/ps.2009.60.6.812] [Medline: 19487352]
61. McCann TV, Lubman DI, Clark E. First-time primary caregivers' experience of caring for young adults with first-episode psychosis. Schizophr Bull 2011 Mar;37(2):381-388 [FREE Full text] [doi: 10.1093/schbul/sbp085] [Medline: 19679716]

62. Shah J, DePietro B, D'Adamo L, Firebaugh M, Laing O, Fowler LA, et al. Development and usability testing of a chatbot to promote mental health services use among individuals with eating disorders following screening. Int J Eat Disord 2022 Sep 18;55(9):1229-1244 [FREE Full text] [doi: 10.1002/eat.23798] [Medline: 36056648]
63. Greer S, Ramo D, Chang Y, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]
64. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
65. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 04;47(1):33 [FREE Full text] [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
66. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
67. Evans-Lacko S, Hahn JS, Peter LJ, Schomerus G. The impact of digital interventions on help-seeking behaviour for mental health problems: a systematic literature review. Curr Opin Psychiatry 2022 May 01;35(3):207-218. [doi: 10.1097/YCO.0000000000000788] [Medline: 35579875]
68. McFarlane WR, Dixon L, Lukens E, Lucksted A. Family psychoeducation and schizophrenia: a review of the literature. J Marital Fam Ther 2003 Apr;29(2):223-245. [doi: 10.1111/j.1752-0606.2003.tb01202.x] [Medline: 12728780]
69. Lucksted A, McFarlane W, Downing D, Dixon L. Recent developments in family psychoeducation as an evidence-based practice. J Marital Fam Ther 2012 Jan;38(1):101-121. [doi: 10.1111/j.1752-0606.2011.00256.x] [Medline: 22283383]
70. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010 Nov;61(11):1099-1105. [doi: 10.1176/ps.2010.61.11.1099] [Medline: 21041348]
71. Kopelovich SL, Stiles B, Monroe-DeVita M, Hardy K, Hallgren K, Turkington D. Psychosis REACH: effects of a brief CBT-informed training for family and caregivers of individuals with psychosis. Psychiatr Serv 2021 Nov 01;72(11):1254-1260. [doi: 10.1176/appi.ps.202000740] [Medline: 34015942]

Abbreviations

Brief-COPE: Brief Coping Orientation to Problems Experienced Inventory
CGPS-R: Caregiver Prime Screen–Revised
CSES: Coping Self-Efficacy Scale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ECI: Experiences of Caregiving Inventory
FQ: Family Questionnaire
GHQ-12: General Health Questionnaire, 12-item version
IPQSR: Illness Perception Questionnaire for Schizophrenia: Relatives' version
KAST: Knowledge About Schizophrenia Test
mHealth: mobile health
NAMI: National Alliance on Mental Illness
PEPPNET: Psychosis-Risk and Early Psychosis Program Network

Edited by J Torous; submitted 05.07.23; peer-reviewed by W Tumuhimbise, G Strudwick; comments to author 15.08.23; revised version received 09.10.23; accepted 21.10.23; published 30.11.23.

Please cite as:
Buck B, Wingerson M, Whiting E, Snyder J, Monroe-DeVita M, Ben-Zeev D
User-Centered Development of Bolster, an mHealth Intervention for Early Psychosis Caregivers: Needs Assessment, Prototyping, and Field Trial
JMIR Ment Health 2023;10:e50522
URL: https://mental.jmir.org/2023/1/e50522
doi:10.2196/50522
PMID:38032692

Original Paper

The Feasibility of Implementing Remote Measurement Technologies in Psychological Treatment for Depression: Mixed Methods Study on Engagement

Valeria de Angel[1,2], BSc, MSc; Fadekemi Adeleye[3], BSc; Yuezhou Zhang[4], MSc; Nicholas Cummins[4], PhD; Sara Munir[5], BSc; Serena Lewis[1,6], BSc; Estela Laporta Puyal[7,8], MSc; Faith Matcham[1,9], CPsychol, PhD; Shaoxiong Sun[4], PhD; Amos A Folarin[2,4,10,11,12], PhD; Yatharth Ranjan[13], MSc; Pauline Conde[13], BSc; Zulqarnain Rashid[13], PhD; Richard Dobson[1,2], PhD; Matthew Hotopf[1,2], PhD

[1]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[2]NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, United Kingdom

[3]Department of Psychology, King's College London, London, United Kingdom

[4]Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[5]Lewisham Talking Therapies, South London and Maudsley NHS Foundation Trust, London, United Kingdom

[6]Department of Psychology, University of Bath, Bath, United Kingdom

[7]Biomedical Signal Interpretation and Computational Simulation Group, Aragón Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, Zaragoza, Spain

[8]Centro de Investigación Biomédica en Red of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain

[9]School of Psychology, University of Sussex, Brighton, United Kingdom

[10]Institute of Health Informatics, University College London, London, United Kingdom

[11]Health Data Research UK London, University College London, London, United Kingdom

[12]NIHR Biomedical Research Centre at University College London Hospitals, University College London Hospitals NHS Foundation Trust, London, United Kingdom

[13]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

Corresponding Author:
Valeria de Angel, BSc, MSc
Department of Psychological Medicine
Institute of Psychiatry, Psychology and Neuroscience
King's College London
E3.08, 3rd Floor East Wing
de Crespigny park
London, SE5 8AF
United Kingdom
Phone: 44 20 7848 0002
Email: valeria.de_angel@kcl.ac.uk

Abstract

Background: Remote measurement technologies (RMTs) such as smartphones and wearables can help improve treatment for depression by providing objective, continuous, and ecologically valid insights into mood and behavior. Engagement with RMTs is varied and highly context dependent; however, few studies have investigated their feasibility in the context of treatment.

Objective: A mixed methods design was used to evaluate engagement with active and passive data collection via RMT in people with depression undergoing psychotherapy. We evaluated the effects of treatment on 2 different types of engagement: study attrition (engagement with study protocol) and patterns of missing data (engagement with digital devices), which we termed data availability. Qualitative interviews were conducted to help interpret the differences in engagement.

Methods: A total of 66 people undergoing psychological therapy for depression were followed up for 7 months. Active data were gathered from weekly questionnaires and speech and cognitive tasks, and passive data were gathered from smartphone sensors and a Fitbit (Fitbit Inc) wearable device.

Results: The overall retention rate was 60%. Higher-intensity treatment (χ^2_1=4.6; *P*=.03) and higher baseline anxiety ($t_{56.28}$=−2.80, 2-tailed; *P*=.007) were associated with attrition, but depression severity was not ($t_{50.4}$=−0.18; *P*=.86). A trend toward significance was found for the association between longer treatments and increased attrition (*U*=339.5; *P*=.05). Data availability was higher for active data than for passive data initially but declined at a sharper rate (90%-30% drop in 7 months). As for passive data, wearable data availability fell from a maximum of 80% to 45% at 7 months but showed higher overall data availability than smartphone-based data, which remained stable at the range of 20%-40% throughout. Missing data were more prevalent among GPS location data, followed by among Bluetooth data, then among accelerometry data. As for active data, speech and cognitive tasks had lower completion rates than clinical questionnaires. The participants in treatment provided less Fitbit data but more active data than those on the waiting list.

Conclusions: Different data streams showed varied patterns of missing data, despite being gathered from the same device. Longer and more complex treatments and clinical characteristics such as higher baseline anxiety may reduce long-term engagement with RMTs, and different devices may show opposite patterns of missingness during treatment. This has implications for the scalability and uptake of RMTs in health care settings, the generalizability and accuracy of the data collected by these methods, feature construction, and the appropriateness of RMT use in the long term.

(JMIR Ment Health 2023;10:e42866) doi:10.2196/42866

KEYWORDS

depression; anxiety; digital health; wearable devices; smartphone; passive sensing; mobile health; mHealth; digital phenotyping; mobile phone

Introduction

Background

Depression is a leading cause of disability, with associated physical comorbidities and increased health care costs [1,2]. Psychological therapy is a recommended first-line treatment for mild to moderate depression [3,4]; however, approximately 50% of people do not recover following intervention [5,6]. Remote measurement technologies (RMTs) such as smartphones and wearables may assist in the treatment of depression to improve patient outcomes by detecting changes in its key behavioral aspects.

RMTs, by generating unobtrusive, continuous, and objective measures of behavior and physiology, could overcome the pitfalls of the current clinical outcome measurements, which rely on patient recall and infrequent symptom scales. Furthermore, they could help establish clinical objectives for treatment, such as a target amount of physical activity or regular sleep-time schedule and serve as indicators of whether treatments targeting particular behaviors have been effective. RMTs may also uncover digital phenotypes to identify people who are more or less responsive to certain treatments, paving the way for increased personalization of mental health care [7]. Finally, RMTs could improve patient and clinician experience of psychotherapy by strengthening communication, helping support the emotional and cognitive needs of patients and enhancing self-awareness [8].

RMTs generally apply 2 types of data collection methods: *active* and *passive*. *Active* data collection requires conscious user engagement, such as responding to mood scales, questionnaires, or speech tasks delivered to a participant's phone. *Passive* data collection refers to the automatic capture of information via device-embedded sensors that require minimal input from users [9]; for example, accelerometers on a fitness tracker automatically detect physical activity. Used in combination, active and passive monitoring provide a way to capture continuous, ecologically valid, and high-resolution measures of signs and symptoms related to depression.

The extent to which these methods can be successfully implemented in health care and used in treatment depends on their feasibility and acceptability as tools for collecting longitudinal data in clinical populations. The feasibility of using RMTs is generally evaluated by measuring 2 broad parameters of engagement: attrition from longitudinal studies and data availability, which is the amount of usable data contributed by individuals through task completion or device use and, therefore, the opposite of missing data [10].

The measurement and reporting of attrition is relevant not only because attrition threatens the generalizability of longitudinal studies but also because it informs implementation efforts by mirroring the potential uptake and engagement within clinical settings. Much of the current research on attrition focuses on active data collection, with passive sensing being underreported. In general, studies have short follow-up periods, with systematic reviews finding a median follow-up period of 7 days for active data [11] and between 7 and 14 days for passive data [12], limiting their ability to be generalized to psychotherapy contexts, which usually span weeks. In addition, the context in which data collection occurs is key to understanding the difference in attrition rates between RMT studies. For example, a review of self-referral studies found, on average, 50% attrition in the first 15 days and varied retention rates depending on factors such as the presence and type of illness studied [13], whereas clinical trials on digital-based psychotherapy found similar attrition but at a much slower pace [14]. By contrast, large studies with dedicated recruitment resources have achieved attrition rates as low as 20% even if follow-up sessions were conducted after 2 years [15]. Therefore, if implementation of RMTs within health care is the aim, research on long-term attrition in active and passive data collection in the context of psychotherapy is critical.

Work on data availability has generally focused on active approaches [16-18], leaving passive sensing underresearched

and underreported [12]. Given that both approaches require varying amounts of input and commitment from the user, *missingness* is likely to vary in the extent to which it occurs at random and may be differentially affected by individual differences [19]. This, in turn, has implications for the integrity of the constructed variables and for understanding the potential sources of biases in the data. Sparse active data points on mood questionnaires can affect how ground truth is determined, whereas less passive data can result in inaccuracies in how features are derived and the resulting data analysis (refer to Currey and Torous [20] for an example of this).

Objective

We sought to explore the feasibility of using RMTs in a clinical setting to help uncover potential implementation and scaling issues, the resolution of which is crucial for widespread adoption. This study used a mixed methods design to evaluate the long-term engagement with active and passive approaches to the remote monitoring of mood and behavior in people with depression undergoing psychotherapy. Applying the framework developed by White et al [10], we focused on 2 forms of engagement as feasibility aspects of interest. The aims were (1) to measure engagement with the research protocol through recruitment and attrition rates, (2) measure engagement with RMTs through passive and active data availability rates and identify data streams that are more vulnerable to missing data, (3) assess the possible effect of treatment on both types of engagement, and (4) use the information gathered from qualitative interviews to aid in the explanation of the quantitative engagement data.

Methods

Study Design

This study was a fully remote, mixed methods, prospective cohort study with repeated measures over a 7-month period, designed to evaluate the feasibility and acceptability of using remote data collection methods in people undergoing treatment; the full protocol has been reported elsewhere [21]. The quantitative measures included recurrent clinical questionnaires and continuous digital sensor data. Qualitative measures comprised semistructured interviews that adopted an inductive approach to thematic analysis.

Recruitment and Setting

Participants were drawn from Improving Access to Psychological Therapies (IAPT) services in South London and Maudsley National Health Service Foundation Trust, United Kingdom, a publicly funded outpatient program providing psychological treatments for adults with mild to moderate mental health disorders. IAPT services provide treatment at both high and low intensities (refer to Table S1 in Multimedia Appendix 1 for details), the allocation of which is based on several factors, including patient needs, preferences, and diagnosis. High-intensity therapy comprises approximately 10 to 12 sessions, whereas low-intensity therapy comprises approximately 6 to 8 sessions. These are usually delivered 1 week apart and can be web-based or face-to-face, depending on clinician availability and patient preference.

Sample

A total of 66 treatment-seeking adults with depression were recruited from their local IAPT services' waiting list, which provided the study information, and screened for eligibility either over the phone by a researcher or through a web-based self-screening tool. The sample size was determined by the primary aims and followed the general recommendation for samples of 50 to 60 participants to assess feasibility outcomes [22]. Recruitment and data collection were conducted between June 2020 and March 2022. We included adults with a current episode of depression, as measured by the Mini International Neuropsychiatric Interview [23], who owned and did not extensively share an Android (Google LLC) smartphone and were able and willing to use a wrist-worn device for the duration of the study. The exclusion criteria included a lifetime diagnosis of bipolar disorder, schizophrenia, or schizoaffective disorders, as the digital patterns of these conditions are different from those of depression, and people who were working regular night shifts or were pregnant, as these external factors can cause changes in sleep patterns. Researchers discussed health anxieties with potential participants through unstructured questions. On the basis of these discussions, those who believed that their health anxieties may worsen with continuous behavioral monitoring were excluded.

Ethics Approval

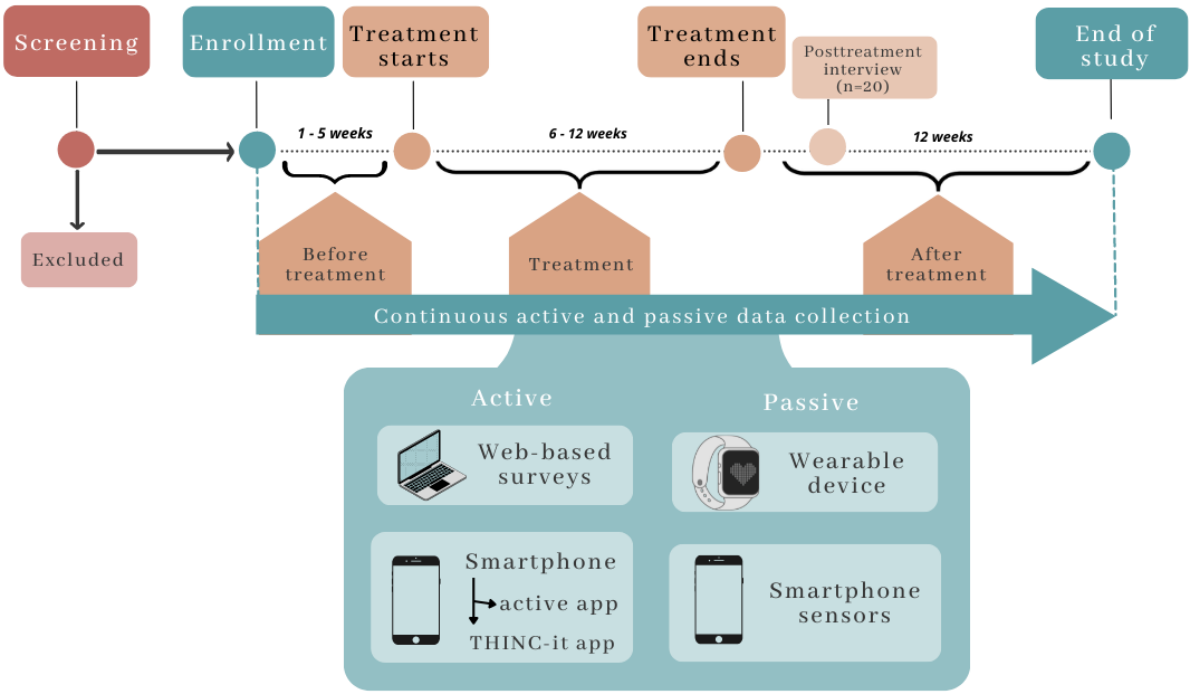
This study was reviewed and given favorable opinion by the London Westminster Research Ethics Committee and received approval from the Health Research Authority (reference number 20/LO/0091).

Procedures

Overview

Details of the measures, technology, and procedures have been covered in depth in a previous publication [21]. The methods described in this section refer to the primary aims and outcomes of the original study protocol. Therefore, the measures presented herein are relevant to this analysis. Overall, the participants were enrolled in the study at least a week before their first therapy session. The researchers had no control over the treatment provided. Consequently, the enrolled participants had different waiting list times, treatment lengths, and treatment intensities. They were followed up throughout treatment and up to 3 months after treatment using smartphone apps and a wrist-worn device (Fitbit Charge 3 or 4 [Fitbit Inc]). Therefore, active and passive data were collected for approximately 7 months, but this depended on the treatment length, which varied from person to person. The study procedures are depicted by Figure 1.

Figure 1. The study timeline for participants, from screening to the end of the study.



Baseline Session

After providing signed informed consent, the participants provided in-depth baseline sociodemographic and clinical data related to their current and previous physical and mental health conditions, family history, treatment status, phone use, and social and physical activity levels. Clinical measures included the Patient Health Questionnaire 9-item [24], a widely validated depression questionnaire, and the Generalized Anxiety Disorder 7-item scale questionnaire [25]. Participants were then guided through the installation and setup of the 4 apps used in this study, which have been detailed in the next section.

Data Collection

Overview

Active and passive data collection began from the baseline session, and data were acquired from a variety of sources. The underlying infrastructure for data collection and storage was the Remote Assessment of Disease and Relapse (RADAR)–base platform, developed by the RADAR-Central Nervous System Consortium [26].

Passive measures were gathered from (1) Fitbit wearable device sensors and (2) smartphone sensors, and active measures were gathered from (3) web-based surveys and (4) smartphone apps.

Passive Measures

Passive measures were collected from 2 devices. First, the participants were provided with a Fitbit and downloaded the Fitbit app, which provided a user interface where they could track their own activity. The data extracted through the Fitbit Application Programming Interface for use in this study were related to sleep, physical activity, heart rate, and step count. The participants used their own Android smartphone and were asked to download the RADAR-base passive RMT app, a purpose-built app that collects smartphone sensor data. Only data streams with a fixed sampling rate allow for the calculation of missing data, as it provides the total number of expected data points in a period, which serves as the denominator for the total number of observed data points. These data streams were acceleration, nearby Bluetooth device detection, and GPS. GPS coordinates were obfuscated by adding a participant-specific random number as a reference point, and the relative change in location was calculated from there; therefore, an individual's home address or precise geographic location could not be gathered.

Active Measures

Overall, active data were collected through 2 methods (Figure 1): web-based surveys and smartphone-based tasks sent via apps. Web-based surveys were clinical measures delivered by email via the REDCap (Research Electronic Data Capture; Vanderbilt University) software, a web-based platform for research that is conducted through a browser [27]. The smartphone-based active data were collected through (1) clinical questionnaires, (2) a series of speech tasks delivered directly to the participant's phone via a custom-built app (the RADAR active RMT app), and (3) validated cognitive assessments in gamified format requiring a separate app, the THINC-it app [28]

Cognitive tasks were completed monthly, whereas the speech task, which required the participants to record themselves reading a short text [29] and answering a question aloud, was delivered fortnightly. All active measures were rotated weekly such that the tasks took an average of 10 minutes per week to complete, except for 1 week in a month, when the THINC-it task increased the completion time by approximately 15 minutes. The participants were notified when it was time to carry out the tasks. Details of all active measures can be found elsewhere [21].

Posttreatment Interview

The first 20 participants who completed the therapy and agreed to participate in an optional posttreatment interview were included in the qualitative analysis. This was a 30-minute semistructured interview conducted on the web examining the participants' experiences of using RMTs during psychotherapy for depression. To reduce potential social desirability bias, interviews were conducted by researchers who had little to no previous contact with their interviewee.

Statistical Analysis

Quantitative Data

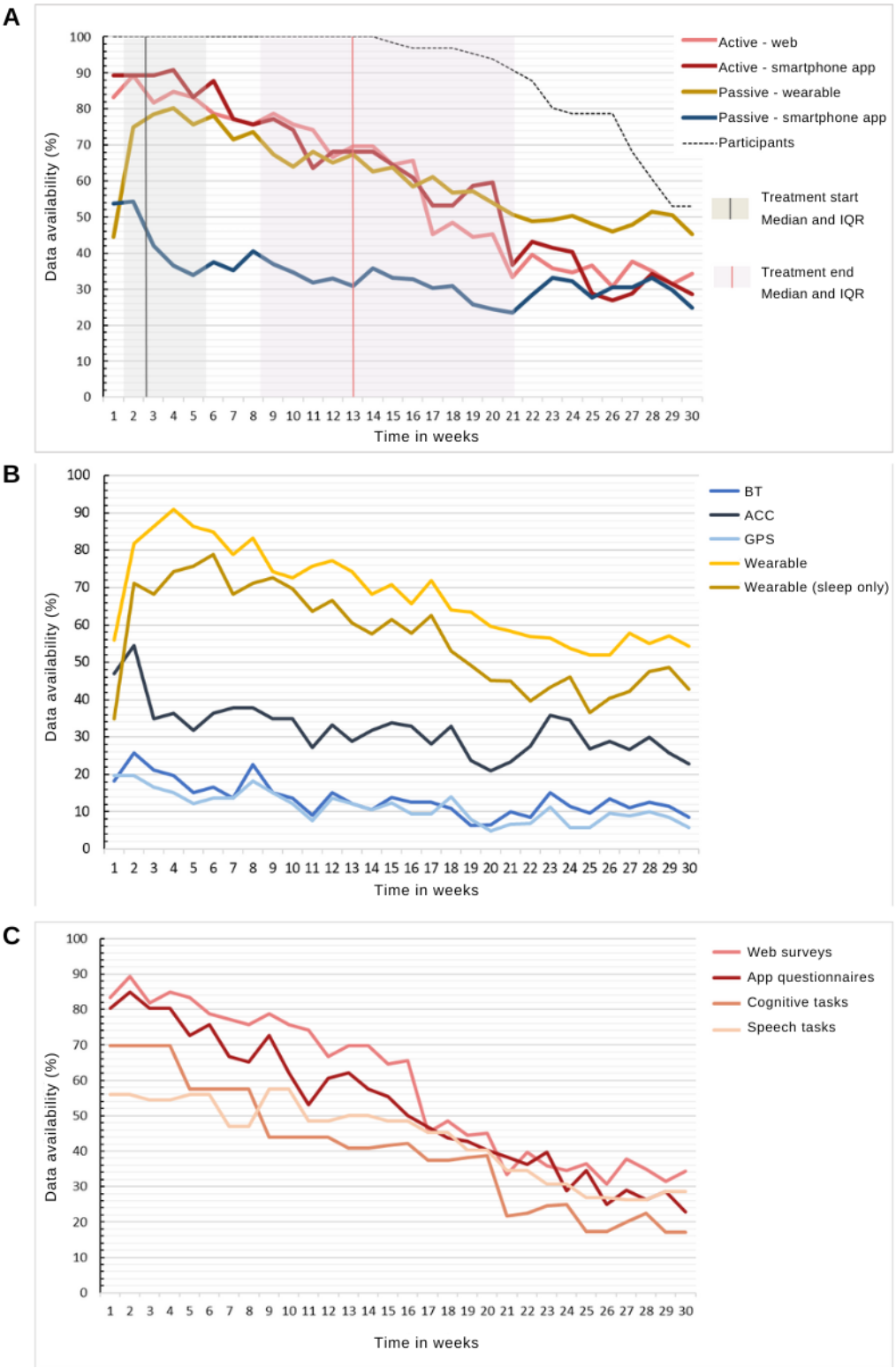
Quantitative data were collected regarding the following parameters of engagement:

1. Study engagement: the main outcome of attrition is defined as the division of study participants into those who completed the study ("completers") and those who did not because of withdrawal or loss to follow-up ("non-completers"). To determine whether symptom severity at baseline was associated with attrition, 2-tailed *t* tests were performed to compare the mean severity of clinical measures taken at baseline, namely the Patient Health Questionnaire 9-item and Generalized Anxiety Disorder 7-item, across the study completion groups. The Shapiro-Wilk test was used in all cases to test for normality distributions in variables, and if this assumption was violated, nonparametric tests were used. All other assumptions for 2-tailed *t* test calculations were met.
   To test the effect of treatment characteristics on attrition, completers were compared with noncompleters in terms of treatment length and treatment intensity. The Mann-Whitney *U* test was used for the continuous variable *treatment length* given the violation of parametric assumptions, and the chi-square test was used to compare frequencies across low- and high-intensity therapy. To examine the role of overall time in the study as a confounder (given its association with treatment length and treatment intensity), we tested its potential association with attrition by conducting a 2-tailed *t* test on the mean study length across the completion groups.
2. Engagement with RMTs: engagement with RMTs was measured as the total number of data points available out of the total number of data points expected. In terms of active data, this was calculated as the number of active tasks completed out of the total number of tasks delivered. In terms of passive data, this was calculated as the number of hours in which there was at least one data point divided by the total number of hours in a day. This was then averaged to a weekly statistic.
   Logistic regression analyses were performed to assess whether being in treatment influenced the magnitude of data availability. The 3 treatment conditions were before treatment, treatment, or after treatment. Given the expected reduction in data availability over time owing to study fatigue, we adjusted these analyses for time in weeks, age, and gender. The weeks selected for analysis had to have at least 10 participants in each treatment condition; therefore, before treatment versus treatment status comparisons involved weeks 3 to 8, and after treatment versus treatment status comparisons comprised weeks 8 to 24.

Missing Data Thresholds

We established missing data thresholds as follows. Passive data required at least one data point per hour for at least 8 hours per day to be considered available. The total number of available hours per week were calculated, and weeks with at least 50% of available hours were deemed available. Active data were sampled weekly; therefore, active data availability for each participant was defined as the completion of at least one active task that week. The proportion of participants in the study with available data each week has been presented by the dotted line in Figure 2A.

Figure 2. (A) Data availability by type of data. Y-axis 1 shows the percentage of people who contributed data out of the total number of available participants. Participant numbers are plotted against the secondary y-axis. (B) Data availability by passive data stream. This shows the proportion of participants with available data, averaged per week. Data were deemed available if there was at least one data point available per hour on at least 8 hours a day. (C) Data availability by active data. This shows the proportion of participants with at least one active data task completed, averaged per week. IQR: Interquartile range, ACC: accelerometer; BT: Bluetooth.



Qualitative Data

Transcriptions of the recordings of the semistructured interviews were checked for accuracy by a second researcher and analyzed using a deductive approach to thematic analysis, with the iterative categorization technique [30]. The deductive approach was used in favor of an inductive approach, as certain core themes in this field have been previously reported [8,31]. These were used as initial frameworks from which to organize the initial coding, as we anticipated that these concepts would also emerge from the current data, but flexibility was given to reorganize these codes as they applied to the current data.

Overarching themes, such as device engagement and the impact of treatment, were preestablished according to the quantitative objectives of the study.

All quantitative data processing and analyses were performed using R (version 4.0.2, R Core Team), and qualitative data were analyzed using NVivo (released in March 2020, QSR International).

Results

Study Engagement: Recruitment and Attrition

Over 900 people were contacted, and of these, 66 (7.3%) were finally enrolled (Figure 3). Of the 66 enrolled individuals, 40 (61%) completed the study. Sample characteristics are presented in Table 1 and show that our sample was similar in demographic proportions to the total IAPT population in South London in terms of age, gender, ethnicity structures [32], and employment status [33].

Table 2 shows the means, medians, and proportions for those who completed the study versus those who did not on treatment-related variables. A chi-square test of independence was performed to examine the relationship between treatment intensity and attrition. The relationship between these variables was significant: N=66, χ^2_1=4.6; *P*=.03. The participants who received low-intensity treatment were more likely to complete the study than those who received high-intensity treatment.

No significant differences were found between the attrition groups across the sample characteristics of age, gender, ethnicity, educational level, employment status, and previous experience with digital health tools. A Mann-Whitney *U* test was conducted to determine whether there was a difference in treatment length between the attrition groups. The results indicated a trend toward significance in terms of the difference in treatment length between the groups (*W*=339.5; *P*=.05). A significance threshold of *P*<.05 would not regard this observation as evidence for a significant difference in treatment length between completers and noncompleters, where longer treatments would be associated with attrition. These associations cannot be accounted for by symptom severity or overall time in the study, given that study length was not associated with attrition, and the severity of depression or anxiety was associated with neither treatment length nor treatment intensity. *t* tests (2-tailed) revealed that the severity of anxiety (t_{56}=−2.80; *P*=.007), but not depression (t_{50}=−0.18; *P*=.86), was associated with attrition such that higher anxiety at baseline was associated with higher attrition levels (Table 2). These associations are mapped in Figure S1 in Multimedia Appendix 1.

Figure 3. Recruitment flowchart.

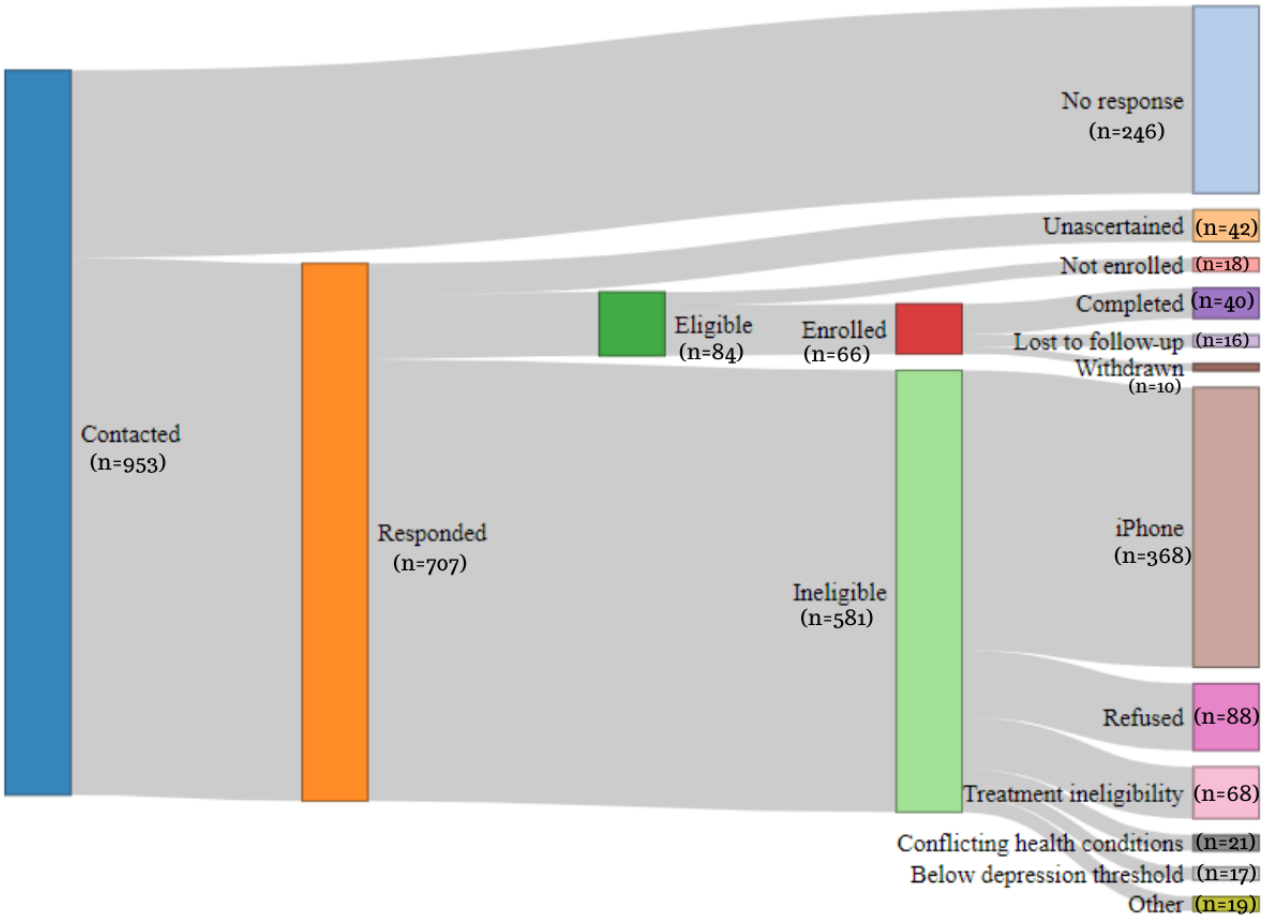


Table 1. Study sample characteristics (N=66).

	Values
Age (years), mean (SD)	34.6 (11.1)
Gender, n (%)	
Woman	40 (61)
Man	23 (35)
Nonbinary	3 (5)
Ethnicity, n (%)	
Asian or Asian British	2 (3)
Black, African, Caribbean, or Black British	11 (17)
Middle Eastern	1 (2)
Mixed or multiple ethnic groups	6 (9)
White British	37 (56)
White (other)	9 (14)
Education level, n (%)	
Secondary education	26 (39)
Degree-level education or diploma (eg, BSc[a] and BA[b])	26 (39)
Postgraduate degree (eg, MSc[c], MA[d], and PhD[e])	14 (21)
Employment status, n (%)	
Paid employment	42 (64)
Unpaid employment	4 (6)
Unemployment	12 (18)
Furlough	3 (5)
Student	4 (6)
Retired	1 (2)
Previous experience with digital health tools, n (%)	55 (83)
Psychiatric comorbidities, n (%)	
0 (single psychiatric diagnosis)	17 (26)
1	12 (18)
≥2	37 (56)
Therapy intensity, n (%)	
Low intensity	33 (50)
High intensity	32 (48)
Treatment data (time in weeks), mean (SD)	
Treatment start lag	4.7 (4.7)
Treatment length[f]	11.6 (6.5)
Posttreatment follow-up	14.5 (6.4)
Total study	29.6 (6.6)
PHQ-9[g]	16.7 (5.1)
GAD-7[h]	13.3 (4.7)

[a]BSc: Bachelor of Science.

[b]BA: Bachelor of Arts.

[c]MSc: Master of Science.

[d]MA: Master of Arts.

[e]PhD: Doctor of Philosophy.

[f]Treatment length is the number of weeks between the first and last sessions, and not the total number of sessions.

[g]PHQ-9: Patient Health Questionnaire 9-item.

[h]GAD-7: Generalized Anxiety Disorder 7-item.

Table 2. Summary statistics and statistical analyses of associations among completion groups, explanatory variables, and covariates.

	Attrition		Analysis			
	Completers (n=40)	Noncompleters (n=26)	Test	*df*	Test statistic	*P* value
Treatment intensity (%)						
Low intensity	75.76	24.24	Chi-square test	1	4.6	.03
High intensity	46.88	53.13	—[a]	—	—	—
Treatment length (weeks), median (IQR)						
Length	7 (5-12.25)	12 (6-22.25)	Mann-Whitney *U* test	N/A[b]	339.5	.05
Covariates, mean (SD)						
Anxiety severity[c]	12.33 (4.66)	15.4 (4.06)	*t* test	56.28	-2.80	.007
Depression severity[d]	16.65 (5.08)	16.88 (5.17)	*t* test	50.40	-0.18	.86
Study length (weeks)	29.38 (6.02)	29.92 (7.56)	*t* test	45.02	-0.31	.76

[a]Not available.

[b]N/A: not applicable.

[c]Anxiety was measured with the Generalized Anxiety Disorder 7-item scale questionnaire.

[d]Depression was measured using the Patient Health Questionnaire 9-item.

Engagement With RMTs: Data Availability

The four main types of data collected were (1) wearable passive data, (2) smartphone-based passive data, (3) smartphone-based active data, and (4) web-based active data. Figure 2 shows how these data types vary in terms of their availability in the study, where 100% data completion would mean that the participants supplied, on average, 100% of the data that week. The data availability for smartphone-based passive data was between 20% and 40% for the duration of the study. Fitbit-based passive data and both active data streams had a similar proportion of data availability, but the rate of decline was lower for wearable passive data than for active data. To describe the missing data patterns across the passive data streams, the proportion of participants who provided sensor data for at least 8 hours is plotted in Figure 2B. There was no established threshold for the minimum acceptable quantity of passive data necessary to perform an analysis of these data. Therefore, we established a limit of 8 hours per day as an acceptable threshold for missing data, as the main daily activities, such as work and sleep, can be broken down into 8-hour cycles. GPS location was the passive data stream most vulnerable to missing data, followed by Bluetooth (Figure 2B). The active data stream most vulnerable to missing data for the first 10 weeks was speech and the THINC-it cognitive task thereafter (Figure 2C).

To study whether being in treatment affected data availability, the amount of data contributed by the participants who were actively receiving treatment was compared with that contributed by those who were in either pretreatment (on the waiting list) or posttreatment. Logistic regression, adjusted for time in weeks, revealed minor effects of treatment status on data availability. Significant differences in data availability were found for active smartphone and passive wearable data. When comparing those in treatment with those in pretreatment (Table 3), we found that the odds of those in treatment having active app smartphone were 2.54 times that of those on the waiting list having active app data, regardless of the time in the study. Conversely, there was a 54% decrease in the odds of contributing Fitbit data for those in treatment compared with those in pretreatment. In summary, more clinical questionnaires were completed while in treatment than while on the waiting list, but more Fitbit data were available while on the waiting list than during treatment.

Table 3. Odds ratios (ORs) for data availability in treatment as compared with that in pretreatment for overlapping weeks, adjusting for time (in weeks), age, and gender. A positive absolute difference in ORs show that those in treatment had more data availability.

Data streams	Treatment versus pretreatment (weeks 3-8)[a]		
	Absolute difference (unadjusted), %	OR (95% CI)	*P* value
Active			
Web based	−4.40	0.89 (0.44-1.79)	.76
Smartphone	+[b]9.24	2.34 (1.16-4.78)	.01
Passive			
Wearable	−9.94	0.46 (0.21-0.95)	.04
Smartphone	−6.31	0.72 (0.44-1.16)	.17

[a]Weeks 3 to 8 when n for all groups was <10.

[b]The absolute difference in available data (as a %) between treatment and pretreatment conditions. That is, there was 9.24% more smartphone data during treatment versus pretreatment.

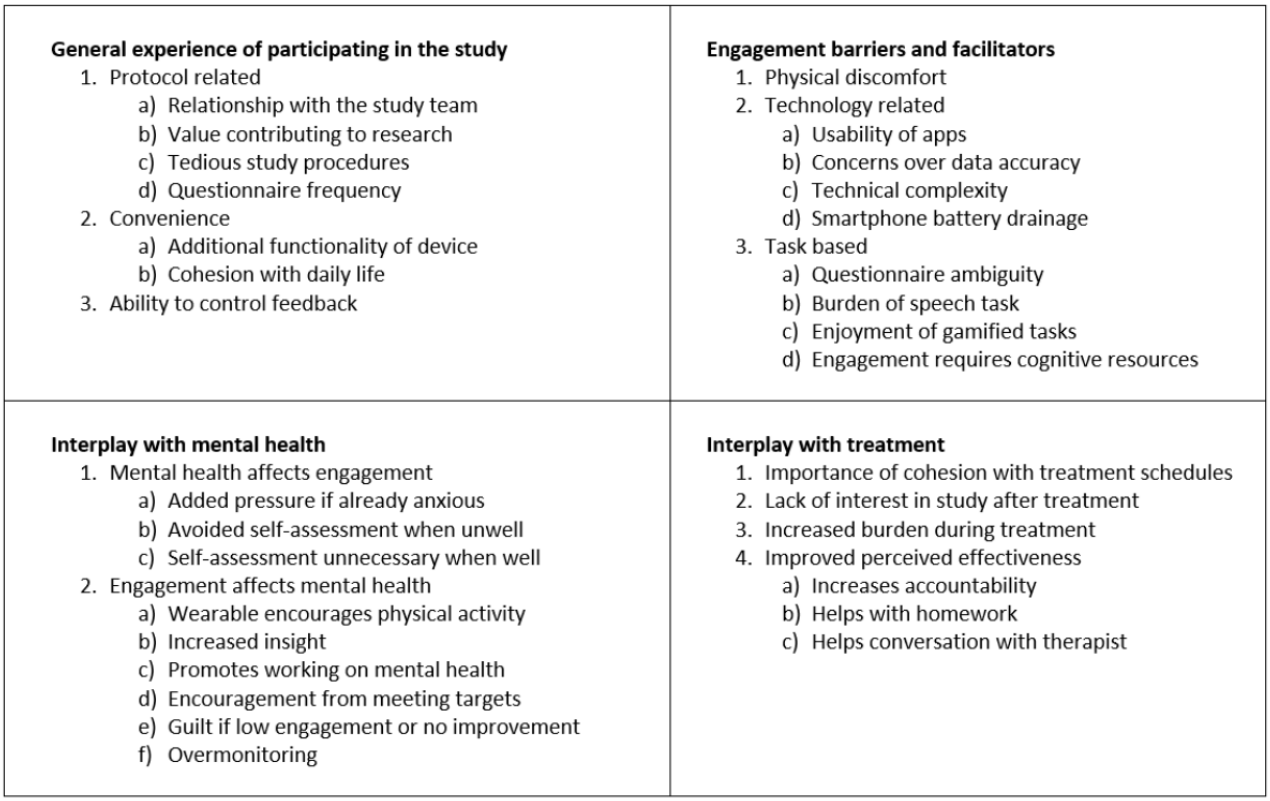
Qualitative Evaluation

A total of 4 major themes related to the study aims were developed from the 20 semistructured interviews, as shown in Figure 4. Quotes associated with each subtheme can be found in Table S2 in Multimedia Appendix 1, and quantified participant responses are described in Figure S2.

The first major theme was the general participation experience. Protocol-related subthemes revolve around the idea that having a good relationship with the study team improves their general experience. A strong motivator of engagement was knowing that they were contributing to research, but experience was dampened by tedious study procedures in the form of repetitive and high-frequency questionnaires. Given the differences in preferences over when to receive feedback, how to receive feedback, and how much feedback to receive regarding the participants' measured mood and behavior, it was thought that the flexibility to control these would have improved the experience.

The level of engagement with the apps and devices was affected by physical discomfort of wearing a Fitbit; technology-related issues, which relate to any technical challenges, such as battery issues and measurement accuracy; and the tasks themselves, specifically their complexity and enjoyability, which added burden or ease to their engagement.

Figure 4. Four major themes were developed from the interviews; each subpanel shows the numbered minor themes and subthemes.



The interplay between mental health and engagement was found to be bidirectional. For example, when participants were unwell, some reported avoiding the self-reflection required by the questionnaires, whereas others experienced this only when feeling well. On the one hand, some reported an improvement through the encouragement of health-promoting behaviors, whereas others experienced guilt or anxiety from obsessing over data, especially if no improvement was apparent.

As for the interplay between treatment and device engagement, the participants felt that the study had to integrate well with their treatment schedules. Treatment milestones were found to affect engagement with the study, with some people losing interest after therapy and others finding it harder to remain engaged with the increased burden of treatment. RMTs were seen to promote treatment effectiveness by increasing accountability with the therapist, providing targets to complete homework, and helping therapeutic conversations. The following section draws upon these results to aid the interpretation of engagement patterns.

Discussion

Principal Findings

We evaluated the feasibility of using both active and passive data collection methods in the psychological treatment for depression. We examined recruitment, engagement with the study protocol through attrition, and engagement with the technology through patterns of missing data. We then used qualitative interviews to help interpret the feasibility data and gain insight into such data patterns.

Study Engagement: Recruitment and Attrition

Recruitment rates were low in proportion to the number of people initially contacted, which is in line with previous depression studies that found recruitment challenging, with large variations in success rates [34]. Digital health studies present potential participants with additional concerns, including unfamiliarity with technology and privacy misgivings [31], which may have contributed to low uptake]. Remote sensing studies tend to show a higher recruitment uptake than this study, for example, the study by Matcham et al [15], a discrepancy that could be largely explained by the exclusion of iPhone (Apple Inc) users and requirement to time recruitment with the start of treatment. Importantly, the similarity of our sample to the target population in key demographic aspects provides some reassurance that uptake was equitable across the main sample characteristics.

A retention of 60% after 7 months is markedly lower than the 94% retention after a 9-month follow-up in a sample at different stages of recurrent major depressive disorder [15], yet it is higher than the 50% retention after 15 days in a self-referral study with little researcher contact [13] and the 53% retention in studies on psychotherapy treatments [35]. Some of the key differences between these studies that can help explain the differences in engagement rates stem from the context in which the studies were carried out, participant burden, and questionnaire frequency, the latter 2 being subthemes emerging from our interviews.

We also found that treatment length and intensity affected attrition, as did higher baseline anxiety, with longer and more intense treatment being associated with early disengagement. In line with the finding from the participant interviews that participation burden increases during treatment, our quantitative analysis found that treatment characteristics and symptom severity influence retention. This could be partly because of the competing cognitive resources between engagement with RMT and treatment tasks, as described in the interviews. For example, other studies have found that constant feedback from health devices may worsen health anxieties [8], which may disproportionately impact those with higher anxiety compared with those with depression.

The main implications for engagement with the study protocol relate to scalability, generalizability, and digital divide. On the one hand, slow recruitment may reflect a low readiness among patients to sign up for remote monitoring within health care services, with implications for increased scale-up costs and staff training. Higher attrition in people with clinical and treatment-related complexities may result in them deriving fewer benefits from the implementation of RMTs than their counterparts. In addition, studies using these samples may have a higher risk of attrition bias than those on less complex treatments. This limitation to the generalizability of research findings owing to bias would, therefore, disproportionally affect those with more complex needs, widening the digital divide.

Engagement With RMTs: Data Availability

The availability of data from the wearables and active questionnaires showed a similar pattern of decline over time. By contrast, smartphone-based passive data showed a low but stable data pattern. Some data streams, such as GPS and Bluetooth for passive data and speech and cognitive tests for active data, are more likely to be incomplete.

Some forms of data collection place a greater burden on the user than others. Therefore, it was expected that active data collection forms that have the highest participant burden would have a faster pace of decline as a function of time and cause study fatigue [36]. Therefore, it was unsurprising that the active data streams that contributed the fewest data points were speech and cognitive tasks, which were lengthier and, according to our interviews, more cognitively demanding. Although wearables require very little engagement, they still involve some level of action: they must be worn, charged, and synchronized. According to the participant interviews, 50% of those interviewed chose not to wear the Fitbit because of comfort and privacy issues, among other reasons. By contrast, passive apps are unobtrusive in their data collection, as they do not require a regular smartphone user to deviate from their usual behavior, and, therefore, produced a more stable pattern of data availability, which seemed to be less affected by study burden.

The passive data streams most vulnerable to missing data were the GPS and Bluetooth sensors. However, other passive sensing studies on mental health have found the opposite pattern, with more data being available from GPS than from accelerometers [15,20,37]. Sensor noncollection can occur for multiple reasons, including participants turning off the data permissions or the sensor itself. GPS and Bluetooth are sensors that can be easily

switched off from a smartphone's main setting page and may be seen as more intrusive forms of monitoring. This was supported by the finding that 35% of those interviewed felt "monitored" by the apps and the recurrence of "privacy" as a theme in RMT research for health care [38].

When comparing the participants in treatment with those on the waiting list, there was an increased completion of active tasks during therapy. In the current sample, the participants expressed the benefits of having a cohesive experience with RMTs and treatment such that completing active tasks during treatment helped with homework, promoted working on their mental health, and sparked conversations with their therapist. The literature on self-management in digital health shows that, despite the potential for added burden, there is a disposition for symptom tracking during treatment [39]. Conversely, people in treatment had less Fitbit wear time than those on the waiting list. The increased self-awareness that comes from tracking health with the Fitbit can be demotivating if there are no evident improvements in health outcomes such as sleep and physical activity [8], which might increase the likelihood of participants removing the device to avoid feelings of guilt and internal pressure [40].

The implications for engagement with RMTs relate to the integrity of the data collected and the differing acceptability thresholds for different devices. Low data availability means that the features derived from passive data may lack accuracy and could lead to false interpretations; however, even from the same device, different sensors contribute different amounts of data. Given that multiple sensor combinations are used to infer different aspects of behavior, accurate feature construction may require longer data collection windows for certain sensors depending on their target behavior. Although data imputation methods may help address some of these issues, increasing data availability using engagement strategies is likely to yield more accurate results. Several suggestions have been proposed by Currey and Torous [20], including overcoming the tendency of smartphones to halt data collection when apps are idle by including active components in passive apps.

Additionally, if different devices (eg, smartphones vs wearables) produce different levels of data availability, this may have implications for the appropriateness of their use, depending on the purpose and length of data collection. Smartphone-based data collection for long-term monitoring is only appropriate if the resources are available to support increased data availability strategies; otherwise, the amount of data may be too scarce to be informative. If such strategies are in place, smartphone-based data, despite having a lesser overall amount, may be a more suitable option than wearables for long-term monitoring since wearables initially provide more data, but that amount gradually decreases over time. This study demonstrated wearables to be a feasible method for collecting activity and sleep data, 2 core items in psychotherapy for depression, for at least 32 consecutive weeks, before data availability falls below the 40% mark. Therefore, this method may be superior to smartphone-based data in a naturalistic context involving relatively longer-term treatments. However, it is important to consider strategies to increase user engagement with technology that take a patient-centered approach, including selecting measures that are meaningful to patients [41].

Limitations and Future Directions

This was a longitudinal cohort study, so the comparison groups of treatment versus nontreatment differed in more ways than only the exposure to treatment and different treatment intensities. The participants in the treatment group were compared with those in the same week of the study but who had yet to start treatment; delayed treatment start was related to treatment intensity, clinical risk, catchment area for the health care center, and symptom severity. Despite our efforts to account for these variables in the analysis, there is a possibility of residual confounders. Future studies could quantify the components of treatment that are related to poorer engagement.

Engagement with RMTs is broadly defined as data availability, which assumes that the occurrence of missing data is because of the participants deliberately disengaging. However, despite presenting some evidence of personal and clinical characteristics related to data availability, missing data can also be completely missing at random because of software errors. These factors may affect data streams differently based on technical factors. Future research could determine the nature of missingness by mapping technical issues to missing data.

There is no standard method to establish a threshold for "missing data." In this study, we justify a minimum of 8 hours of passive data and at least one active task completed; however, other studies (eg, the study by Matcham et al [15]) considered data availability as a single point of data per hour. It is critical to understand how much missing data are admissible before the integrity of the data is affected so that there can be an accurate characterization of the behavior. A single point of active data may describe a symptom experience for the previous 2 weeks, whereas a single passive data point covers a second's worth of activity. Therefore, rates of missingness need to be interpreted with this relativity in mind, and future studies should work toward establishing acceptable thresholds for data availability for each behavioral feature under study.

This study has shown engagement differences between data collection types and a difference in engagement between those in treatment and those on the waiting list; however, despite the statistical significance, future work should attempt to establish whether these differences are clinically meaningful.

Finally, the COVID-19 pandemic has given rise to a rapid adoption of technology, especially in the health care sector [42]. This is likely to have had an impact on people's attitudes toward digital tools for health monitoring and, consequently, engagement with RMTs. As a result, this study may have picked up on higher technology acceptance, or conversely, technology fatigue, as a factor of time.

Conclusions

We investigated the feasibility of remote collection methods in the psychological treatment for depression and reported the extent to which it was feasible to collect active and passive data via RMTs in a population with depression within a health care setting. Uptake was low but equal across the main demographic

categories and, therefore, broadly representative of the target population. Retention in our study was low, but comparable with retention rates in psychological therapy [35]. Treatment characteristics such as length and intensity were associated with attrition, as was higher baseline anxiety, suggesting that patients undergoing more complex treatment may perceive fewer benefits from long-term remote monitoring. In addition, different data streams showed different levels of missing data despite being gathered from the same device, implying that different sensors may require different data collection protocols to ensure sufficient data for accurate feature construction. Being in treatment also affected RMT engagement in different ways, depending on the device, with Fitbit contributing less data during treatment but active tasks being completed more often. Future work should establish acceptable thresholds for data availability for different sensors and devices to ensure a minimum requirement for the integrity of RMT data and investigate which aspects of treatment are related to poorer engagement. Finally, successful implementation of RMTs requires more than the user engagement measures presented in this study; however, adopting these user engagement measures is a key first step.

Acknowledgments

This research was reviewed by a team with experience of mental health problems and their caregivers, trained to advise on research through the Feasibility and Acceptability Support Team for Researchers: a free, confidential service in England provided by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre.

This study represents independent research funded by the National Institute for Health and Care Research Biomedical Research Centre in South London and Maudsley National Health Service (NHS) Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, National Institute for Health and Care Research, or Department of Health and Social Care. The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript.

The authors would like to acknowledge the staff and students of King's College London and the South London and Maudsley Improving Access to Psychological Therapies services in Lewisham, Croydon, and Lambeth who contributed to and supported this project. The authors convey special thanks to Katherine Tallent, Areej Elgaziari, Rebecca Chapman, and Alice Pace at Improving Access to Psychological Therapies for their help in recruiting and coordinating the link between their services and our project and to Nor Jamil, Lucy Hyam, and Alice Nynabb at King's College London for their hard work on recruitment and the overall running of the study.

RD was supported by the following: (1) NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, United Kingdom; (2) Health Data Research United Kingdom, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation, and Wellcome Trust; (3) The BigData@Heart consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement number 116074 (this joint undertaking receives support from the European Union's Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations and is chaired by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and European Society of Cardiology) (4) the NIHR University College London Hospitals Biomedical Research Centre; (5) the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London; (6) the UK Research and Innovation London Medical Imaging and Artificial Intelligence Centre for Value Based Healthcare; and (7) the NIHR Applied Research Collaboration South London (NIHR Applied Research Collaboration South London) at King's College Hospital NHS Foundation Trust.

Conflicts of Interest

MH is the principal investigator of the Remote Assessment of Disease and Relapse–Central Nervous System program, a precompetitive public-private partnership funded by the Innovative Medicines Initiative and the European Federation of Pharmaceutical Industries and Associations.

Multimedia Appendix 1
Supplementary information to the paper, including information on treatments and quantitative and qualitative analyses.
[DOCX File , 137 KB - mental_v10i1e42866_app1.docx]

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]

2. Leone M, Kuja-Halkola R, Leval A, D'Onofrio BM, Larsson H, Lichtenstein P, et al. Association of youth depression with subsequent somatic diseases and premature death. JAMA Psychiatry 2021 Mar 01;78(3):302-310 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.3786] [Medline: 33295952]
3. Gelenberg AJ, Freeman MP, Markowitz JC, Rosenbaum JF, Thase ME, Trivedi MH, et al. Practice Guideline for the Treatment of Patients with Major Depressive Disorder. 3rd edition. Washington, DC, USA: American Psychiatric Association; 2010.
4. Depression in adults: treatment and management - NICE Guideline. National Institute for Health and Care Excellence. 2022 Jun 29. URL: https://www.nice.org.uk/guidance/ng222/chapter/Recommendations#treatment-for-a-new-episode-of-less-severe-depression [accessed 2022-07-13]
5. Clark DM. Realizing the mass public benefit of evidence-based psychological therapies: the IAPT program. Annu Rev Clin Psychol 2018 May 07;14:159-183 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050817-084833] [Medline: 29350997]
6. More than half of patients who finished psychological therapy recovered in 2017-18. National Health Service Digital. 2018 Nov 20. URL: https://digital.nhs.uk/news/2018/more-than-half-of-patients-who-finished-psychological-therapy-recovered-in-2017-18 [accessed 2022-02-23]
7. Passos IC, Ballester P, Rabelo-da-Ponte F, Kapczinski F. Precision psychiatry: the future is now. Can J Psychiatry 2022 Jan;67(1):21-25 [FREE Full text] [doi: 10.1177/0706743721998044] [Medline: 33757313]
8. de Angel V, Lewis S, White KM, Matcham F, Hotopf M. Clinical targets and attitudes toward implementing digital health tools for remote measurement in treatment for depression: focus groups with patients and clinicians. JMIR Ment Health 2022 Aug 15;9(8):e38934 [FREE Full text] [doi: 10.2196/38934] [Medline: 35969448]
9. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
10. White KM, Williamson C, Bergou N, Oetzmann C, de Angel V, Matcham F, et al. A systematic review of engagement reporting in remote measurement studies for health symptom tracking. NPJ Digit Med 2022 Jun 29;5(1):82 [FREE Full text] [doi: 10.1038/s41746-022-00624-7] [Medline: 35768544]
11. de Girolamo G, Barattieri di San Pietro C, Bulgari V, Dagani J, Ferrari C, Hotopf M, et al. The acceptability of real‐time health monitoring among community participants with depression: a systematic review and meta‐analysis of the literature. Depress Anxiety 2020 Apr 27;37(9):885-897 [FREE Full text] [doi: 10.1002/da.23023]
12. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022 Jan 11;5(1):3 [FREE Full text] [doi: 10.1038/s41746-021-00548-8] [Medline: 35017634]
13. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
14. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Mar 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
15. Matcham F, Leightley D, Siddi S, Lamers F, White KM, Annas P, RADAR-CNS consortium. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022 Mar 21;22(1):136 [FREE Full text] [doi: 10.1186/s12888-022-03753-1] [Medline: 35189842]
16. Goldberg SB, Buck B, Raphaely S, Fortney JC. Measuring psychiatric symptoms remotely: a systematic review of remote measurement-based care. Curr Psychiatry Rep 2018 Aug 28;20(10):81. [doi: 10.1007/s11920-018-0958-z] [Medline: 30155749]
17. Gual-Montolio P, Martínez-Borba V, Bretón-López JM, Osma J, Suso-Ribera C. How are information and communication technologies supporting routine outcome monitoring and measurement-based care in psychotherapy? A systematic review. Int J Environ Res Public Health 2020 May 02;17(9):3170 [FREE Full text] [doi: 10.3390/ijerph17093170] [Medline: 32370140]
18. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
19. Matcham F, Carr E, White KM, Leightley D, Lamers F, Siddi S, RADAR-CNS consortium. Predictors of engagement with remote sensing technologies for symptom measurement in Major Depressive Disorder. J Affect Disord 2022 Aug 01;310:106-115. [doi: 10.1016/j.jad.2022.05.005] [Medline: 35525507]
20. Currey D, Torous J. Increasing the value of digital phenotyping through reducing missingness: a retrospective analysis. medRxiv 2022 May 17 [FREE Full text] [doi: 10.1101/2022.05.17.22275182]
21. de Angel V, Lewis S, Munir S, Matcham F, Dobson R, Hotopf M. Using digital health tools for the Remote Assessment of Treatment Prognosis in Depression (RAPID): a study protocol for a feasibility study. BMJ Open 2022 May 06;12(5):e059258 [FREE Full text] [doi: 10.1136/bmjopen-2021-059258] [Medline: 35523486]
22. Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. J Clin Epidemiol 2012 Mar;65(3):301-308. [doi: 10.1016/j.jclinepi.2011.07.011] [Medline: 22169081]

23. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-57. [Medline: 9881538]
24. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
25. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
26. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, RADAR-CNS Consortium. RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 2019 Aug 01;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]
27. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
28. McIntyre RS, Best MW, Bowie CR, Carmona NE, Cha DS, Lee Y, et al. The THINC-integrated tool (THINC-it) screening assessment for cognitive dysfunction: validation in patients with major depressive disorder. J Clin Psychiatry 2017 Jul;78(7):873-881. [doi: 10.4088/JCP.16m11329] [Medline: 28858441]
29. International Phonetic Association. Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet. Cambridge, UK: Cambridge University Press; 1999.
30. Neale J. Iterative categorization (IC): a systematic technique for analysing qualitative data. Addiction 2016 Jun;111(6):1096-1106 [FREE Full text] [doi: 10.1111/add.13314] [Medline: 26806155]
31. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, RADAR-CNS Consortium. Barriers to and facilitators of engagement with mHealth technology for remote measurement and management of depression: qualitative analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]
32. Harwood H, Rhead R, Chui Z, Bakolis I, Connor L, Gazard B, et al. Variations by ethnicity in referral and treatment pathways for IAPT service users in South London. Psychol Med (forthcoming) 2021 Aug 02:1-12. [doi: 10.1017/S0033291721002518] [Medline: 34334151]
33. Psychological Therapies, Annual Reports on the use of IAPT services. National Health Service Digital. 2021. URL: https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services [accessed 2022-07-08]
34. Liu Y, Pencheon E, Hunter RM, Moncrieff J, Freemantle N. Recruitment and retention strategies in mental health trials - a systematic review. PLoS One 2018 Aug 29;13(8):e0203127 [FREE Full text] [doi: 10.1371/journal.pone.0203127] [Medline: 30157250]
35. Wierzbicki M, Pekarik G. A meta-analysis of psychotherapy dropout. Prof Psychol Res Pr 1993 May;24(2):190-195 [FREE Full text] [doi: 10.1037/0735-7028.24.2.190]
36. Dineley J, Lavelle G, Leightley D, Matcham F, Siddi S, Peñarrubia-María MT, et al. Remote smartphone-based speech collection: acceptance and barriers in individuals with major depressive disorder. In: Proceedings of the 2021 Interspeech. 2021 Presented at: Interspeech '21; August 30-September 3, 2021; Brno, Czechia p. 631-635 URL: https://www.isca-speech.org/archive/interspeech_2021/dineley21_interspeech.html [doi: 10.21437/interspeech.2021-1240]
37. Kiang MV, Chen JT, Krieger N, Buckee CO, Alexander MJ, Baker JT, et al. Sociodemographic characteristics of missing data in digital phenotyping. Sci Rep 2021 Jul 29;11(1):15408 [FREE Full text] [doi: 10.1038/s41598-021-94516-7] [Medline: 34326370]
38. Zhang H, Ibrahim A, Parsia B, Poliakoff E, Harper S. Passive social sensing with smartphones: a systematic review. Computing (forthcoming) 2022 Aug 12 [FREE Full text] [doi: 10.1007/s00607-022-01112-2]
39. Niendam TA, Tully LM, Iosif AM, Kumar D, Nye KE, Denton JC, et al. Enhancing early psychosis treatment using smartphone technology: a longitudinal feasibility and validity study. J Psychiatr Res 2018 Jan;96:239-246. [doi: 10.1016/j.jpsychires.2017.10.017] [Medline: 29126059]
40. Kerner C, Goodyear V. The motivational impact of wearable healthy lifestyle technologies: a self-determination perspective on Fitbits with adolescents. Am J Health Educ 2017 Jul 28;48(5):287-297 [FREE Full text] [doi: 10.1080/19325037.2017.1343161]
41. Manta C, Patrick-Lake B, Goldsack J. Digital measures that matter to patients: a framework to guide the selection and development of digital measures of health. Digit Biomark 2020 Sep 15;4(3):69-77 [FREE Full text] [doi: 10.1159/000509725] [Medline: 33083687]
42. Hutchings R. The impact of Covid-19 on the use of digital technology in the NHS. Nuffield Trust. 2020 Aug. URL: https://www.nuffieldtrust.org.uk/files/2020-08/the-impact-of-covid-19-on-the-use-of-digital-technology-in-the-nhs-web-2.pdf [accessed 2022-09-09]

Abbreviations

IAPT: Improving Access to Psychological Therapies
RADAR: Remote Assessment of Disease and Relapse
REDCap: Research Electronic Data Capture
RMT: remote measurement technology

Edited by J Torous; submitted 21.09.22; peer-reviewed by M Kapsetaki, S Difrancesco; comments to author 05.11.22; revised version received 10.11.22; accepted 26.11.22; published 24.01.23.

Please cite as:
de Angel V, Adeleye F, Zhang Y, Cummins N, Munir S, Lewis S, Laporta Puyal E, Matcham F, Sun S, Folarin AA, Ranjan Y, Conde P, Rashid Z, Dobson R, Hotopf M
The Feasibility of Implementing Remote Measurement Technologies in Psychological Treatment for Depression: Mixed Methods Study on Engagement
JMIR Ment Health 2023;10:e42866
URL: https://mental.jmir.org/2023/1/e42866
doi:10.2196/42866
PMID:36692937

Original Paper

The Effectiveness of a Smartphone Intervention Targeting Suicidal Ideation in Young Adults: Randomized Controlled Trial Examining the Influence of Loneliness

Lauren McGillivray[1*], PhD; Nicholas Keng-Meng Hui[2*], BMed; Quincy J J Wong[3], PhD; Jin Han[1], PhD; Jiahui Qian[1], PhD; Michelle Torok[1], PhD

[1]Black Dog Institute, University of New South Wales, Sydney, Australia

[2]Faculty of Medicine and Health, University of New South Wales, Sydney, Australia

[3]School of Psychology, Western Sydney University, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Michelle Torok, PhD
Black Dog Institute
University of New South Wales
Hospital Rd
Sydney, 2031
Australia
Phone: 61 2 9065 9133
Email: m.torok@unsw.edu.au

Abstract

Background: Loneliness is commonly reported by young people and has been shown to contribute to the rapid onset and escalation of depression and suicidal ideation during adolescence. Lonely people may also be particularly susceptible to disengaging from treatment early given the likelihood of their more complex clinical profiles leading to cognitive fatigue. While a smartphone intervention (LifeBuoy) has been shown to effectively reduce suicidal ideation in young adults, poor engagement is a well-documented issue for this therapeutic modality and has been shown to result in poorer treatment outcomes.

Objective: This study aims to determine whether loneliness affects how young people experiencing suicidal ideation engage with and benefit from a therapeutic smartphone intervention (LifeBuoy).

Methods: A total of 455 community-based Australian young adults (aged 18-25 years) experiencing recent suicidal ideation were randomized to use a dialectical behavioral therapy–based smartphone intervention (LifeBuoy) or an attention-matched control app (LifeBuoy-C) for 6 weeks. Participants completed measures of suicidal ideation, depression, anxiety, and loneliness at baseline (T0), post intervention (T1), and 3 months post intervention (T2). Piecewise linear mixed models were used to examine whether loneliness levels moderated the effect of LifeBuoy and LifeBuoy-C on suicidal ideation and depression across time (T0 to T1; T1 to T2). This statistical method was then used to examine whether app engagement (number of modules completed) influenced the relationship between baseline loneliness and suicidal ideation and depression across time.

Results: Loneliness was positively associated with higher levels of overall suicidal ideation (*B*=0.75, 95% CI 0.08-1.42; *P*=.03) and depression (*B*=0.88, 95% CI 0.45-1.32; *P*<.001), regardless of time point or allocated condition. However, loneliness did not affect suicidal ideation scores across time (time 1: *B*=1.10, 95% CI –0.25 to 2.46; *P*=.11; time 2: *B*=0.43, 95% CI –1.25 to 2.12; *P*=.61) and depression scores across time (time 1: *B*=0.00, 95% CI –0.67 to 0.66; *P*=.99; time 2: *B*=0.41, 95% CI –0.37 to 1.18; *P*=.30) in either condition. Similarly, engagement with the LifeBuoy app was not found to moderate the impact of loneliness on suicidal ideation (*B*=0.00, 95% CI –0.17 to 0.18; *P*=.98) or depression (*B*=–0.08, 95% CI –0.19 to 0.03; *P*=.14).

Conclusions: Loneliness was not found to affect young adults' engagement with a smartphone intervention (LifeBuoy) nor any clinical benefits derived from the intervention. LifeBuoy, in its current form, can effectively engage and treat individuals regardless of how lonely they may be.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619001671156; https://tinyurl.com/yvpvn5n8

International Registered Report Identifier (IRRID): RR2-10.2196/23655

(JMIR Ment Health 2023;10:e44862) doi:10.2196/44862

KEYWORDS

loneliness; suicidal ideation; suicide prevention; digital therapeutics; smartphone intervention; apps

Introduction

Suicidal ideation is a serious condition affecting approximately 11% of all young people [1,2] and has been shown to increase the risk of suicide attempt and suicide death approximately 2-fold [2-4]. Understanding which young people have an elevated risk of ideation is critical to improving early detection and shaping early intervention initiatives that can effectively prevent premature mortality in young people. There are many reasons why young people may experience suicidal ideation, including mood and affective disorders [5], trauma [6], and relationship issues [7], among other factors. One condition that has not received much attention in understanding suicidal trajectories is loneliness, which is a common manifestation of and highly comorbid with complex mental health and trauma experiences [5,6,8,9].

Loneliness has been defined as the unpleasant subjective experience of feeling isolated when a discrepancy exists between the actual and the desired levels of interpersonal relationships [10]. Research has found that loneliness is more prevalent in adolescents and young adults than in working-aged adults [11] and that it can contribute to the rapid onset and escalation of depression [8,9] and suicidal ideation [11,12] during adolescence. Findings from The Young Australian Loneliness Survey [13] found that one-quarter of young people surveyed (ages 12-25 years) reported clinically significant levels of loneliness, and young adults in particular (ages 18-25 years) were lonelier than adolescents (ages 12-17 years). These findings suggest that loneliness may be developmentally important to the onset and escalation of suicidal ideation in young people. Accordingly, further confirmatory research to establish the impacts of loneliness on young people's mental health is warranted.

Recent research examining the social and health impacts of the COVID-19 pandemic suggests there is a growing epidemic of loneliness among young people [14]. Many public health measures implemented to contain the pandemic enforced physical isolation from friends and family, and may have led to disproportionate negative mental health effects for young people [15-17], as social connections with peers play a vital role in their healthy development [18]. For example, there is evidence of worsening rates of depression [19] and marked increases in hospital presentations for suicidal ideation and self-harm [20] among young people since the onset of the pandemic. There are, however, major gaps in our understanding of how loneliness may be contributing to these trends, particularly how it may have impacted help-seeking behaviors for mental health issues [21]. Though limited, there is some evidence that more severe loneliness is associated with greater external or professional help-seeking, driven by a diminished perception of being able to self-manage mental health symptoms [21,22]. Related to these knowledge gaps on links between loneliness and help-seeking, there is a dearth of research examining how loneliness may affect how young people interact with, and benefit from, help they do access given potential comorbidities with complex psychopathology [10,11]. Specifically, there is no research yet on this issue in relation to digital interventions for mental health, which increased in availability during the pandemic when face-to-face services were limited [23].

Meta-analytic research on the effectiveness of self-guided digital health interventions for suicide prevention shows that these digital tools can help to reduce the severity of suicidal ideation [24-27], and there is also evidence that the level of engagement with digital health interventions is a significant positive moderator of clinical benefit [28,29]. However, issues of poor engagement with digital health interventions are well documented, with an average of 70% of users not completing all treatment modules [28], yet reasons for poor engagement are not fully understood. The few research studies that have investigated predictors of engagement have focused on the intervention itself (eg, desirable content or interface [30-32]) rather than on user characteristics. Attrition data from clinical trials have noted higher dropout rates among individuals with more clinically severe profiles (eg, severe depression) due to the psychological effort required to attain treatment goals [29,33]. Given that loneliness is a significant and independent risk factor for complex psychopathology, including depression and suicidal ideation [8,9,11,12,34-36], lonely people may be particularly susceptible to disengaging from treatment early. Establishing what role loneliness plays in how young people engage with and subsequently benefit from digital health interventions will help to advance current understandings of whether specific strategies are needed to support adherence, which is particularly relevant to ensuring supports are being accessed optimally during periods of heightened loneliness and mental ill health, such as health epidemics or pandemics.

This study presents a secondary and post hoc exploratory analysis of data from a randomized controlled trial that tested the efficacy of a therapeutic smartphone intervention (LifeBuoy) against an attention-matched placebo control app in supporting young adults to self-manage suicidal ideation. The trial took place during the COVID-19 pandemic. The main analyses showed that LifeBuoy was associated with significant reductions in the severity of suicidal ideation in the intervention condition compared to the control group [25]; however, a better understanding of the fundamental person-specific characteristics that may influence outcomes will help inform who self-guided digital health interventions should be targeted to:

- Aim 1: Determine whether baseline loneliness severity moderates change in suicidal ideation and depression scores across time among young adults in both the intervention and control conditions
- Aim 2: Examine whether engagement with LifeBuoy (defined as the number of modules completed) moderates

the impact of loneliness on suicidal ideation and depression in the treatment (exposure) condition

Methods

This paper is a secondary analysis of a clinical trial of the LifeBuoy smartphone intervention, with the full trial details, safety protocols [37], and main outcomes reported elsewhere [25]. The trial protocol was prospectively registered on the Australian New Zealand Clinical Trials Registry (ACTRN12619001671156) and has also been published elsewhere [37]; the LifeBuoy intervention was registered with the Therapeutic Goods Administration Clinical Trial Notification scheme (CT-2020-CTN-00256-1-v1).

Ethics Approval

Ethics approval for this study was obtained from the University of New South Wales Human Research Ethics Committee (HC190764). All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. As such, informed consent to participate was obtained from all participants prior to data collection.

Participants and Recruitment

The study was a 2-arm parallel, double-blind, randomized controlled trial conducted at the Black Dog Institute (BDI), Australia. Participants were recruited from the community via targeted advertisements posted on Facebook between May 11 and May 22, 2020 (Multimedia Appendix 1). Eligibility criteria were as follows: aged between 18 and 25 years, fluent in English, own and have access to a smartphone, currently residing in Australia, and having experienced suicidal thoughts in the past 12 months. Participants were excluded if they reported being diagnosed with a psychotic disorder ever or had attempted suicide in the 1 month prior to trial registration due to safety considerations. Participants undergoing other mental health treatments were not excluded from the trial. The study was approved by the University of New South Wales Human Research Ethics Committee (HC190764) and required written opt-in consent from participants.

Procedure

After completing the baseline survey, participants were assigned randomly (2:2) to the intervention or control condition within a block design (four per block). Randomization was performed using a computer-generated algorithm integrated into BDI's digital trial management platform. All study participants and investigators were masked from condition assignment until the completion of the final survey.

The trial was run online using this trial management platform, which also supported the recruitment portal, including online consent, screening, registration, data collection, and access to the apps (see [25] for full details). Participants were required to complete these closed (password-protected) surveys within 7 days at baseline (T0, day 0), post intervention (T1, day 43), and 3 months post intervention (T2, day 132). After completing the baseline assessment, participants were sent a link (via email and SMS text message) to download the free app randomly allocated to them and informed that they had 6 weeks to access this app. No reminders/prompts or support were provided to use the apps.

Intervention

LifeBuoy is a brief self-guided smartphone intervention consisting of seven models grounded in dialectical behavior therapy and acceptance and commitment therapy. Each module takes approximately 5 minutes to complete, and participants had the opportunity to return to them as frequently as needed over the course of 6 weeks. Participants were required to complete each module sequentially before moving on to the next. Each module introduced and practiced knowledge and skills relating to distress tolerance, emotional regulation, and goal setting through interactive learning exercises. A detailed description of LifeBuoy has been published elsewhere along with screenshots of the app [25].

Control

A matched-attention placebo app (LifeBuoy-C) was developed for the control condition (details published elsewhere) [25]. LifeBuoy-C provided generalized mental health information in a similar format and time expectancy to the intervention app to control for potential digital placebo effects [38]. LifeBuoy-C mirrored the module structure, user interface, and graphic design of LifeBuoy but contained no therapeutic content.

Measures

Suicidal Ideation (Primary Outcome)

Suicidal ideation was measured using the Suicidal Ideation Attributes Scale (SIDAS) [39]. The SIDAS comprises five items rated on an 11-point scale that assess the frequency, severity, impact, and controllability of suicidal ideation. Total scores range from 0 to 50, with higher scores signifying more severe suicidal ideation. Scores ≥21 indicate a high risk for suicidal behavior. The scale has demonstrated acceptable internal consistency (Cronbach α=.91) [39].

Loneliness

Loneliness was measured using the Three-Item Loneliness Scale (TILS) [40], which assesses social exclusion, lack of companionship, and social isolation. The TILS is based on the University of California, Los Angeles Loneliness Scale [40] and has demonstrated good internal consistency (Cronbach α=.72) [40]. Items are rated on a 3-point Likert scale from hardly ever (1) to often (3), with a higher combined score (ranging from 3 to 9) indicating greater loneliness.

Depression

Depression was measured using the Patient Health Questionnaire (PHQ-9). The PHQ-9 assesses the nine diagnostic criteria of depression from the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition) [41]. Respondents rate their responses on a 4-point Likert scale that ranges from not at all (0) to nearly every day (3). Total scores range from 0 to 27, with higher scores indicating greater depression severity. The scale has demonstrated good internal consistency (α>.80) [42].

App Engagement

The LifeBuoy app recorded details about the number of modules completed and automatically uploaded this data to central servers. Module completion was recorded at the time of baseline (T0) and at the postintervention (T1) and 3-month postintervention follow-up (T2).

Statistical Analysis

To determine whether the loneliness levels of participants at baseline moderated the effect of the LifeBuoy and LifeBuoy-C apps on suicidal ideation (SIDAS) and depressive symptoms (PHQ-9) across time (T0 to T1; T1 to T2), piecewise linear mixed models for repeated measures analyses were used. Thus, loneliness (TILS), two time variables reflecting T0 to T1 and T1 to T2, condition (LifeBuoy, LifeBuoy-C), and relevant interaction terms between these variables were specified as fixed effects in models, along with an intercept as a random effect. An identity covariance matrix was used, and degrees of freedom were estimated using Satterthwaite's method. The linear mixed model approach can include all participants in the analyses, even those with missing follow-up data points, under the missing-at-random assumption, and therefore accords with the intention-to-treat principle.

To examine within the LifeBuoy condition whether app engagement affected the relationship between loneliness and each of the key outcomes (suicidal ideation, depression), piecewise linear mixed models for repeated measures analyses were conducted similar to the above approach described. In this case, app engagement (the number of modules completed), loneliness (TILS), two time variables reflecting T0 to T1 and T1 to T2, and relevant interaction terms between these variables were specified as fixed effects in models, along with an intercept as a random effect. Broader engagement data (number of modules completed by condition and median time spent in the app by condition) has been reported elsewhere [25]. Safety data and harms have also been reported elsewhere [24].

For all analyses, P values ≤.05 were considered significant. For all linear mixed models examined, residual plots indicated no distinct patterns, and the distribution of residuals in histograms approximated normality, indicating model assumptions were met. SPSS 25.0 (IBM Corp) was used for all analyses. Attrition analyses were conducted and reported elsewhere [25].

Results

User Statistics

In total, 455 young adults completed the baseline survey and were randomized to either the intervention condition (n=228) or the control condition (n=227). No participants formally withdrew or were excluded from the study. Demographic and clinical profiles of the total sample are described elsewhere, along with a CONSORT (Consolidated Standards for Reporting Trials) trial flow diagram [25]. The mean loneliness score at baseline was 7.17 (SD 1.83), with no significant differences between the intervention and control conditions (7.19 vs 7.16; F=0.89; t_{453}=0.16; P=.44). The mean SIDAS score was 22.61 (SD 8.18), which indicates clinically relevant levels of suicidal ideation (scores ≥21), with 136 (59.6%) participants in the intervention condition and 129 (57.1%) participants in the control condition meeting this criterion. The mean PHQ-9 score was 17.15 (SD 5.64), indicating that the sample, on average, met the clinical threshold for moderate-severe depression (scores ≥15). A total of 152 (66.7%) participants in the intervention condition and 156 (69.3%) participants in the control condition met the moderate-severe depression threshold. There were 202 (90.2%) participants in the intervention condition and 199 (88.4%) participants in the control condition who endorsed ever receiving mental health treatment. In the intervention condition, there was no significant difference in the number of modules completed between those who have (mean 6.80, SD 4.41) and have not (mean 6.77, SD 2.84) ever received mental health treatment ($F_{1,222}$<1; P=.98). In the control condition, there was no significant difference in the number of modules completed between those who have (mean 6.84, SD 4.96) and have not (mean 5.81, SD 3.82) ever received mental health treatment ($F_{1,223}$=1.04; P=.31).

In the intervention condition, there was a significant positive correlation between baseline loneliness and PHQ-9 scores (r=0.32; P<.001) but not between loneliness and SIDAS scores (r=–0.03; P=.67). In the control condition, significant positive correlations were observed between baseline loneliness and SIDAS scores (r=0.16; P=.02) and loneliness and PHQ-9 scores (r=0.27; P<.001).

Loneliness at T0 as a Moderator of the Effect of LifeBuoy and LifeBuoy-C on Suicidal Thinking and Depression

For the model with suicidal thinking (SIDAS) as the dependent variable (Table 1), there was a significant main effect of loneliness (P=.03), indicating that higher levels of loneliness were positively associated with higher levels of suicidal thinking overall, irrespective of condition or time point. There was also a significant time 1 × condition interaction (P=.02), as expected given the primary outcomes analyses previously reported [25] reflecting a nonsignificant change in suicidal thinking in the LifeBuoy-C condition (B=–3.23, 95% CI –10.48 to 4.01; z=–0.87; P=.38; d=–0.38) but a significant decrease in suicidal thinking in the LifeBuoy condition (B=–7.84, 95% CI –14.76 to –0.92; z=–2.22; P=.03; d=–1.00), resulting in a between condition effect size (d=0.42) at T1. There were no other significant effects. For the model with depression (PHQ-9) as the dependent variable (Table 1), there was a significant main effect of loneliness (P<.001), indicating that higher levels of loneliness were positively associated with higher levels of depression overall, irrespective of condition or time point. There were no other significant effects.

Table 1. Loneliness as a moderator of the effect of LifeBuoy and LifeBuoy-C on the SIDAS and PHQ-9.

	DV[a]: SIDAS[b]		DV: PHQ-9[c]	
	B (95% CI)	*P* value	*B* (95% CI)	*P* value
Intercept	16.98 (12.03 to 21.94)	<.001	10.86 (7.66 to 14.06)	<.001
Time 1[d]	–0.75 (–8.01 to 6.51)	.84	–2.01 (–5.62 to 1.59)	.27
Time 2[e]	1.15 (–8.01 to 10.32)	.80	2.19 (–1.90 to 6.27)	.29
Condition	6.71 (–0.05 to 13.48)	.05	–0.53 (–4.88 to 3.82)	.81
Loneliness	0.75 (0.08 to 1.42)	*.03*[f]	0.88 (0.45 to 1.32)	*<.001*
Time 1 × condition	–12.52 (–22.56 to –2.49)	*.02*	–0.92 (–5.87 to 4.02)	.71
Time 2 × condition	–2.35 (–14.96 to 10.26)	.71	–1.57 (–7.30 to 4.15)	.59
Time 1 × loneliness	–0.35 (–1.33 to 0.63)	.49	–0.14 (–0.62 to 0.34)	.57
Time 2 × loneliness	–0.36 (–1.59 to 0.87)	.57	–0.32 (–0.87 to 0.23)	.25
Condition × loneliness	–0.87 (–1.78 to 0.05)	.06	0.06 (–0.53 to 0.64)	.85
Time 1 × condition × loneliness	1.10 (–0.25 to 2.46)	.11	0.00 (–0.67 to 0.66)	.99
Time 2 × condition × loneliness	0.43 (–1.25 to 2.12)	.61	0.41 (–0.37 to 1.18)	.30

[a]DV: dependent variable.

[b]SIDAS: Suicidal Ideation Attributes Scale.

[c]PHQ-9: Patient Health Questionnaire.

[d]Time 1: postintervention.

[e]Time 2: 3-month follow-up.

[f]Italicized *P* values denote statistical significance.

App Engagement as a Moderator of Loneliness' Relationship With Suicidal Thinking and Depression in the LifeBuoy Condition

Among the LifeBuoy condition (n=228), the mean number of app modules completed was 6.84 (SD 4.27), and loneliness scores at baseline were not significantly correlated with module completion (r=0.05; P=.42). In examining whether loneliness moderated the relationship between the number of LifeBuoy modules completed and clinical benefits, there were no significant effects in the model with suicidal thinking (SIDAS) as the dependent variable (Table 2). For the model with depression (PHQ-9) as the dependent variable (Table 2), there was only a significant main effect of loneliness (P<.001), indicating that higher levels of loneliness were positively associated with higher levels of depression overall, irrespective of time point or engagement.

Table 2. App engagement as a moderator of loneliness' relationship with the SIDAS and PHQ-9 in the LifeBuoy condition.

	DV[a]: SIDAS[b]		DV: PHQ-9[c]	
	B (95% CI)	*P* value	*B* (95% CI)	*P* value
Intercept	23.50 (13.48 to 33.52)	<.001	6.44 (0.34 to 12.54)	.04
Time 1[d]	–13.81 (–30.82 to 3.20)	.11	0.56 (–6.97 to 8.08)	.88
Time 2[e]	–10.00 (–32.45 to 12.45)	.38	1.76 (–7.18 to 10.70)	.70
Loneliness	–0.13 (–1.46 to 1.19)	.84	1.47 (0.66 to 2.27)	<.001
Modules	0.03 (–1.32 to 1.39)	.96	0.60 (–0.23 to 1.42)	.16
Time 1 x loneliness	0.83 (–1.37 to 3.02)	.46	–0.53 (–1.51 to 0.45)	.29
Time 2 x loneliness	1.02 (–1.88 to 3.92)	.49	–0.14 (–1.32 to 1.03)	.81
Time 1 x modules	0.08 (–2.17 to 2.33)	.94	–0.54 (–1.56 to 0.48)	.30
Time 2 x modules	1.22 (–1.70 to 4.14)	.41	–0.17 (–1.38 to 1.03)	.78
Loneliness x modules	0.00 (–0.17 to 0.18)	.99	–0.08 (–0.19 to 0.03)	.14
Time 1 x loneliness x modules	–0.01 (–0.30 to 0.28)	.94	0.06 (–0.07 to 0.19)	.38
Time 2 x loneliness x modules	–0.13 (–0.51 to 0.25)	.50	0.03 (–0.12 to 0.19)	.68

[a]DV: dependent variable.

[b]SIDAS: Suicidal Ideation Attributes Scale.

[c]PHQ-9: Patient Health Questionnaire.

[d]Time 1: postintervention.

[e]Time 2: 3-month follow-up.

Discussion

Principal Findings

This study investigated whether loneliness played a role in moderating the efficacy of and engagement with a therapeutic smartphone intervention designed to help young people manage suicidal ideation. Baseline loneliness scores were not found to significantly moderate changes in suicidal ideation or depression severity post intervention or at the 3-month follow-up in either condition. The severity or extent of loneliness at baseline also had no effect on how young adults self-engaged with the LifeBuoy app, contrary to prior research that suggests that lonelier individuals may lack self-efficacy to manage mental health issues [21,22].

These results are consistent with the existing literature examining the relationship between loneliness and mental health to the extent that they suggest that loneliness is associated with more complex psychopathology [8,9,11,12,34-36]. In this study, this relationship was evidenced by significant positive correlations between loneliness and suicidal ideation and depression scores. However, these findings go beyond confirming simple relationships between loneliness and mental health. This study shows that despite loneliness being associated with poorer mental health, it does not appear to be a characteristic or condition that moderates how young people engage with, nor benefit from, a digital health intervention. Though prior studies have posited that positive associations between depression and treatment dropout may be moderated by the psychological effort required to sustain treatment adherence [31,43], this study suggests that loneliness is not an independent nor significant factor that exacerbates this nonadherence, at least for brief smartphone interventions.

One potential explanation for the null association is that there was a high prevalence of moderate-severe levels of suicidal ideation and depression across the total sample, which may have obscured or dampened the effects of transdiagnostic conditions such as loneliness. Alternatively, the proportion of participants reporting clinically severe suicidal ideation (SIDAS score ≥21) may be important to consider in efforts to understand why loneliness did not emerge as a unique predictor of engagement nor benefit in this study. The LifeBuoy app was purposefully designed to improve ideation [25], and the trial was marketed to young people for whom ideation was currently or recently a salient concern. As such, LifeBuoy may have been seen as an intervention highly relevant to the needs of the young people who participated in the trial—and need is a strong intrinsic motivator for engagement [44]. Indeed, young people in both conditions engaged similarly well with the smartphone apps over the course of the trial [25]. Actual high levels of motivation or need to engage with the app may have overridden any cognitive barriers typically associated with loneliness. Examining the role of loneliness in a more clinically heterogenous or diverse sample of young adults may help to clarify the relationships between mental ill health, loneliness, and engagement with treatment. As this is the first study to examine the role of loneliness in how young people experiencing suicidal ideation benefit from a digital health intervention, caution must be taken against concluding that loneliness is not an important factor in understanding engagement with, and benefit from, self-guided treatment.

Overall, our findings show that participants engaged with and benefited from LifeBuoy, irrespective of loneliness severity, which shows that LifeBuoy, in its current form, can effectively engage and treat individuals despite how lonely they may be. This finding strengthens the potential for digital health interventions to become an established mental health treatment method for all young people experiencing suicidal ideation. These findings also suggest that a digital therapeutic such as LifeBuoy might be suitable for large-scale delivery during future pandemics or environmental crises when loneliness is expected to increase.

The study findings also add new knowledge about which person-specific characteristics or conditions may be important to target using digital health interventions to enhance adherence and clinical benefits. Loneliness, as an isolated factor, may not be important to uniquely target future digital health intervention development when samples or users present with complex psychopathology.

Although this study found no significant link between baseline loneliness and depression, previous research has shown that greater depression severity can hinder engagement with and benefit from digital health interventions [31,43]. Future studies could explore whether additional social factors such as social connectedness and social anxiety act to moderate the relationship between loneliness, suicidal ideation, and depression.

Strengths and Limitations

The study has several strengths. To our knowledge, this study is the first to examine whether loneliness—a known risk factor for suicidal ideation and depression—affects how young people engage with a digital health intervention. Accordingly, this study contributes new knowledge about the person-specific characteristics that may need to be considered when designing and delivering such interventions. The data for this study was derived from a well-designed randomized controlled trial of a therapeutic smartphone intervention involving 455 young adults and thus provided a large robust sample that supports the reliability of the current findings and conclusions.

There are some limitations to consider. As patient information was collected independently via a web-based survey, self-reporting bias could affect the findings via over- or underestimated effects. The TILS was used because it is a brief validated measure that can reduce survey length and minimize attrition; however, the limited number of items and the small score range (3-9) may not appropriately capture nuance or heterogeneity in loneliness. Future studies focusing primarily on loneliness should consider using more comprehensive scales. Given that loneliness was not a primary or major secondary outcome of the main trial examining the efficacy of LifeBuoy, measures that differentiate between social and emotional loneliness, such as the DeJong-Gierveld Loneliness Scale [45], were not considered and would have allowed for more nuanced analyses of the role of loneliness in digital health interventions. Information on current engagement with mental health treatment was not collected, although whether participants ever received mental health treatment was assessed, and app engagement did not differ on this variable in both intervention and control conditions. Future studies similar to this study should assess both current and past mental health treatment of participants and determine the potential impact of these variables on app engagement. Broader limitations relating to the primary randomized controlled trial have been discussed elsewhere [25].

Conclusion

Loneliness did not affect how young people engaged with or benefited from a digital health intervention (LifeBuoy) shown to reduce suicidal ideation in young adults. Despite a significant positive association between loneliness and depression levels, loneliness did not affect young people's engagement with LifeBuoy nor moderate clinical outcomes. As the first study to explore how loneliness may affect young people's engagement with and benefits from a targeted digital health intervention, this study has developed new insights into whom these self-guided interventions may work best for. Future studies should analyze the interplay of additional factors such as motivation, social connectedness, and social anxiety, as well as distinguish between social and emotional loneliness using more refined measures.

Acknowledgments

We would like to thank the young adults with a lived experience of suicide who participated in this study and the Black Dog Institute IT and development teams who developed LifeBuoy and helped set up the trial.

The parent trial of the Lifebuoy smartphone app was funded by the Australian Rotary Health Mental Health of Young Australians Research Grant Scheme (RG192840) and a National Health and Medical Research Council Early Career Fellowship, which was awarded to MT (RG1138710). The development of the LifeBuoy smartphone app was funded by a philanthropic grant awarded by the S & C Roth Family Foundation (RG20210528).

Data Availability

The data sets generated or analyzed during this study are not publicly available due to the sensitive nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Recruitment advertisement.
[DOCX File , 53 KB - mental_v10i1e44862_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 7256 KB - mental_v10i1e44862_app2.pdf]

References

1. Borges G, Benjet C, Orozco R, Medina-Mora M. The growth of suicide ideation, plan and attempt among young adults in the Mexico City metropolitan area. Epidemiol Psychiatr Sci 2017 Dec;26(6):635-643 [FREE Full text] [doi: 10.1017/S2045796016000603] [Medline: 27523098]
2. Voss C, Ollmann TM, Miché M, Venz J, Hoyer J, Pieper L, et al. Prevalence, onset, and course of suicidal behavior among adolescents and young adults in Germany. JAMA Netw Open 2019 Oct 02;2(10):e1914386 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.14386] [Medline: 31664450]
3. Nock MK, Green JG, Hwang I, McLaughlin KA, Sampson NA, Zaslavsky AM, et al. Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement. JAMA Psychiatry 2013 Mar;70(3):300-310 [FREE Full text] [doi: 10.1001/2013.jamapsychiatry.55] [Medline: 23303463]
4. Ribeiro JD, Franklin JC, Fox KR, Bentley KH, Kleiman EM, Chang BP, et al. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. Psychol Med 2016 Jan;46(2):225-236 [FREE Full text] [doi: 10.1017/S0033291715001804] [Medline: 26370729]
5. Gili M, Castellví P, Vives M, de la Torre-Luque A, Almenara J, Blasco MJ, et al. Mental disorders as risk factors for suicidal behavior in young people: a meta-analysis and systematic review of longitudinal studies. J Affect Disord 2019 Feb 15;245:152-162. [doi: 10.1016/j.jad.2018.10.115] [Medline: 30390504]
6. Barzilay R, Moore TM, Calkins ME, Maliackel L, Jones JD, Boyd RC, et al. Deconstructing the role of the exposome in youth suicidal ideation: trauma, neighborhood environment, developmental and gender effects. Neurobiol Stress 2021 May;14:100314 [FREE Full text] [doi: 10.1016/j.ynstr.2021.100314] [Medline: 33869680]
7. Zygo M, Pawłowska B, Potembska E, Dreher P, Kapka-Skrzypczak L. Prevalence and selected risk factors of suicidal ideation, suicidal tendencies and suicide attempts in young people aged 13-19 years. Ann Agric Environ Med 2019 Jun 17;26(2):329-336. [doi: 10.26444/aaem/93817] [Medline: 31232067]
8. Danneel S, Nelemans S, Spithoven A, Bastin M, Bijttebier P, Colpin H, et al. Internalizing problems in adolescence: linking loneliness, social anxiety symptoms, and depressive symptoms over time. J Abnorm Child Psychol 2019 Oct;47(10):1691-1705. [doi: 10.1007/s10802-019-00539-0] [Medline: 30937813]
9. Vanhalst J, Klimstra TA, Luyckx K, Scholte RHJ, Engels RCME, Goossens L. The interplay of loneliness and depressive symptoms across adolescence: exploring the role of personality traits. J Youth Adolesc 2012 Jun;41(6):776-787. [doi: 10.1007/s10964-011-9726-7] [Medline: 22045508]
10. Peplau LA. Perspective on loneliness. In: Perlman D, editor. Loneliness: A Sourcebook of Current Theory, Research and Therapy. New York: John Wiley and Sons; 1982:1-18.
11. McClelland H, Evans JJ, Nowland R, Ferguson E, O'Connor RC. Loneliness as a predictor of suicidal ideation and behaviour: a systematic review and meta-analysis of prospective studies. J Affect Disord 2020 Sep 01;274:880-896. [doi: 10.1016/j.jad.2020.05.004] [Medline: 32664029]
12. Bennardi M, Caballero FF, Miret M, Ayuso-Mateos JL, Haro JM, Lara E, et al. Longitudinal relationships between positive affect, loneliness, and suicide ideation: age-specific factors in a general population. Suicide Life Threat Behav 2019 Feb;49(1):90-103. [doi: 10.1111/sltb.12424] [Medline: 29210215]
13. Lim MH, Eres R, Peck C. The young Australian loneliness survey: understanding loneliness in adolescence and young adulthood. VicHealth. 2019. URL: https://www.vichealth.vic.gov.au/-/media/ResourceCentre/PublicationsandResources/Social-connection/The-young-Australian-loneliness-survey-Report.pdf [accessed 2023-03-20]
14. Sampogna G, Giallonardo V, Del Vecchio V, Luciano M, Albert U, Carmassi C, et al. Loneliness in young adults during the first wave of COVID-19 lockdown: results from the multicentric COMET study. Front Psychiatry 2021;12:788139 [FREE Full text] [doi: 10.3389/fpsyt.2021.788139] [Medline: 34955932]
15. Groarke JM, Berry E, Graham-Wisener L, McKenna-Plumley PE, McGlinchey E, Armour C. Loneliness in the UK during the COVID-19 pandemic: cross-sectional results from the COVID-19 Psychological Wellbeing Study. PLoS One 2020;15(9):e0239698 [FREE Full text] [doi: 10.1371/journal.pone.0239698] [Medline: 32970764]
16. Kendrick K, Isaac M. Mental health impact of COVID-19: Australian perspective. Indian J Psychiatry 2020 Sep;62(Suppl 3):S373-S376 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_853_20] [Medline: 33227062]

17. Karageorghis CI, Bird JM, Hutchinson JC, Hamer M, Delevoye-Turrell YN, Guérin SMR, et al. Physical activity and mental well-being under COVID-19 lockdown: a cross-sectional multination study. BMC Public Health 2021 May 27;21(1):988 [FREE Full text] [doi: 10.1186/s12889-021-10931-5] [Medline: 34039306]
18. Houghton S, Hattie J, Carroll A, Wood L, Baffour B. It hurts to be lonely! Loneliness and positive mental wellbeing in Australian rural and urban adolescents. J Psychologists Counsellors Schools 2016 Feb 04;26(1):52-67. [doi: 10.1017/jgc.2016.1]
19. Wang Y, Kala MP, Jafar TH. Factors associated with psychological distress during the coronavirus disease 2019 (COVID-19) pandemic on the predominantly general population: a systematic review and meta-analysis. PLoS One 2020;15(12):e0244630 [FREE Full text] [doi: 10.1371/journal.pone.0244630] [Medline: 33370404]
20. Sara G, Wu J, Uesi J, Jong N, Perkes I, Knight K, et al. Growth in emergency department self-harm or suicidal ideation presentations in young people: comparing trends before and since the COVID-19 first wave in New South Wales, Australia. Aust N Z J Psychiatry 2023 Jan;57(1):58-68 [FREE Full text] [doi: 10.1177/00048674221082518] [Medline: 35266405]
21. Hom MA, Stanley IH, Joiner TE. Evaluating factors and interventions that influence help-seeking and mental health service utilization among suicidal individuals: a review of the literature. Clin Psychol Rev 2015 Aug;40:28-39. [doi: 10.1016/j.cpr.2015.05.006] [Medline: 26048165]
22. Kleinberg A, Aluoja A, Vasar V. Social support in depression: structural and functional factors, perceived control and help-seeking. Epidemiol Psychiatr Sci 2013 Dec;22(4):345-353 [FREE Full text] [doi: 10.1017/S2045796013000504] [Medline: 24063718]
23. Dang A, Dang D, Rane P. The expanding role of digital therapeutics in the post-COVID-19 era. Open COVID J 2021 May 21;1(1):32-37. [doi: 10.2174/2666958702101010032]
24. Hetrick SE, Yuen HP, Bailey E, Cox GR, Templer K, Rice SM, et al. Internet-based cognitive behavioural therapy for young people with suicide-related behaviour (Reframe-IT): a randomised controlled trial. Evid Based Ment Health 2017 Aug;20(3):76-82. [doi: 10.1136/eb-2017-102719] [Medline: 28701336]
25. Torok M, Han J, McGillivray L, Wong Q, Werner-Seidler A, O'Dea B, et al. The effect of a therapeutic smartphone application on suicidal ideation in young adults: findings from a randomized controlled trial in Australia. PLoS Med 2022 May;19(5):e1003978 [FREE Full text] [doi: 10.1371/journal.pmed.1003978] [Medline: 35639672]
26. Torok M, Han J, Baker S, Werner-Seidler A, Wong I, Larsen ME, et al. Suicide prevention using self-guided digital interventions: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2020 Jan;2(1):e25-e36 [FREE Full text] [doi: 10.1016/S2589-7500(19)30199-2] [Medline: 33328037]
27. Kahl BL, Miller HM, Cairns K, Giniunas H, Nicholas M. Evaluation of ReachOut.com, an unstructured digital youth mental health intervention: prospective cohort study. JMIR Ment Health 2020 Oct 15;7(10):e21280 [FREE Full text] [doi: 10.2196/21280] [Medline: 33055066]
28. Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079 [FREE Full text] [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
29. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
30. Gulliver A, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Batterham PJ. Predictors of acceptability and engagement in a self-guided online program for depression and anxiety. Internet Interv 2021 Sep;25:100400 [FREE Full text] [doi: 10.1016/j.invent.2021.100400] [Medline: 34026569]
31. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
32. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726 [FREE Full text] [doi: 10.1017/S0033291715000665] [Medline: 25881626]
33. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009 Feb 23;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]
34. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
35. Mushtaq R, Shoib S, Shah T, Mushtaq S. Relationship between loneliness, psychiatric disorders and physical health ? A review on the psychological aspects of loneliness. J Clin Diagn Res 2014 Sep;8(9):WE01-WE04 [FREE Full text] [doi: 10.7860/JCDR/2014/10077.4828] [Medline: 25386507]
36. Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. Psychol Aging 2010 Jun;25(2):453-463 [FREE Full text] [doi: 10.1037/a0017216] [Medline: 20545429]

37. Han J, McGillivray L, Wong QJ, Werner-Seidler A, Wong I, Calear A, et al. A mobile health intervention (LifeBuoy App) to help young people manage suicidal thoughts: protocol for a mixed-methods randomized controlled trial. JMIR Res Protoc 2020 Oct 27;9(10):e23655 [FREE Full text] [doi: 10.2196/23655] [Medline: 33107832]
38. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]
39. van Spijker BAJ, Batterham PJ, Calear AL, Farrer L, Christensen H, Reynolds J, et al. The suicidal ideation attributes scale (SIDAS): community-based validation study of a new scale for the measurement of suicidal ideation. Suicide Life Threat Behav 2014 Aug;44(4):408-419. [doi: 10.1111/sltb.12084] [Medline: 24612048]
40. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging 2004;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]
41. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. J Affect Disord 2021 Jan 15;279:473-483. [doi: 10.1016/j.jad.2020.09.131] [Medline: 33126078]
42. Kroenke K, Wu J, Yu Z, Bair MJ, Kean J, Stump T, et al. Patient Health Questionnaire Anxiety and Depression Scale: initial validation in three clinical trials. Psychosom Med 2016;78(6):716-727 [FREE Full text] [doi: 10.1097/PSY.0000000000000322] [Medline: 27187854]
43. Farrer LM, Griffiths KM, Christensen H, Mackinnon AJ, Batterham PJ. Predictors of adherence and outcome in internet-based cognitive behavior therapy delivered in a telephone counseling setting. Cognitive Ther Res 2013 Oct 27;38(3):358-367. [doi: 10.1007/s10608-013-9589-1]
44. Reiss S. Extrinsic and intrinsic motivation at 30: unresolved scientific issues. Behav Anal 2005;28(1):1-14 [FREE Full text] [doi: 10.1007/BF03392100] [Medline: 22478436]
45. Grygiel P, Humenny G, Rębisz S. Using the De Jong Gierveld Loneliness Scale with early adolescents: factor structure, reliability, stability, and external validity. Assessment 2019 Mar;26(2):151-165. [doi: 10.1177/1073191116682298] [Medline: 27932403]

Abbreviations

BDI: Black Dog Institute
CONSORT: Consolidated Standards of Reporting Trials
PHQ-9: Patient Health Questionnaire
SIDAS: Suicidal Ideation Attributes Scale
TILS: Three-Item Loneliness Scale

Edited by J Torous; submitted 06.12.22; peer-reviewed by R Melia, M Marciniak; comments to author 30.01.23; revised version received 21.02.23; accepted 23.02.23; published 30.03.23.

Please cite as:
McGillivray L, Keng-Meng Hui N, Wong QJJ, Han J, Qian J, Torok M
The Effectiveness of a Smartphone Intervention Targeting Suicidal Ideation in Young Adults: Randomized Controlled Trial Examining the Influence of Loneliness
JMIR Ment Health 2023;10:e44862
URL: https://mental.jmir.org/2023/1/e44862
doi:10.2196/44862
PMID:36995752

Original Paper

Tailoring a Digital Mental Health Program for Patients With Sickle Cell Disease: Qualitative Study

Cara Nikolajski[1], MPH, PhD; Julia O'Brien[2], RN, PhD; Emily Nardo[3,4], MS; Eva Szigethy[5], MD, PhD; Charles Jonassaint[3,4], MHS, PhD

[1]Center for High-Value Health Care, University of Pittsburgh Medical Center, Pittsburgh, PA, United States

[2]Department of Health and Community Systems, School of Nursing, University of Pittsburgh, Pittsburgh, PA, United States

[3]Center for Research on Health Care, University of Pittsburgh, Pittsburgh, PA, United States

[4]Center for Behavioral Health and Smart Technology, University of Pittsburgh, Pittsburgh, PA, United States

[5]Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Charles Jonassaint, MHS, PhD
Center for Research on Health Care
University of Pittsburgh
230 McKee Place
Suite 600
Pittsburgh, PA, 15213
United States
Phone: 1 412 586 9850
Email: cjonassaint@pitt.edu

Abstract

Background: Depression and other mental health disorders are prevalent among people living with chronic health conditions. Although digital cognitive behavioral therapy (CBT) is considered an effective treatment, African American individuals are less likely to engage in and adhere to digital therapies for mental health disorders compared with White individuals.

Objective: The aim of this study was to understand digital CBT mental health treatment perceptions and preferences of African American individuals with sickle cell disease (SCD).

Methods: African American individuals with SCD from various US locations were invited to participate in a series of focus groups. Participants were introduced to a health coach–supported mental health app and then asked a series of questions about the usability and appeal of the program as well as, more generally, what would make a digital mental health program effective for them. The authors reviewed the focus group transcripts and conducted a qualitative analysis of the results.

Results: A total of 25 people participated in 5 focus groups. Overall, 5 primary themes emerged regarding how app content and related coaching could be modified to enhance digital CBT engagement. These themes included connection with others living with SCD, the personalization of app content and coaching, characteristics of coaches, journaling and pain tracking, and considerations for optimal engagement.

Conclusions: Enhancing the user experience by making digital CBT tools relevant to patient populations is critical for optimizing program engagement and its uptake. Our findings highlight potential strategies to modify and design digital CBT tools for users with SCD and may also be applicable to patients with other chronic conditions.

Trial Registration: ClinicalTrials.gov NCT04587661; https://clinicaltrials.gov/ct2/show/NCT04587661

(JMIR Ment Health 2023;10:e44216) doi:10.2196/44216

KEYWORDS

mental health; sickle cell disease; digital health; cognitive behavioral therapy; digital cognitive behavioral therapy

Introduction

Background

African American individuals are more likely to experience severe depression than White individuals but less likely to receive treatment [1,2]. Depression and other mental health disorders are even more common in African American individuals living with chronic health conditions [3,4]. One particularly vulnerable population is individuals with sickle cell disease (SCD), a heritable condition that affects millions of people worldwide and >100,000 people in the United States, most of whom are of African descent [5]. Among patients with SCD, 38.8% present with mental health disorders [6]. A review of the literature on mental health symptoms in SCD found that 22% to 57% of individuals with SCD report clinically significant depressive symptoms depression when screened [7]. Similarly, anxiety disorders are also high in this population. In a worldwide study of quality of life among patients with SCD, 38% reported having symptoms of anxiety [8].

Depression and anxiety disorders not only lead to poorer quality of life but in patients with SCD, depression is associated with increased pain frequency and duration, pain-related hospitalizations, opioid dependence, medical costs, and risk of mortality [9,10]. Despite the high prevalence of mental health disorders and the links among depression, anxiety, and SCD-related health outcomes, mental health is rarely assessed, let alone treated as a part of routine SCD care. This is a major gap in care, given that effective and cost-efficient mental health treatments, such as cognitive behavioral therapy (CBT), exist but are often not part of comprehensive SCD care [11-15].

There are several barriers to the implementation of face-to-face behavioral health care in SCD clinics. Access to CBT is limited because of the expense of providing face-to-face therapy in low-resource clinical settings and a shortage of available CBT-trained therapists, especially those who are culturally competent to work with hard-to-reach minority populations. Even when CBT is available, therapy is time-consuming and requires traveling to a clinic, which is challenging for adult patients with major social and economic challenges [16-18] and for the parents of adolescent patients [19].

However, technology may be able to overcome these barriers and increase access to CBT among patients with SCD. Over the past 2 decades, face-to-face CBT has been adapted so that it can be delivered over the phone; over the internet via a desktop computer; and more recently, via mobile devices. These programs often combine a digital education component with personalized support, usually through a health coach [20,21]. Digital CBT makes it possible to deliver high-quality, evidence-based behavioral mental health treatments at scale in SCD care, in which face-to-face CBT may not be feasible [14,22]. The main limitation of digital CBT in SCD care is that its uptake and engagement are often poor among minority race and ethnic groups, and even when made available, patients with SCD may not use digital CBT [13]. One digital CBT trial showed that African American individuals in primary care with depression and anxiety were less likely than White individuals to start the digital CBT program (75% vs 87%) and complete all available modules (29% vs 43%) [12,13]. One potential explanation for racial differences in digital CBT uptake and engagement is that Black patients may find that the content is not relevant or relatable to them and does not represent their real-world cultural and health experience [14]. However, no qualitative studies have examined barriers to and strategies for improving the implementation of digital CBT in routine SCD care.

Objectives

The objective of this study was to conduct a series of qualitative focus groups to determine the factors that promote the successful implementation and uptake of a health coach–supported digital CBT program for mental health in SCD care. Among patients with SCD, adolescents and young adults are at high risk for depression and anxiety [23-25], have an increased risk of complications [26], and have the highest rate of acute care encounters and rehospitalizations [27]. Therefore, we focused on improving digital CBT uptake and engagement primarily among adolescents and young adults with SCD. The focus group findings described herein will be used to tailor an existing health coach–supported digital CBT program to enhance its implementation among individuals with SCD.

Methods

Study Design

We used a convenience sampling strategy, working with SCD-focused community-based organizations to recruit focus group participants residing in diverse locations across the United States. Participants were referred to the study team through these organizations and were contacted by the study team via email. Owing to the geographical dispersion of our sample and to maintain the safety of participant and study team during the COVID-19 pandemic, focus groups were internet based. The target participants for these focus groups were adolescents or young adults who reported having an “interest in or experience with mental health.” The target size for these focus group ranged from 8 to 10 participants. The focus group facilitator verbally provided potential participants with information regarding elements of informed consent before the start of each focus group (eg, risks and benefits, participation procedures, data use, and the voluntary nature of the project), and after verbal consent was obtained, the focus groups proceeded using an institutional review board–approved focus group guide. Our focus group guide included questions regarding the impact of SCD on mental health and their attitudes and preferences regarding digital tools designed to support management of mental health symptoms. During the focus groups, participants were shown screenshots of a health coach–supported digital CBT program called RxWell, which is offered by the University of Pittsburgh Medical Center Health Plan, to promote conversation related to the program’s elements, layout, and graphics. The app is publicly available on the Google Play and Apple app stores and is free of cost for University of Pittsburgh Medical Center Health Plan members. For each element of the digital CBT program (ie, in-app techniques, symptom tracking, and digital coaching), participants were asked to express their general perceptions of each element, why they would or would not engage with various program

components, their perceptions of the look and relatability of the program, and their preferences regarding coach communication. Focus group discussions were transcribed verbatim and participants received a US $30 food service gift card for their participation.

Qualitative Analysis

A total of 3 coders participated in the analysis. The primary qualitative coder led the development of the initial code book. The codebook draft was tested by 2 secondary coders who used the first 2 transcripts. All 3 coders then met to talk through and to finalize the codebook. The primary and one of the secondary coders coded all 5 transcripts, meeting after each focus group to discuss similarities and differences. The third coder adjudicated all the disagreements. We did not record precise frequencies of comments because of the difficulty in quantifying focus group dynamics (eg, a participant may be the first to comment, with others providing affirmation through head nods or other forms of nonverbal agreement). Thus, we have chosen to use words, such as “some” or “many,” to provide a meaningful reflection of the commonality of responses without quantification. Individual comments are noted as such or provided as a quotation.

Ethics Approval

The study protocol was approved by the University of Pittsburgh Human Research Protections Office (Study20070307) and was conducted in compliance with the ethical standards of the responsible institution on human participants as well as in accordance with the Declaration of Helsinki.

Results

Participant Characteristics

We recruited 25 African American individuals with SCD for this study. On the basis of their availability, the individuals participated in one of five 90-minute focus groups conducted between December 2020 and March 2021. The focus group size ranged from 3 to 9 participants. Participants were, on average, aged 23 years (range 16-47 years with 1 missing); 52% (13/25) were female, 32% (8/25) were male, 4% (1/25) were nonbinary, and 12% (3/25) did not disclose their gender. Only one participant had a previous exposure to RxWell.

Salient Themes

Overview

In total, 5 primary themes emerged, suggesting that app content and coaching could be modified to enhance digital CBT engagement in this population. Each theme is described in the subsequent section along with examples of quotes from focus group participants that align with each theme (Textbox 1).

Textbox 1. Focus group themes and example quotes.

Connection with others living with sickle cell disease

- "I've always struggled with finding like-minded people instead of trained professionals...I think it would also be beneficial if you had groups on there [the app] with people who have sickle cell that can share their thoughts and experiences."
- "I think it would also be cool if there was, like, a blog portion where...people with sickle cell specifically, could jot their thoughts...'This is how I deal with this,' that type of thing.'"
- "If this app was generalized [meant for general audiences], maybe there can be a link to the specialized sickle cell app."

Personalization of app content and coaching

- "This makes me think of data, and chatbots, and a bunch of things actually...The coach was recommending different, like, techniques that you could so...the app could also do that based on how your depression trends have been. Like, maybe it'll encourage you more to pick up certain different techniques to do more often."
- "I feel like a welcome screen...you could put in all of that information...how old are you, your name, if you have a chronic illness..."
- "Instead of having me go to the app, you know, maybe my coach can text me and say, boom, 'Okay [participant's name], how are you doing today? How are you feeling?'...Because some people feel like they don't have anybody."

Characteristics of coaches

- "For me to use it, it [the coach] would have to be...a [sic] actual trained professional not just, um, some customer service type. I want it to be someone who went to school for it, you know?"
- "I probably prefer someone who has experience with conditions that I have. So, if they can relate better – whether they have the condition themselves, or they've studied, or anything like that...For me personally, I think as long as they, like, have a good understanding of what I'm going through...should be good."
- "I also really like the coaching aspect...I think, like, if they were to have, like, a personalized bio—what would be most appealing is, like, just stating why users should trust them and why we should take their advice."
- "I already built my trust with my other coach, and I have to do it again with another coach. So, one set coach would be good like everybody else is saying."

Journaling and pain tracking

- "I've taken a lot of surveys that say, like, 'not at all' and 'several days' like that. I think I would just personally change how that's constructed a little bit to give it more of a journal feel."
- "It could be useful to have a pain scale as well on the app. So, you can kind of track your pain and if it's connected to your mood."

Considerations for optimal engagement

- "I have a really short attention span for things, so I get really bored with things easily. So, um, I really don't use those parts of the app [the techniques]. But another part of the app is you get connected with a coach...My coach just kind of checks in and asks me questions about my goals...That's been really helpful to always have someone just checking in...I don't want to say I'm advanced, but I've been kind of doing this CBT work and stuff with my therapists for years. So, none of this is new to me. And rehashing it is not very helpful to me."
- "I also think with our age group, specifically, I think it could be really helpful because most people our age...do a have smartphone. They do use apps. And I think that a person might be more inclined to use this app before they would go to, like, a therapist or speak with someone in person."
- "I don't know if I would sustain the use though...I guess it depends on everybody's need, right?"

Connection With Other People With SCD

Living with SCD was described as an isolating experience, and some participants expressed having difficulty finding opportunities to connect with individuals with similar lived experience. Some participants mentioned the potential benefit of a message board or chat function, in which individuals with SCD could connect and discuss experiences of dealing with the disease and depression, anxiety, or stress. Another group of participants acknowledged that it may be difficult to tailor the digital CBT program for people with a specific disease or condition. They suggested including links to alternative sites and resources within the app so that users could easily navigate as desired.

Personalization of App Content and of Coaching

All focus groups described elements of personalization that they felt would enhance their engagement with the digital CBT program. Most ideas pertained to the coaching component, but a few participants mentioned small ways in which the app itself could be personalized, such as including the use of the user's name throughout.

There were a number of suggestions for coaching personalization: (1) adding an intake survey that allows users to provide personal information and preferences that the coaches can use to tailor the experience; (2) integrating a chatbot into the program that could collect user data and further personalize the experience; (3) integrating some discussion of SCD into coach communications but not focusing on SCD as the sole

contributor to depression, anxiety, or stress; (4) having coaches send text messages occasionally to check in with users so that they do not always have to receive communications through the app; and (5) using other forms of messaging beyond typical coach check-in messages (eg, motivational quotes and pictures) to engage and support the mental health of users.

Characteristics of Coaches

The participants had mixed views on the coach's characteristics and experience. Some people preferred someone with a formal mental health training, whereas others wanted to connect with someone with similar lived experience (eg, one with SCD or another chronic condition). Otherwise, participants had little preference related to coach's characteristics, as long as the coaches were down-to-earth and nonjudgmental and did not use scripted responses. In addition, some participants felt that access to a brief coach biography would be beneficial for building a trusting relationship with the coach. Finally, the participants agreed that engaging with a single coach over time would be optimal.

Journaling and Pain Tracking

Several participants mentioned the potential benefits of incorporating a journaling element into the digital CBT program to provide more personal narrative regarding their day-to-day experiences and contributors to worsening or improved mood. This was also mentioned as an opportunity to log more detailed information beyond the standard validated depression and anxiety symptom assessments used in the digital CBT program. In addition, several participants acknowledged that pain was a common SCD symptom and contributed to the exacerbations of mental health symptoms; thus, in-app pain tracking and medication reminders would provide additional benefits.

Considerations for Optimal Engagement

Engagement with the digital mental health program involves interacting with the app to learn and practice new techniques, the completion of monthly symptom assessments, and regular engagement with a coach. Although more frequent engagements are recommended, there are no clear guidelines on the optimal level of engagement. There was substantial variability among the participants in terms of reporting how much they would engage with the program. Some participants mentioned that they might only engage with the program once or twice a week, only when needed, for a short period, or not at all, whereas others with more chronic depression or anxiety symptoms would consider using the app for a longer period. These differences in user needs may require a reconsideration of how to optimize engagement for various subgroups of digital CBT tool users.

Discussion

Principal Findings

People with SCD are at a high risk of developing mental health disorders [7,10,28,29]. Digital CBT has been found to be an effective, scalable, and cost-efficient method for treating mental health problems among patients with diverse psychiatric disorders [30,31]. Despite positive evidence for digital CBT, African American patients are less likely to use and engage with these types of interventions compared with White patients [13]. To fill this gap and to guide the design of these programs, it is critical to request insights from patients regarding what would be engaging and useful to them. Therefore, this study conducted a series of focus groups with adolescents and adults with SCD to better understand their perceptions of digital CBT, identify features that would enhance engagement in our digital CBT program, and take initial steps in developing a framework to guide the design of digital interventions that target underrepresented groups, such as patients with SCD. The focus groups revealed 5 themes that described best practices to promote engagement with a digital CBT program for mental health.

First, patients crave social connections and desire to feel a sense of belonging. Patients with SCD often express a fear of rejection when disclosing their SCD status and desire understanding and support from their peers. Symptoms of SCD impede engagement in social activities, employment, and school, and concerns regarding their peers' perceptions may discourage individuals with SCD from establishing relationships or seeking support [32-34]. Both adolescents and adults recognize the importance of all forms of social support in their lives; however, social support from other individuals with SCD plays a unique role, because these connections provide the opportunity to empathize with each other's experiences and share strategies and skills for coping with SCD.

The success and relevance of social support features in these apps have varied based on the target population [35-38], hence there is a need to evaluate the potential of interventions that increase social support to improve mental health among patients with SCD. Features that allow for increased social connection and encourage social identity among patients with SCD may yield additional mental health benefits that could lead to a further reduction in depression and other mental health disorders [39,40]. Social connections and support between peers living with SCD provide an opportunity to develop a social support network outside the app user's family and health care team [35]. Many participants in this study recommended tools that allow connection with other app users with similar struggles and experiences. Incorporating components of social connections, such as digital support groups or peer connect features, may enhance the efficacy of digital CBT.

The presence of a health coach may be another way to foster social connection. Health coaching or real-time support has been shown to be a critical component of the effectiveness of digital CBT [30,31]. However, there is no evidence to detail how digital health coaching can be tailored to SCD or other minority populations. The adoption of techniques to ensure personalized treatment may be beneficial. These include using the patient's name throughout, tailoring information specific to the individual user, including multiple modes of communication, and addressing patient concerns that are not directly related to disease pathology.

The literature varies, but several characteristics contribute to the effectiveness of peer coaches. Overall, it seems that the background, attitude, and empathy of the coach are the most important for building a relationship with the participant. Ideally,

coaches should share a similar background with the participant, either sharing the same chronic illness or having a similar chronic illness that shares management skills, or the coaches should come from the same community as the participant [41]. Community health workers, who are also a source of social support and may encourage behavior change or provide coaching that supports behavior change, are the most effective when they share a common background with the community they serve [41]. In addition, prior research on other chronic disease populations suggests that successful peer coaches may not need to have high disease-related self-efficacy; in fact, peer coaches who have struggled may be less intimidating for participants and may be more successful in building rapport with their mentee [42].

Our participants also expressed mixed preferences regarding health coach qualifications, with some wanting coaching from a person with SCD, while others said that a professional would be preferred. The varied needs and expectations of patients with SCD may affect what kind of health coach is the most suitable. In particular, patients who lack peer social support may benefit greatly from a health coach who has experience in managing SCD whom they can confide in. Likewise, those who already have strong social networks may prefer professionals. Preferences for a peer versus professional coach could be tied to what concerns are most salient for the individual participant, for example an eHealth intervention that used advanced practice registered nurses (APRNs) found that patients were comfortable texting the APRN regarding physical symptoms but were less comfortable discussing personal problems, even when invited to share such experiences [43].

Participants in this study did not want the digital CBT intervention to focus exclusively on SCD as the sole source of depression, anxiety, or stress, and the ability to talk to a peer coach regarding personal problems unrelated to SCD may be crucial for some participants. However, it is worth noting that these preferences are not universal. Jacob et al [43] reported that some patients with SCD expressed appreciation knowing that the APRNs cared and were available to talk even though they may not have chosen to discuss personal problems, and some enjoyed having conversations with the APRNs regarding their personal life in addition to communicating regarding their SCD. Although patient preferences regarding coach qualifications may vary, this finding reinforces the importance of the coach's interpersonal skills, whether they are a peer with SCD or a professional.

Given the variability in participants' preferences for coaches, the consideration of each participant's preferences during coach assignment could be beneficial in increasing engagement. Providing access to the background and expertise of health coaches may foster connections between users and coaches. For example, a successful study of peer coaches for patients with diabetes that saw improvements in hemoglobin A_{1C} used a booklet that included a photo and brief description of each coach to elicit participants' preferences [42]. A similar guidebook could be used in the future to pair digital CBT SCD participants with their coaches. Our participants also expressed the importance of having one health coach as the main point of contact for digital CBT during the course of the intervention. This approach may also be crucial to maintaining engagement, as consistency gives both coach and participant time to build a trusting relationship, and gives the coach the opportunity to learn more about the individual participant and provide the personalized support that the participant desires. Attempting to match participants and coaches as much as possible based on availability and preference may help prevent attrition and increase participant engagement in the digital CBT intervention.

Journaling and symptoms tracking were features that patients felt were important for inclusion in a mental health app. Several SCD studies have implemented e-diaries that provide an ecological momentary assessment that quantitatively tracks real-time disease-related symptoms [22,44]. Several participants in this study indicated that quantitative tracking of symptoms, such as pain, activity, mood, and sleep, is beneficial and an app feature that they thought is important for understanding their health. However, participants also emphasized the importance of narrative journaling to fully understand the factors associated with their health and wellness. To the best of our knowledge, no study has implemented a narrative journaling feature in mobile mental health apps for SCD. Journal features may provide users with an alternative method for reflection and increase self-awareness while providing additional mental health benefits [45,46]. Expressive writing is a well-known therapeutic intervention with both physical and psychological health benefits [47]. Thus, including a journaling feature in mental health apps, implemented as open-ended text or voice entry, may not only help patients identify factors associated with their symptoms but may also provide a therapeutic effect via emotional disclosure. The developers of behavioral mental health apps for this population should consider not only having an e-diary feature that allows for quantitative tracking of symptoms but also a narrative journal feature that allows more open-ended expression and documentation of an individual's day-to-day experiences.

Limitations

The sample obtained for this study was relatively small, with one focus group having only 3 of the 10 scheduled participants in attendance. Thus, some data may not encompass the unique preferences and needs of this population. In addition, this study intentionally focused on individuals with SCD, with a sample age range primarily limited to individuals aged <30 years (only one participant was aged >30 years). Older individuals may have different perspectives and preferences than younger individuals. Therefore, although much of the presented information may apply to other chronic disease populations and age groups, careful consideration should be given when translating these findings to other populations.

Digital CBT Recommendations and Conclusions

On the basis of the findings of this study, we have broad recommendations for designing digital CBT programs for adolescents and young adults with SCD.

Social Connection

Individuals with SCD look for ways to seek support from peers and find communities. Consider implementing a forum, support

groups, scoreboards, or other features that allow users to connect, communicate, and see the activities of other participants within the digital CBT app.

Personalization

Users feel more motivated to use digital CBT programs when the apps feel more personalized. This includes automation that allows for personalization, for example, having a chatbot calling the user by name and the suggestion of topics that are relevant to the user based on prior activity. Features that the user can control and make their own, such as choices for the welcome screen or theme of the app, colors, creating an avatar, etc, also provide an opportunity for customization.

Health Coach

Digital CBT health coaches need to have a background knowledge regarding SCD (ie, some basic training and knowledge of the condition, exposure to community) as well as adequate training in mental health and behavioral treatments. A person with lived experience as a patient with SCD, as well as mental health training, would be ideal, but when not available, having a background knowledge regarding SCD may be sufficient. Personalized human support along with digital education is necessary.

Symptom Monitoring

Digital CBT apps should have features such as journals, logbooks, or other spaces with the capacity to monitor symptoms and daily thoughts. This allows the user to identify connections between daily events and mental and physical symptoms.

Content

For young adults and adolescents, it is important to have quick, engaging, and multimedia-rich content. Content should be "bite sized," with enough new content and variety to satisfy diverse learning styles and information preferences among users. A one-size-fits-all approach has not been successful. Finally, the length of lessons should consider the cognitive load and competing demands on attention.

Optimizing digital CBT for individuals with SCD is vital for successful implementation of this intervention modality. Patients who perceive digital CBT as irrelevant, unrelatable, or culturally insensitive are less likely to initiate and sustain engagement in this treatment. The results of this study provide important insights into ways to design and implement digital CBT to advance the engagement in and effectiveness of digital CBT among individuals living with SCD. Personalizing the content delivered and coaching as much as possible, creating a desirable space for social connection between patients, and creating and delivering content that is relevant to patients with SCD while also acknowledging other contributors to mental health are all characteristics that will increase the successful implementation of digital CBT in routine SCD care. Researchers and developers should consider these findings as potential strategies for optimizing digital mental health support for this and other vulnerable populations.

Acknowledgments

This work was supported by National Institute of Mental Health Grant #R34MH125152 and National Institute of Nursing Research Grant #T32NR008857.

Data Availability

Data supporting the findings of this study are available upon reasonable request from the corresponding author, CJ.

Authors' Contributions

CJ, ES, and CN contributed to the study conception, design, and data collection. All authors contributed to the interpretation of the data and writing of the manuscript. All the authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Neighbors HW, Caldwell C, Williams DR, Nesse R, Taylor RJ, Bullard KM, et al. Race, ethnicity, and the use of services for mental disorders: results from the National Survey of American Life. Arch Gen Psychiatry 2007 Apr 01;64(4):485-494 [FREE Full text] [doi: 10.1001/archpsyc.64.4.485] [Medline: 17404125]
2. Williams DR, González HM, Neighbors H, Nesse R, Abelson JM, Sweetman J, et al. Prevalence and distribution of major depressive disorder in African Americans, Caribbean blacks, and non-Hispanic whites: results from the National Survey of American Life. Arch Gen Psychiatry 2007 Mar 01;64(3):305-315 [FREE Full text] [doi: 10.1001/archpsyc.64.3.305] [Medline: 17339519]
3. Stockdale S, Lagomasino IT, Siddique J, McGuire T, Miranda J. Racial and ethnic disparities in detection and treatment of depression and anxiety among psychiatric and primary health care visits, 1995-2005. Med Care 2008 Jul;46(7):668-677 [FREE Full text] [doi: 10.1097/MLR.0b013e3181789496] [Medline: 18580385]

4. Himle JA, Baser RE, Taylor RJ, Campbell RD, Jackson JS. Anxiety disorders among African Americans, blacks of Caribbean descent, and non-Hispanic whites in the United States. J Anxiety Disord 2009 Jun;23(5):578-590 [FREE Full text] [doi: 10.1016/j.janxdis.2009.01.002] [Medline: 19231131]
5. Brousseau DC, Panepinto JA, Nimmer M, Hoffmann RG. The number of people with sickle-cell disease in the United States: national and state estimates. Am J Hematol 2010 Jan;85(1):77-78 [FREE Full text] [doi: 10.1002/ajh.21570] [Medline: 20029951]
6. Jonassaint CR, Lukombo I, Feldman R, Driscoll J, Eack SM, Abebe KZ, et al. Differences in the prevalence of mental health disorders among Black American adults with sickle cell disease compared to those with non-heritable medical conditions or no medical conditions. Br J Haematol 2022 Feb 03;196(4):1059-1068. [doi: 10.1111/bjh.17962] [Medline: 34859423]
7. Jonassaint CR, Jones VL, Leong S, Frierson GM. A systematic review of the association between depression and health care utilization in children and adults with sickle cell disease. Br J Haematol 2016 Jul 15;174(1):136-147 [FREE Full text] [doi: 10.1111/bjh.14023] [Medline: 26991317]
8. Osunkwo I, Andemariam B, Minniti CP, Inusa BP, El Rassi F, Francis-Gibson B, et al. Impact of sickle cell disease on patients' daily lives, symptoms reported, and disease management strategies: results from the international Sickle Cell World Assessment Survey (SWAY). Am J Hematol 2021 Apr 01;96(4):404-417 [FREE Full text] [doi: 10.1002/ajh.26063] [Medline: 33264445]
9. Grant MM, Gil KM, Floyd MY, Abrams M. Depression and functioning in relation to health care use in sickle cell disease. Ann Behav Med 2000 Jun;22(2):149-157. [doi: 10.1007/bf02895779]
10. Pecker LH, Darbari DS. Psychosocial and affective comorbidities in sickle cell disease. Neurosci Lett 2019 Jul 13;705:1-6. [doi: 10.1016/j.neulet.2019.04.011] [Medline: 30965108]
11. Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The effectiveness of computerized cognitive behavioural therapy in routine care. Br J Clin Psychol 2006 Nov;45(Pt 4):499-514. [doi: 10.1348/014466505X84782] [Medline: 17076960]
12. Jonassaint CR, Belnap BH, Huang Y, Karp JF, Abebe KZ, Rollman BL. Racial differences in the effectiveness of internet-delivered mental health care. J Gen Intern Med 2020 Feb 19;35(2):490-497 [FREE Full text] [doi: 10.1007/s11606-019-05542-1] [Medline: 31745855]
13. Jonassaint CR, Gibbs P, Belnap BH, Karp JF, Abebe KZ, Rollman BL. Engagement and outcomes for a computerised cognitive-behavioural therapy intervention for anxiety and depression in African Americans. BJPsych Open 2017 Jan 02;3(1):1-5 [FREE Full text] [doi: 10.1192/bjpo.bp.116.003657] [Medline: 28058109]
14. Jonassaint C, Kang C, Prussien KV, Yarboi J, Sanger MS, Wilson JD, et al. Feasibility of implementing mobile technology-delivered mental health treatment in routine adult sickle cell disease care. Transl Behav Med 2020 Feb 03;10(1):58-67 [FREE Full text] [doi: 10.1093/tbm/iby107] [Medline: 30508141]
15. Palermo T, Dudeney J, Santanelli JP, Carletti A, Zempsky WT. Feasibility and acceptability of internet-delivered cognitive behavioral therapy for chronic pain in adolescents with sickle cell disease and their parents. J Pediatr Hematol Oncol 2018 Mar;40(2):122-127 [FREE Full text] [doi: 10.1097/MPH.0000000000001018] [Medline: 29176462]
16. Edwards CL, Scales MT, Loughlin C, Bennett GG, Harris-Peterson S, De Castro LM, et al. A brief review of the pathophysiology, associated pain, and psychosocial issues in sickle cell disease. Int J Behav Med 2005 Sep;12(3):171-179. [doi: 10.1207/s15327558ijbm1203_6] [Medline: 16083320]
17. Barakat L, Smith-Whitley K, Ohene-Frempong K. Treatment adherence in children with sickle cell disease: disease-related risk and psychosocial resistance factors. J Clin Psychol Med Settings 2002;9(3):201-209. [doi: 10.1023/A:1016047210623]
18. Thomas V, Taylor LM. The psychosocial experience of people with sickle cell disease and its impact on quality of life: qualitative findings from focus groups. Br J Health Psychol 2002 Sep;7(Part 3):345-363. [doi: 10.1348/135910702760213724] [Medline: 12614505]
19. Alberts NM, Law EF, Chen AT, Ritterband LM, Palermo TM. Treatment engagement in an internet-delivered cognitive behavioral program for pediatric chronic pain. Internet Interv 2018 Sep;13:67-72 [FREE Full text] [doi: 10.1016/j.invent.2018.07.005] [Medline: 30206521]
20. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec 15;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
21. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
22. Schatz J, Schlenz AM, McClellan CB, Puffer ES, Hardy S, Pfeiffer M, et al. Changes in coping, pain, and activity after cognitive-behavioral training: a randomized clinical trial for pediatric sickle cell disease using smartphones. Clin J Pain 2015 Jun;31(6):536-547 [FREE Full text] [doi: 10.1097/AJP.0000000000000183] [Medline: 25503599]
23. Kayle M, Docherty SL, Sloane R, Tanabe P, Maslow G, Pan W, et al. Transition to adult care in sickle cell disease: a longitudinal study of clinical characteristics and disease severity. Pediatr Blood Cancer 2019 Jan 24;66(1):e27463 [FREE Full text] [doi: 10.1002/pbc.27463] [Medline: 30251318]

24. Benton TD, Ifeagwu JA, Smith-Whitley K. Anxiety and depression in children and adolescents with sickle cell disease. Curr Psychiatry Rep 2007 Apr 26;9(2):114-121. [doi: 10.1007/s11920-007-0080-0] [Medline: 17389120]
25. Jerrell JM, Tripathi A, McIntyre RS. Prevalence and treatment of depression in children and adolescents with sickle cell disease. Prim Care Companion CNS Disord 2011 Mar 10:22m03310. [doi: 10.4088/pcc.10m01063]
26. Quinn C, Rogers ZR, McCavit TL, Buchanan GR. Improved survival of children and adolescents with sickle cell disease. Blood 2010 Apr 29;115(17):3447-3452 [FREE Full text] [doi: 10.1182/blood-2009-07-233700] [Medline: 20194891]
27. Brousseau DC, Owens PL, Mosso AL, Panepinto JA, Steiner CA. Acute care utilization and rehospitalizations for sickle cell disease. JAMA 2010 Apr 07;303(13):1288-1294. [doi: 10.1001/jama.2010.378] [Medline: 20371788]
28. Hasan S, Hashmi S, Alhassen M, Lawson W, Castro O. Depression in sickle cell disease. J Natl Med Assoc 2003 Jul;95(7):533-537. [Medline: 12911250]
29. Jenerette C, Funk M, Murdaugh C. Sickle cell disease: a stigmatizing condition that may lead to depression. Issues Ment Health Nurs 2005 Dec 09;26(10):1081-1101. [doi: 10.1080/01612840500280745] [Medline: 16284000]
30. Fitzsimmons-Craft EE, Taylor CB, Graham AK, Sadeh-Sharvit S, Balantekin KN, Eichen DM, et al. Effectiveness of a digital cognitive behavior therapy-guided self-help intervention for eating disorders in college women: a cluster randomized clinical trial. JAMA Netw Open 2020 Aug 03;3(8):e2015633 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.15633] [Medline: 32865576]
31. Rollman BL, Herbeck Belnap B, Abebe KZ, Spring MB, Rotondi AJ, Rothenberger SD, et al. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: a randomized clinical trial. JAMA Psychiatry 2018 Jan 01;75(1):56-64 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.3379] [Medline: 29117275]
32. Adegbola M, Barnes DM, Opollo JG, Herr K, Gray J, McCarthy AM. Voices of adults living with sickle cell disease pain. J Natl Black Nurses Assoc 2012 Dec;23(2):16-23 [FREE Full text] [Medline: 25003809]
33. Bulgin D, Tanabe P, Jenerette C. Stigma of sickle cell disease: a systematic review. Issues Ment Health Nurs 2018 Aug 13;39(8):675-686 [FREE Full text] [doi: 10.1080/01612840.2018.1443530] [Medline: 29652215]
34. Poku BA, Caress A, Kirk S. Adolescents' experiences of living with sickle cell disease: an integrative narrative review of the literature. Int J Nurs Stud 2018 Apr;80:20-28. [doi: 10.1016/j.ijnurstu.2017.12.008] [Medline: 29353708]
35. Crosby LE, Ware RE, Goldstein A, Walton A, Joffe NE, Vogel C, et al. Development and evaluation of iManage: a self-management app co-designed by adolescents with sickle cell disease. Pediatr Blood Cancer 2017 Jan 30;64(1):139-145 [FREE Full text] [doi: 10.1002/pbc.26177] [Medline: 27574031]
36. Gomolin A, Lebouché B, Engler K, Vedel I. Optimizing smartphone intervention features to improve chronic disease management: a rapid review. Health Informatics J 2020 Sep 12;26(3):1795-1809 [FREE Full text] [doi: 10.1177/1460458219891377] [Medline: 31826703]
37. Saberi P, Siedle-Khan R, Sheon N, Lightfoot M. The use of mobile health applications among youth and young adults living with HIV: focus group findings. AIDS Patient Care STDS 2016 Jun;30(6):254-260 [FREE Full text] [doi: 10.1089/apc.2016.0044] [Medline: 27214751]
38. Senn TE, Braksmajer A, Coury-Doniger P, Urban MA, Carey MP. Mobile technology use and desired technology-based intervention characteristics among HIV+ Black men who have sex with men. AIDS Care 2017 Apr 18;29(4):423-427 [FREE Full text] [doi: 10.1080/09540121.2016.1220479] [Medline: 27535069]
39. Haslam S, Haslam C, Cruwys T, Jetten J, Bentley SV, Fong P, et al. Social identity makes group-based social connection possible: implications for loneliness and mental health. Curr Opin Psychol 2022 Feb;43:161-165. [doi: 10.1016/j.copsyc.2021.07.013] [Medline: 34403958]
40. Werner-Seidler A, Afzali MH, Chapman C, Sunderland M, Slade T. The relationship between social support networks and depression in the 2007 national survey of mental health and well-being. Soc Psychiatry Psychiatr Epidemiol 2017 Dec 9;52(12):1463-1473. [doi: 10.1007/s00127-017-1440-7] [Medline: 28889230]
41. Hsu LL, Green NS, Donnell Ivy E, Neunert CE, Smaldone A, Johnson S, et al. Community health workers as support for sickle cell care. Am J Prev Med 2016 Jul;51(1 Suppl 1):S87-S98 [FREE Full text] [doi: 10.1016/j.amepre.2016.01.016] [Medline: 27320471]
42. Rogers EA, Hessler DM, Bodenheimer TS, Ghorob A, Vittinghoff E, Thom DH. Diabetes peer coaching: do "better patients" make better coaches? Diabetes Educ 2014 Nov 20;40(1):107-115. [doi: 10.1177/0145721713513178] [Medline: 24258250]
43. Jacob E, Stinson J, Duran J, Gupta A, Gerla M, Ann Lewis M, et al. Usability testing of a smartphone for accessing a web-based e-diary for self-monitoring of pain and symptoms in sickle cell disease. J Pediatr Hematol Oncol 2012 Jul;34(5):326-335 [FREE Full text] [doi: 10.1097/MPH.0b013e318257a13c] [Medline: 22627570]
44. McClellan CB, Schatz JC, Puffer E, Sanchez CE, Stancil MT, Roberts CW. Use of handheld wireless technology for a home-based sickle cell pain management protocol. J Pediatr Psychol 2009 Jun 23;34(5):564-573. [doi: 10.1093/jpepsy/jsn121] [Medline: 19029141]
45. Tan TT, Tan MP, Lam CL, Loh EC, Capelle DP, Zainuddin SI, et al. Mindful gratitude journaling: psychological distress, quality of life and suffering in advanced cancer: a randomised controlled trial. BMJ Support Palliat Care 2021 Jul 08:bmjspcare-2021-003068. [doi: 10.1136/bmjspcare-2021-003068] [Medline: 34244182]

46. Tanabe P, Porter J, Creary M, Kirkwood E, Miller S, Ahmed-Williams E, et al. A qualitative analysis of best self-management practices: sickle cell disease. J Natl Med Assoc 2010 Nov;102(11):1033-1041. [doi: 10.1016/s0027-9684(15)30730-6] [Medline: 21141291]
47. Pennebaker JW. Expressive writing in psychological science. Perspect Psychol Sci 2018 Mar 09;13(2):226-229. [doi: 10.1177/1745691617707315] [Medline: 28992443]

Abbreviations

APRN: advanced practice registered nurse
CBT: cognitive behavioral therapy
SCD: sickle cell disease

Edited by J Torous; submitted 16.11.22; peer-reviewed by Y Virella Pérez, T Timakum; comments to author 19.12.22; revised version received 11.02.23; accepted 13.02.23; published 06.04.23.

Please cite as:
Nikolajski C, O'Brien J, Nardo E, Szigethy E, Jonassaint C
Tailoring a Digital Mental Health Program for Patients With Sickle Cell Disease: Qualitative Study
JMIR Ment Health 2023;10:e44216
URL: https://mental.jmir.org/2023/1/e44216
doi:10.2196/44216
PMID:37023443

Review

Efficacy of Virtual Care for Depressive Disorders: Systematic Review and Meta-analysis

Crystal Edler Schiller[1], PhD; Julianna Prim[1], PhD; Anna E Bauer[1], PhD; Linda Lux[1], MPA; Laura Claire Lundegard[1], BSc; Michelle Kang[1]; Samantha Hellberg[1], MA; Katherine Thompson[1], PhD; Theresa Webber[1], BA; Adonay Teklezghi[1], BA; Noah Pettee[1]; Katherine Gaffney[1], MD; Gabrielle Hodgins[1], MD; Fariha Rahman[1], BSc; J Nikki Steinsiek[1], MD, MPH; Anita Modi[1], BSc; Bradley N Gaynes[1], MD, MPH

Department of Psychiatry, University of North Carolina School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Crystal Edler Schiller, PhD
Department of Psychiatry
University of North Carolina School of Medicine
University of North Carolina at Chapel Hill
304 MacNider Hall
Chapel Hill, NC, 27599-7160
United States
Phone: 1 919 966 4810
Email: crystal_schiller@med.unc.edu

Abstract

Background: The COVID-19 pandemic has created an epidemic of distress-related mental disorders such as depression, while simultaneously necessitating a shift to virtual domains of mental health care; yet, the evidence to support the use of virtual interventions is unclear.

Objective: The purpose of this study was to evaluate the efficacy of virtual interventions for depressive disorders by addressing three key questions: (1) Does virtual intervention provide better outcomes than no treatment or other control conditions (ie, waitlist, treatment as usual [TAU], or attention control)? (2) Does in-person intervention provide better outcomes than virtual intervention? (3) Does one type of virtual intervention provide better outcomes than another?

Methods: We searched the PubMed, EMBASE, and PsycINFO databases for trials published from January 1, 2010, to October 30, 2021. We included randomized controlled trials of adults with depressive disorders that tested a virtual intervention and used a validated depression measure. Primary outcomes were defined as remission (ie, no longer meeting the clinical cutoff for depression), response (ie, a clinically significant reduction in depressive symptoms), and depression severity at posttreatment. Two researchers independently selected studies and extracted data using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Risk of bias was evaluated based on Agency for Healthcare and Research Quality guidelines. We calculated odds ratios (ORs) for binary outcomes and standardized mean differences (SMDs) for continuous outcomes.

Results: We identified 3797 references, 24 of which were eligible. Compared with waitlist, virtual intervention had higher odds of remission (OR 10.30, 95% CI 5.70-18.60; N=619 patients) and lower posttreatment symptom severity (SMD 0.81, 95% CI 0.52-1.10; N=1071). Compared with TAU and virtual attention control conditions, virtual intervention had higher odds of remission (OR 2.27, 95% CI 1.10-3.35; N=512) and lower posttreatment symptom severity (SMD 0.25, 95% CI 0.09-0.42; N=573). In-person intervention outcomes were not significantly different from virtual intervention outcomes (eg, remission OR 0.84, CI 0.51-1.37; N=789). No eligible studies directly compared one active virtual intervention to another.

Conclusions: Virtual interventions were efficacious compared with control conditions, including waitlist control, TAU, and attention control. Although the number of studies was relatively small, the strength of evidence was moderate that in-person interventions did not yield significantly better outcomes than virtual interventions for depressive disorders.

(JMIR Ment Health 2023;10:e38955) doi:10.2196/38955

KEYWORDS

depression; virtual; treatment; therapy; efficacy; virtual care; meta-analysis; review; mental health; depressive disorder; virtual intervention; digital intervention; digital health; eHealth; health outcome; digital mental health; health intervention

Introduction

Prior to the COVID-19 pandemic, the lifetime prevalence of major depressive disorder (MDD) was over 20% for adults in the United States [1], and the majority (71%) of cases were untreated [2]. Compared with the prepandemic period, depressive symptoms became over three times more prevalent [3] during the pandemic, with up to 48% of citizens of developed nations reporting clinically significant depression [4]. At the same time, pandemic constraints critically challenged the provision of mental health services. Cost-effective, scalable, affordable, and accessible interventions were urgently needed, and the use of virtual care expanded quickly [5]. However, the efficacy of modern virtual interventions had not been systematically examined. Thus, the aim of this systematic review and meta-analysis was to fill this gap to inform clinical, administrative, and policy decision-making.

Prior systematic reviews and meta-analyses examined the evidence supporting the efficacy of computerized or virtual cognitive behavioral therapy (CBT) for MDD or depressive symptoms compared with no treatment or treatment as usual (TAU) (ie, referring participants to primary care providers or other health clinics in their local community to manage their depressive symptoms). Moreover, meta-analyses [6] and umbrella summaries across meta-analyses [7] have suggested that virtual treatment works at least as well as in-person treatment for those with depressive symptoms. Prior meta-analyses of virtual treatments for adults included studies conducted before 2016, and many included adults with depressive symptoms or various depression and anxiety diagnoses [6,8-10]. Since 2016, individual studies of virtual interventions have proliferated, expanding beyond CBT [11,12], and increased in rigor. As a result, a comparison of modern virtual interventions with not only waitlist or TAU but also with face-to-face interventions [6,13] for adults with MDD is feasible and warranted given that face-to-face psychotherapy had become impractical and, in certain settings, impossible.

Information evaluating whether virtual care is an efficacious alternative to individual, face-to-face intervention with a therapist is needed for clinicians, health systems, payers, and policymakers. In addition, data to guide decisions about which existing virtual interventions are most efficacious for treating depressive disorder are lacking. In the absence of such data, common assumptions about the superiority of in-person treatment have guided clinical decisions and policies regarding depression treatment.

The purpose of this systematic review was to answer three clinically relevant key questions (KQs) for depressive disorders (ie, MDD, persistent depressive disorder, or dysthymia) based on studies conducted in the last 10 years.

KQ1: Does virtual intervention provide better clinical outcomes than no treatment, TAU, or attention control, defined as a rigorous control condition that simulates active treatment without the active ingredient (ie, does it work)?

KQ2: Does in-person intervention provide better outcomes than virtual intervention (ie, is in-person intervention better)?

KQ3: Does one type of virtual intervention provide better outcomes than another type of virtual intervention (ie, what works best)?

The KQs were structured based on Agency for Healthcare and Research Quality (AHRQ) guidance for decision-making related to best practices in treatment [14].

Methods

Design

We used the Cochrane Handbook for Systematic Reviews of Interventions methods [15] and AHRQ guidance for grading the strength of evidence [16]. The protocol for this systematic review and meta-analysis is published in the Open Science Framework [17]. Reporting conforms to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [18].

Search Strategy and Selection Criteria

For this systematic review and meta-analysis, we searched the PubMed, EMBASE, and PsycINFO databases for trials published from January 1, 2010, to October 30, 2021, for MeSH (Medical Subject Headings) and major headings listed in Table A1 of Multimedia Appendix 1. Relevant systematic reviews and meta-analyses were used to identify additional existing literature, and ClinicalTrials.gov was searched to identify unpublished trials.

The study criteria were selected to inform clinical decision-making and policy in the United States. Eligible studies were randomized controlled trials (RCTs) of adults with a clinical diagnosis of MDD, dysthymic disorder, or persistent depressive disorder that tested a virtual psychological intervention for depression in at least one study arm, reported an outcome using a validated depression measure (see Table A2 of Multimedia Appendix 1), and were conducted in countries with a very high human development index (see Table 3 of Multimedia Appendix 1 for a list of eligible countries, [19]). To ensure generalizability of the results to individuals with major depression with access to current technology in the United States, we included studies conducted in similarly highly developed nations. To ensure comparability across studies, we included studies with standard, validated measures of depression, both self-reported and clinician-rated. Because evidence-based treatments for depression differ for children and adults, we excluded studies of children from this review.

References identified through searches were imported into Covidence Systematic Review software (Veritas Health Innovation, Melbourne, Australia). Two reviewers independently

screened the titles and abstracts of all references according to the inclusion and exclusion criteria. Studies included by either reviewer were retrieved for full-text screening by two independent reviewers for eligibility. Discrepancies between reviewers were resolved through discussions and consensus.

Data Extraction

One author extracted summary data from the included trials into standardized forms, and a second senior author (BNG, CES, or LL) checked the data for accuracy. Two authors independently rated the risk of bias across nine domains (see Table A4 in Multimedia Appendix 1) using the Cochrane Risk of Bias tool for RCTs [20] modified for psychotherapy outcome research [21]. Disagreements were resolved by discussion and consensus. Trials with a high risk of bias were excluded, although sensitivity analyses were performed to determine the impact on the results (see Figure A1 in Multimedia Appendix 1).

Data Synthesis and Analysis

Primary outcomes were rates of remission (ie, no longer meeting the clinical cutoff for depression), rates of response (ie, a clinically significant reduction in depressive symptoms), and depression severity at posttreatment. We calculated odds ratios (ORs) with 95% CIs for remission and response, and calculated standardized mean differences (SMDs, Cohen *d*) with 95% CIs for differences in symptom severity between groups. Forest plots were generated for all outcomes with sufficient data.

To determine the appropriateness of quantitative analyses, the senior authors (BNG, CES, LL, AEB) assessed the clinical and methodological heterogeneity of studies under consideration [15]. We performed meta-analyses using the *meta* package (v 4.19-2) in R version 3.6.1 [22] when two or more trials reported data on outcomes of interest with low levels of heterogeneity. Effect sizes were weighted by their inverse variance. To account for variability in the different study populations, we used random-effects models to estimate pooled or comparative effects with three or more studies. However, because the effect estimates from smaller studies (which are generally more prone to bias) are more influential in random-effects models, we used fixed-effects models in analyses with fewer than three studies to ensure that a small study would not overinfluence the estimates [22].

Statistical heterogeneity in effects between studies included in each meta-analysis was assessed by calculating the χ^2 statistic (Q) and the I^2 statistic, assessing the proportion of variation in study estimates due to heterogeneity rather than sampling error [15]. In instances of high heterogeneity, we performed sensitivity analyses to determine the extent to which excluding dissimilar studies changed the overall effect estimates. Most studies had only two study arms (ie, intervention and control); however, two studies had two intervention arms in addition to a waitlist control arm [23,24]. A two-level model was used if there was no significant difference between the three-level model and a two-level model based on a likelihood ratio test [22]. Using AHRQ guidelines [16], we assessed the overall strength of evidence (SOE) considering four factors: directness, consistency, precision, and bias. SOE was assessed by one author (LL) and checked for consensus with two other authors (BNG and CES). We began each SOE assessment with a rating of high and downgraded for factors that reduced the level of confidence. The resulting definitions of high, moderate, low, and insufficient SOE grades are summarized in Table A5 of Multimedia Appendix 1.

Results

Characteristics of Included Studies

Database and manual searching yielded 3797 citations for consideration (Figure 1), 24 of which met the criteria for inclusion in this review. The characteristics of participants included in each study are summarized in Table 1. Participants in all of the included trials were diagnosed with MDD, and those with severe psychiatric comorbidities such as any psychotic disorder or active substance use disorder, bipolar disorder, or acute risk of suicidality were excluded.

Figure 1. Study selection. *See Table A2 in Multimedia Appendix 1.

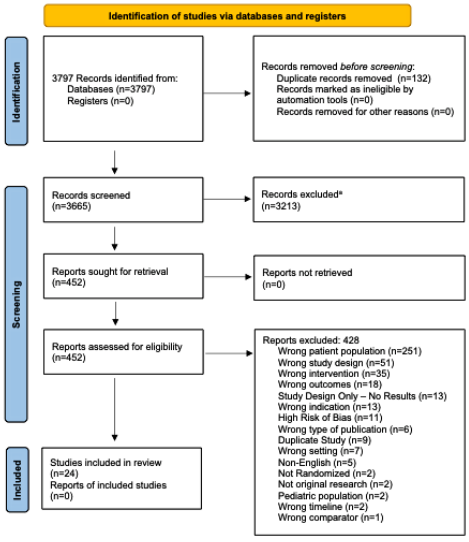


Table 1. Participant characteristics in each included trial.

Reference[a]	MDD[b] diagnostic measure[c]	Average MDD severity at baseline	Intervention condition		Comparison condition		Age (years), mean (SD)	Women, n (%)	Some college[d], n (%)
			Participants, n	Completed posttreatment assessment, n (%)	Participants, n	Completed posttreatment assessment, n (%)			
KQ[e]1a (virtual vs waitlist)									
Berger et al [23]	MINI[f]	Moderate	25	22 (88)	26	22 (85)	39 (14)	36 (70)	32 (63)
Berger et al [23]	MINI	Moderate	25	25 (100)	26	22 (85)	39 (14)	36 (70)	32 (63)
Carlbring et al [25]	SCID[g]	Moderate	40	40 (100)	40	38 (95)	44 (14)	66 (83)	61 (76)
Chan et al [11]	Clinical interview	Moderate	167	109 (65)	153	144 (94)	27 (7)	234 (73)	288 (90)
Johansson et al [26]	SCID	Moderate	27	25 (91)	27	27 (100)	39 (NR[h])	31 (57)	23 (42)
Kenter et al [27]	CIDI[i]	Moderate	136	96 (69)	133	89 (67)	38 (11)	145 (54)	110 (41)
Smith et al [28]	MINI	Moderate	61	36 (59)	68	55 (81)	40 (13)	106 (82)	89 (69)
Vernmark et al [23]	SCID	Moderate	30	29 (97)	29	29 (100)	37 (13)	40 (68)	48 (82)
Vernmark et al [24]	SCID	Moderate	29	27 (93)	29	29 (100)	37 (13)	39 (68)	48 (82)
KQ1b (virtual vs TAU[j])									
Dennis et al [12][k]	SCID	Moderate	120	104 (87)	121	100 (83)	31 (6)	241 (100)	181 (75)
Forsell et al [29]	SCID	Moderate	22	21 (95)	20	18 (90)	30 (5)	42 (100)	30 (71)
Hallgren et al [30]	MINI	Moderate	317	273 (86)	312	256 (82)	43 (12)	472 (75)	390 (62)
Löbner et al [31]	ICD-10[l]	Mild to moderate	320	259 (81)	327	307 (94)	44 (13)	446 (69)	NR
Moreno et al [32]	MINI	Moderate	80	74 (93)	87	85 (98)	44 (12)	149 (89)	28 (17)
Pfeiffer et al [33]	Medical record	Moderate	167	109 (65)	163	129 (79)	52 (15)	66 (20)	281 (85)
Raevuori et al [34]	ICD	Moderate	63	57 (90)	61	51 (84)	25 (NR)	90 (73)	NR
Wozney et al [35][k]	SCID	Moderate to severe	32	26 (81)	30	24 (80)	29 (5)	62 (100)	14 (22)
KQ1c (virtual vs attention control)									
Flygare et al [36]	SCID	Moderate	48	36 (74)	14	11 (76)	45 (12)	47 (76)	NR
Johansson et al [37][m]	MINI	Moderate	46	42 (91)	46	46 (100)	47 (14)	64 (70)	77 (84)
Ly et al [38][m]	MINI	Moderate	40	36 (90)	41	36 (88)	36 (11)	57 (70)	51 (63)
Oehler et al [39]	MINI	Mild to moderate	173	125 (72)	174	127 (73)	42 (12)	274 (79)	229 (66)
Reins et al [40]	SCID	Moderate	65	49 (75)	66	53 (80)	42 (11)	100 (76)	94 (72)
KQ2 (in-person vs virtual)									
Andersson et al [41]	SCID	Moderate	36	33 (92)	33	32 (97)	42 (14)	54 (78)	NR
Egede et al [42]	SCID	Moderate	121	106 (88)	120	108 (90)	64 (5)	5 (2)	NR
Mohr et al [43]	HAMD[n]	Moderate	162	141 (87)	163	151 (93)	NR (NR)	NR	NR
Thase et al [13]	SCID	Moderate	77	67 (87)	77	66 (86)	46 (14)	102 (66)	152 (99)

[a]Each row represents an intervention arm. Some references are listed more than once because they provided data from multiple intervention arms.

[b]MDD: major depressive disorder.

[c]Participants of all trials were diagnosed with MDD.

[d]Some college means any self-reported level of educational attainment greater than high school or equivalent.

[e]KQ: key question.

[f]MINI: Mini International Neuropsychiatric Interview.

[g]SCID: Structured Clinical Interview for Axis-I Disorders.

[h]NR: not reported.

[i]CIDI: Composite International Diagnostic Interview.

[j]TAU: treatment as usual.

[k]All participants were diagnosed with MDD with perinatal onset.

[l]ICD-10: International Classification of Diseases, 10th revision.

[m]Included in systematic review but excluded from meta-analysis due to differences in methods from other studies.

[n]HAMD: Hamilton Rating Scale for Depression.

The characteristics of each trial, including the length of intervention, treatment modality and mode, provider type, and comparison condition, are summarized in Table 2.

Risk of bias assessments across the nine individual domains and an overall summary is presented for each study in Table A4 of Multimedia Appendix 1; detailed information on intervention outcomes is presented in Figures 2-5; and SOE ratings alongside a summary of results are presented in Table 3.

Table 2. Trial characteristics.

Reference[a]	Length of intervention (weeks)	Intervention condition			Comparison condition		
		Modality	Mode	Provider type	Modality	Mode	Provider
KQ[b]1a (virtual vs waitlist)							
Berger et al [23]	10	CBT[c] (Deprexis)	Online intervention, guided	Mental health specialist	Waitlist	NA[d]	None
Berger et al [23]	10	CBT (Deprexis)	Online intervention, unguided	None	Waitlist	NA	None
Carlbring et al [25]	7	ACT[e]/BA[f]	Online intervention, guided	Mental health specialist	Waitlist	NA	None
Chan et al [11]	6	CBT-I	Smartphone intervention, unguided	None	Waitlist	NA	None
Johansson et al [26]	8	CBT	Online intervention, guided	Mental health specialist	Waitlist	NA	None
Kenter et al [27]	8	Problem-solving therapy	Online intervention, guided	Student	Waitlist	NA	None
Smith et al [28]	12	CBT	Online intervention, guided	Mental health specialist	Waitlist	NA	None
Vernmark et al [24]	8	CBT	Individualized email therapy	Mental health specialist	Waitlist	NA	None
Vernmark et al [24]	8	CBT	Online intervention, guided	Mental health specialist	Waitlist	NA	None
KQ1b (virtual vs TAU[g])							
Dennis et al [12]	12	IPT[h]	Telehealth (telephone)	Nurses	TAU	In-person	Nurse
Forsell et al [29]	10	CBT	Online intervention, guided	Mental health specialist	TAU	In-person	OBGYN[i]
Hallgren et al [30]	12	CBT	Online intervention, guided	Mental health specialist	TAU	In-person	PCP[j]
Löbner et al [31]	6	CBT (Moodgym)+TAU	Online intervention, self-guided	PCP	TAU	In-person	PCP
Moreno et al [32]	24	Supportive therapy+medication	Telehealth (video)	Mental health specialist	TAU	In-person	PCP
Pfeiffer et al [33]	12	CBT (Beating the Blues)	Online intervention, guided+TAU	Peer support specialist	TAU+depression workbook	In-person	VA[k] physician
Raevuori et al [34]	8	CBT (Meru Health Program)	Smartphone intervention, guided	Mental health specialist	TAU	In-person	Mental health specialist
Wozney et al [35]	24	CBT (MOM: Managing Our Mood)	Handbook and telephone coaching	Trained coach	TAU	In-person	PCP
KQ1c (virtual vs attention control)							
Flygare et al [36]	8	CBT	Online intervention, guided	Mental health specialist	Psychoed[l]	Online intervention, guided	Mental health specialist
Johansson et al [37][m]	10	Psychodynamic therapy	Online intervention, guided	Mental health specialist	Psychoed	Online intervention, guided	Mental health specialist
Ly et al [38][m]	8	BA	Smartphone, guided	Mental health specialist	Mindfulness	Smartphone, guided	Mental health specialist
Oehler et al [39]	6	CBT (iFight Depression)	Online intervention, guided	Mental health specialist	Progressive muscle relaxation	Online intervention, guided	Mental health specialist

Reference[a]	Length of intervention (weeks)	Intervention condition			Comparison condition		
		Modality	Mode	Provider type	Modality	Mode	Provider
Reins et al [40]	6	CBT (GET.ON Mood Enhancer)	Online intervention, guided	Mental health specialist	Psychoed	Online intervention, unguided	None
KQ2 (in-person vs virtual)							
Andersson et al [41]	8	CBT	In-person group,8 sessions (60 min)	Mental health specialist	CBT	Online intervention, guided	Mental health specialist
Egede et al [42]	8	BA	In-person, 8 sessions (60 min)	Mental health specialist	BA	Telemedicine (video), 8 sessions (60 min)	Mental health specialist
Mohr et al [43]	18	CBT	In-person, 18 sessions (45 min)	Mental health specialist	CBT	Telemedicine (telephone), 18 sessions (45 min)	Mental health specialist
Thase et al [13]	20	CBT	In-person, 20 sessions (50 min)	Mental health specialist	CBT (Good Days Ahead)	Online intervention, guided (Good Days Ahead)	Mental health specialist

[a]Each row represents an intervention arm. Some references are listed more than once because they provided data from multiple intervention arms.
[b]KQ: key question.
[c]CBT: cognitive behavioral therapy.
[d]NA: not applicable.
[e]ACT: acceptance and commitment therapy.
[f]BA: behavioral activation.
[g]TAU: treatment as usual.
[h]IPT: interpersonal therapy.
[i]OBGYN: obstetrician/gynecologist.
[j]PCP: primary care provider.
[k]VA: Veteran's Administration.
[l]Psychoed: psychoeducation.
[m]Included in systematic review but excluded from meta-analysis due to differences in methods from other studies.

Figure 2. Forest plots of virtual intervention compared with waitlist control clinical outcomes. ΔBDI: Change in Beck Depression Inventory Score; ΔPHQ: Change in Patient Health Questionnaire-9 Score; BDI: Beck Depression Inventory; CES-D: Center for Epidemiologic Studies Depression Scale; MADRS-SR: Montgomery–Åsberg Depression Rating Scale – Self-Report Questionnaire; MDD: Major Depressive Disorder; PHQ-9: Patient Health Questionnaire-9.

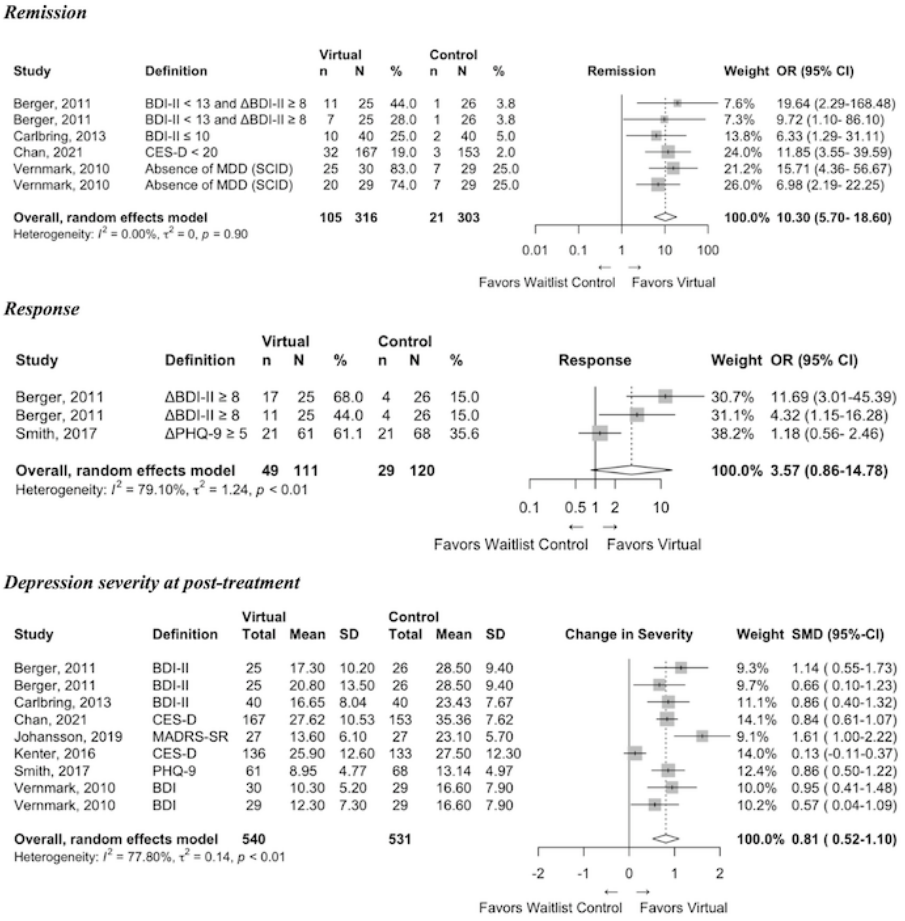


Figure 3. Forest plots of virtual intervention compared with treatment as usual (TAU) clinical outcomes. ΔMADRS: Change in Montgomery–Åsberg Depression Rating Scale Score; ΔMADRS-SR: Change in Montgomery–Åsberg Depression Rating Scale – Self-Report Questionnaire Score; BDI: Beck Depression Inventory; EPDS: Edinburgh Postnatal Depression Scale; MDD: Major Depressive Disorder; MADRS: Montgomery–Åsberg Depression Rating Scale Interview; MADRS-SR: Montgomery–Åsberg Depression Rating Scale – Self-Report Questionnaire; PHQ-9: Patient Health Questionnaire-9.

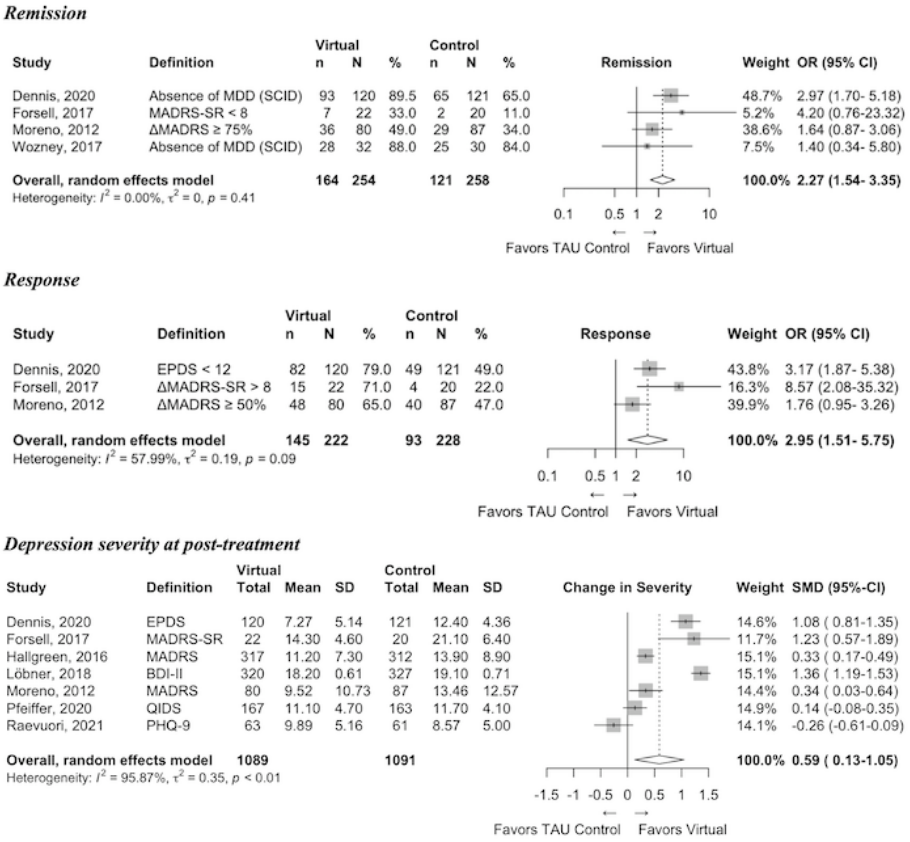


Figure 4. Forest plots for virtual intervention (internet-based cognitive behavioral therapy [iCBT]) compared with virtual sham intervention clinical outcomes. ΔBDI: Change in Beck Depression Inventory Score; BDI: Beck Depression Inventory; HAMD: Hamilton Depression Rating Scale; IDS-SR: Inventory for Depressive Symptomatology – Self-Report; MDD: Major Depressive Disorder; SCID: Semi-Structured Clinical Interview for DSM Disorders.

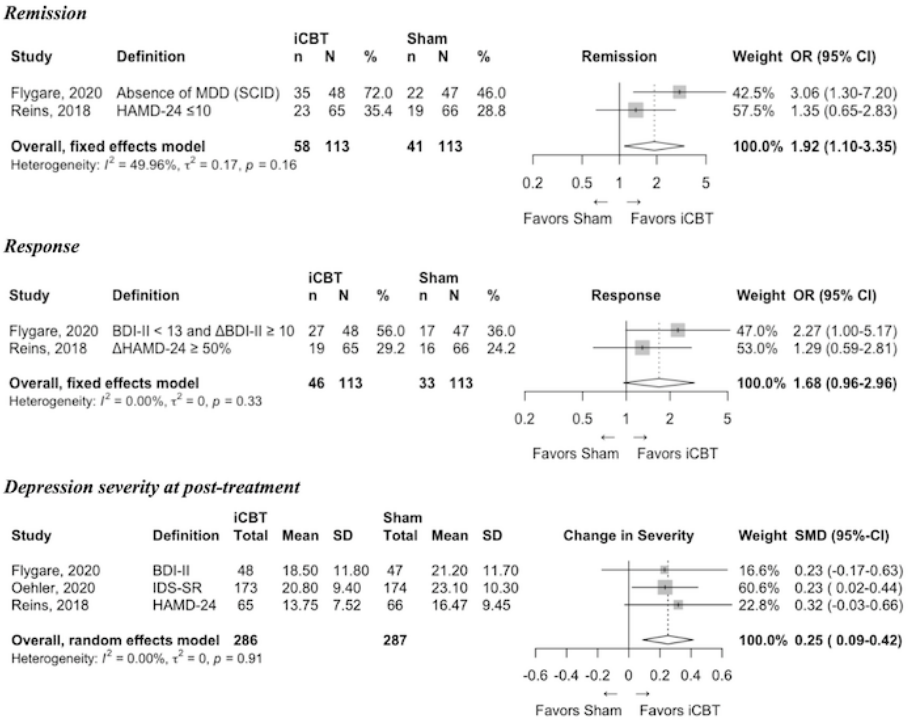


Figure 5. Forest plots for in-person intervention compared with virtual intervention clinical outcomes (key question 2). ΔBDI: Change in Beck Depression Inventory Score; ΔHAMD: Change in Hamilton Depression Rating Scale Score; ΔMADRS: Change in Montgomery–Åsberg Depression Rating Scale Score; BDI: Beck Depression Inventory; HAMD: Hamilton Depression Rating Scale; MDD: Major Depressive Disorder; SCID: Semi-Structured Clinical Interview for DSM Disorders.

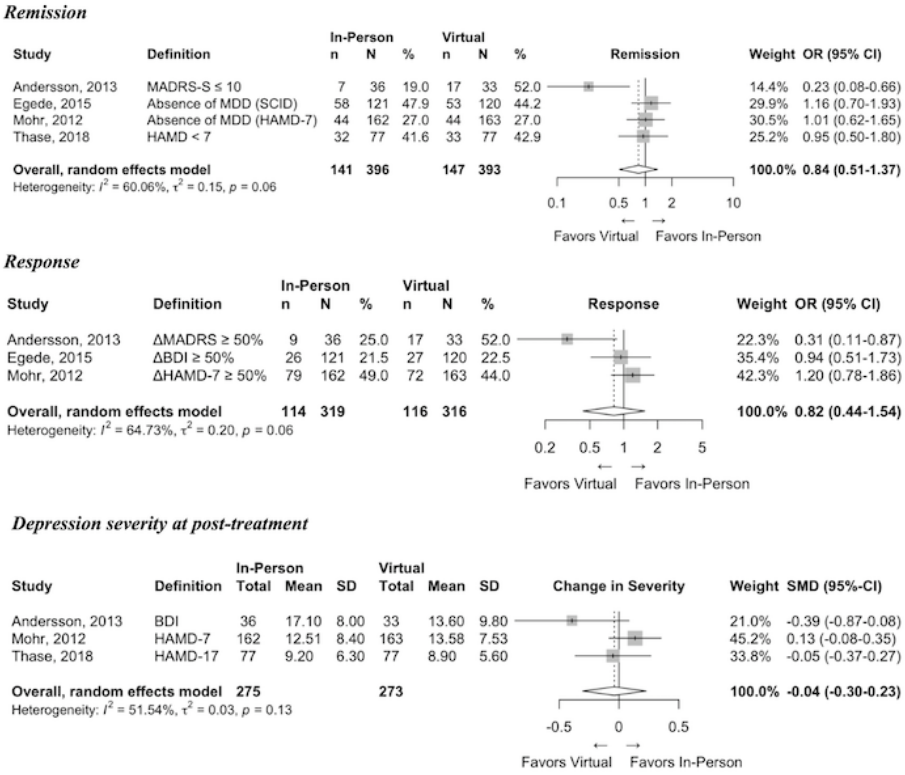


Table 3. Strength of evidence for each outcome organized by key question (KQ).

Outcome	Study design, duration, sample size (N), events[a] (n)	Effect size (95% CI)	Factors that affect the strength of evidence	Overall evidence strength and direction of effect	Findings
KQ1a (virtual vs waitlist)					
Remission	RCT[b] (4 trials), 7-10 weeks, N=564, n=118 enrolled; N=619, n=126 analyzed due to 2 trials each having two intervention arms comparing to one control group	OR[c] 10.30 (5.70 to 18.60)	Low ROB[d], imprecise estimate[e] but high effect (increase), direct consistent (I^2= 0%)	High; virtual intervention>waitlist	The SOE[f] is high that virtual interventions have 10 times higher odds of remission than waitlist
Response	RCT (2 trials), 10-12 weeks, N=195, n=74 enrolled; N=221, n=78 analyzed due to 1 trial having two intervention arms comparing to the same control group	OR 3.57 (0.86 to 14.78)	1 Low, 1 moderate ROB (decrease),imprecise estimate (decrease), direct inconsistent[g] (I^2=79.1%) (decrease)	Low; no statistically significant difference	The SOE is low that there are no substantial differences in response between virtual interventions and waitlist
Depression severity	RCT (7 trials), 7-12 weeks, N=1180, n=1180 enrolled; N=1071, n=1071 analyzed due to 2 trials each having two intervention arms comparing to one control group	SMD[h] 0.81 (0.52 to 1.10)	4 Low, 3 moderate ROB; precise estimate; direct inconsistent (I^2=77.8%) (decrease)	Moderate; virtual intervention>waitlist	The SOE is moderate that virtual interventions have greater reduction in depression severity compared with waitlist
KQ1b (virtual vs TAU[i])					
Remission	RCT (4 trials), 10-24 weeks; N=512, n=285	OR 2.27 (1.54 to 3.35)	Low ROB; imprecise estimate (decrease); direct consistent (I^2=0%)	Moderate; virtual intervention>TAU	The SOE is moderate that virtual interventions have 2 times higher odds of remission than TAU
Response	RCT (3 trials), 10-24 weeks; N=450, n=238	OR 2.95 (1.51 to 5.75)	Low ROB; imprecise estimate (decrease); direct consistent (I^2= 58.0%)	Moderate; virtual intervention>TAU	The SOE is moderate that virtual interventions have 3 times higher odds of response than TAU
Depression severity	RCT (7 trials), 8-24 weeks; N=1533, n=1533	SMD 0.59 (0.13 to 1.05)	5 Low, 2 Moderate ROB; precise estimate; direct inconsistent (I^2=95.9%)	Moderate; virtual intervention>TAU	The SOE is moderate that virtual interventions have greater reduction in depression severity compared with TAU
KQ1c (virtual vs attention control)					
Remission	RCT (2 trials), 6-8 weeks; N=226, n=99	OR 1.92 (1.10 to 3.35)	1 Low, 1 Moderate ROB (decrease); imprecise estimate (decrease); direct consistent (I^2=50%)	Low; virtual CBT[j]>attention control	The SOE is low that virtual CBT has 2 times greater odds of remission than virtual psychoeducation
Response	RCT (2 trials), 6-8 weeks; N=226, n=79	OR 1.68 (0.96 to 2.96)	1 Low, 1 Moderate ROB (decrease); imprecise estimate (decrease); direct consistent (I^2=0%)	Low; no statistically significant difference	The SOE is low there are no substantial differences in response between virtual CBT and virtual psychoeducation
Depression severity	RCT (3 trials), 6-8 weeks; N=573, n=573	SMD 0.25 (0.09 to 0.42)	2 Low, 1 Moderate ROB; precise estimate; direct consistent (I^2=0%)	High; virtual CBT>attention control	The SOE is high that virtual CBT has greater reduction in depression severity compared with virtual psychoeducation
KQ2 (in-person vs virtual)					
Remission	RCT (4 trials), 8-20 weeks; N=789, n=288,	OR 0.84 (0.51 to 1.37)	Low ROB; imprecise estimate (decrease); direct consistent (I^2=60.1%)	Moderate; no statistically significant difference, noninferiority trials[f]	The SOE is moderate that there are no substantial differences in remission between in-person and virtual interventions

Outcome	Study design, duration, sample size (N), events[a] (n)	Effect size (95% CI)	Factors that affect the strength of evidence	Overall evidence strength and direction of effect	Findings
Response	RCT (3 trials), 8-18 weeks; N=635, n=230	OR 0.82 (0.44 to 1.54)	Low ROB; imprecise estimate (decrease); direct consistent (I^2=64.7%)	Moderate; no statistically significant difference, noninferiority trials	The SOE is moderate that there are no substantial differences in response between in-person and virtual interventions
Depression severity	RCT (3 trials), 8-20 weeks; N=548, n=548	SMD –0.04 (–0.30 to 0.23)	Low ROB; precise estimate; direct consistent (I^2=51.5%)	Moderate; no statistically significant difference, noninferiority trials	The SOE is moderate that there are no substantial differences in posttreatment depression severity between in-person and virtual interventions

[a]Based on risk of bias, precision of estimate, directness of comparison, and consistency.

[b]RCT: randomized controlled trial.

[c]OR: odds ratio.

[d]ROB: risk of bias.

[e]Imprecision is based on the number of events <300 events, or n=400 for continuous events or very wide confidence intervals; precision was the primary variable that influenced strength of evidence ratings given that most trials had low risk of bias and were direct and consistent.

[f]SOE: strength of evidence.

[g]Inconsistent was based on I^2>75%.

[h]SMD: standardized mean difference.

[i]TAU: treatment as usual.

[j]CBT: cognitive behavioral therapy.

Efficacy of Virtual Intervention Versus Waitlist Control (KQ1a)

The efficacy of virtual interventions compared with waitlist was assessed in seven double-blinded RCTs [11,23-28] (Table 2). Most trials compared virtual CBT with waitlist [11,23,24,26,28]: four trials examined virtual CBT guided by mental health providers [23,24,26,28], two examined unguided virtual CBT [11,24], and one examined CBT provided via email [24]. Two studies examined virtual adaptations of evidence-based therapies other than CBT (ie, combined acceptance and commitment therapy and behavioral activation [BA] [25] and problem-solving therapy [27]).

Remission was evaluated in four trials [11,23-25]. Two of the trials included comparisons of two different intervention arms against waitlist control groups: Berger et al [23] examined both guided and unguided virtual CBT compared with waitlist, and Vernmark et al [24] examined both guided virtual CBT and CBT provided via email compared with waitlist. Meta-analysis including a total of five comparisons across three studies showed that the odds of remission were 10 times higher (95% CI 5.70-18.60; N=619; high SOE) with virtual intervention compared with waitlist (Figure 2). Response was measured in three separate comparisons across two studies [23,28]. The odds of response did not substantially differ between virtual interventions and waitlist (OR 3.57, 95% CI 0.86-14.87; N=221; low SOE). Depression severity at posttreatment was assessed in seven trials [11,22–27]. Virtual interventions resulted in lower depression severity at posttreatment compared with waitlist (SMD 0.81, 95% CI 0.52-1.10; N=1071; moderate SOE).

Efficacy of Virtual Intervention Versus TAU (KQ1b)

Efficacy of virtual interventions compared with TAU was evaluated in eight double-blinded RCTs [12,29-35] (Table 2). Three of the trials focused on interventions for specific populations: those with perinatal-onset MDD [12,29,35], a majority-male veteran cohort [33], and a Latinx Spanish-speaking population [32] (Table 1). Most virtual interventions involved guided virtual CBT [29-31,33-35]; however, two trials provided synchronous telehealth interventions, including interpersonal therapy delivered by nurses via telephone [12] and supportive therapy plus pharmacotherapy provided by a psychiatrist via video visits [32]. Most TAU study arms consisted of primary care appointments, scheduled on an as-needed basis delivered by physicians [29-33,35].

Remission was evaluated in four trials [12,29,32,35]. The odds of remission were two times higher with virtual intervention compared with TAU (OR 2.27, 95% CI 1.54-3.35; N=512; moderate SOE) (Figure 3). Response was evaluated in three trials [12,29,32]. The odds of response were nearly three times higher with virtual intervention compared with TAU (OR 2.95, 95% CI 1.51-5.75; N=450; moderate SOE). Depression severity was evaluated in seven trials [12,29-34]. Virtual intervention resulted in a lower depression severity at posttreatment compared with TAU (SMD 0.59, 95% CI 0.13-1.05; N=1533; moderate SOE).

Efficacy of Virtual Therapy Versus Attention Control (KQ1c)

Five trials compared a virtual adaptation of an evidence-based intervention (eg, CBT [36,39,40], BA [38], or psychodynamic therapy [37]) with a virtual control (ie, mindfulness [38]) or sham condition (Table 1). Of these, three studies compared

virtual CBT with attention control conditions, which included online psychoeducation [36,40] and progressive muscle relaxation [39]. These three studies were included in one set of meta-analyses based on the consistency in interventions (virtual CBT) and attention control conditions (Figure 4).

Remission and response were assessed in two trials [36,40], both of which compared virtual CBT to virtual psychoeducation and favored virtual CBT in terms of both remission and response (Figure 4). The odds of remission were higher with virtual CBT compared with virtual psychoeducation (OR 1.92, 95% CI 1.10-3.35; N=226; low SOE), whereas there was no statistically significant difference in response rates between virtual CBT and virtual psychoeducation (OR 1.68, 95% CI 0.96-2.96; N=226; low SOE). Depression severity at posttreatment was evaluated in three trials [36,39,40]. All three studies favored virtual CBT compared with an attention control condition. Virtual CBT resulted in lower depression severity at posttreatment compared with virtual psychoeducation (SMD 0.25, 95% CI 0.09-0.42; N=573; high SOE).

Efficacy of In-Person Versus Virtual Intervention (KQ2)

Efficacy of in-person compared with virtual delivery of behavioral therapy, either CBT or BA, was evaluated in four RCTs [13,41-43] (Table 1), three of which were noninferiority trials [41-43] powered to evaluate whether virtual therapy provides at least the same benefit to the patient as in-person therapy. Two of the virtual interventions were guided virtual CBT [13,41] and two were synchronous telehealth interventions [42,43]. One trial compared in-person group CBT with virtual CBT [41].

Remission was evaluated in four trials [13,41,43]. In no study did remission rates for the in-person intervention exceed those seen in virtual interventions (Figure 5). Indeed, one trial reported significantly lower remission rates in the in-person CBT groups compared with virtual CBT (19% vs 52%; P<.005) [41]. The odds of remission with the in-person intervention were not higher than those with the virtual intervention (OR 0.84, 95% CI 0.51-1.37; N=789; moderate SOE). Sensitivity analysis excluding the study of Andersson et al [41], as the only trial comparing group therapy with virtual therapy, similarly indicated no significant difference but resolved the heterogeneity (OR 1.05, 95% CI 0.77-1.43; I^2=0) (Multimedia Appendix 1, Figure A1).

Response was evaluated in three trials [41-43]. None of the trials reported better outcomes for the in-person arm. Indeed, after 8 weeks of intervention, in-person group CBT produced a significantly *lower* response rate than virtual CBT (25% vs 52%; P=.02) [41]. The odds of response with in-person intervention were no higher than those with virtual intervention (OR 0.82, 95% CI 0.44-1.54; N=635; moderate SOE) (Figure 5). Sensitivity analysis dropping the outlier [41] similarly indicated no statistically significant difference (Multimedia Appendix 1, Figure A1).

Depression severity was evaluated in all four trials [13,41-43]; none reported a benefit for in-person compared to virtual intervention. One trial comparing BA delivered in person versus via telehealth reported no statistically significant difference in depressive severity at posttreatment, but did not provide sufficient quantitative data [40] and was thus excluded from meta-analysis. In-person intervention was not associated with lower depression severity at posttreatment compared with virtual interventions (SMD –0.04, 95% CI –0.30 to 0.23; N=548; moderate SOE) (Figure 5) for the remaining three trials.

Comparative Efficacy of Various Virtual Interventions (KQ3)

No trials comparing the efficacy of one virtual intervention with another virtual intervention were identified in our searches that met our inclusion criteria.

Discussion

Principal Results

Virtual intervention for individuals with mild to moderate depressive disorders resulted in higher remission rates and a lower severity of symptoms at posttreatment compared with waitlist, TAU, and attention control conditions. There was no consistent evidence that an in-person intervention is significantly more efficacious than a virtual intervention for depression. Two studies compared telehealth with in-person sessions, while two studies compared virtual behavior therapies with in-person sessions. Despite these methodological differences, heterogeneity between studies was low and sensitivity analyses showed no difference in results if any study was removed. Studies included individuals with mild to moderate depressive disorders across a number of patient populations, including primary care patients, veterans, perinatal women, and Spanish-speaking Latinx individuals, suggesting relatively broad generalizability to depressed populations in countries with a very high human development index. Of note, we found no eligible studies comparing the effectiveness of different active virtual interventions, which is an important research and clinical gap that should be addressed in future trials. Taken together, the results suggest that virtual therapy is an effective method of treatment for mild to moderate depressive disorders. The results further suggest a lack of clear evidence that in-person treatment is superior to virtual treatment for those with mild to moderate depressive disorders without significant comorbidity and living in countries with a very high human development index.

Given the significant limitations in access to evidence-based care in the United States, this represents a potential opportunity to increase access to effective and affordable treatment. Despite the finding that, on average, there is not reliable evidence that in-person treatment is superior to virtual treatment for depressive disorders, critical research to identify which patients benefit most from in-person and virtual treatment has not been done. Some patients may benefit more from in-person therapy than virtual treatment. Many people do not have access to high-speed internet, a private space for virtual sessions, or a home environment that is safe or conductive to engaging in therapy at home. None of the studies included in this review addressed these important individual differences that may differentially impact treatment feasibility, acceptability, and outcomes. As

such, in-person therapy for mood disorders remains an important first-line treatment option. However, virtual therapy can be considered an additional first-line treatment option, particularly for those who prefer it and those without transportation, time, or geographical access to in-person treatment.

Limitations

This systematic review and meta-analysis had important limitations. For some of the outcomes, a low number of events (ie, remission or response) observed across a small number of studies reduced the SOE, particularly in the case of response, which had the lowest number of observations of any outcome assessed and resulted in low to moderate SOE ratings across each key question. Our review was narrowly focused on depression intervention outcomes: all of the trials that met the inclusion criteria focused on interventions for MDD, and despite inclusion criteria of all depressive disorders, none examined other depressive disorders or other common co-occurring conditions such as anxiety disorders. Only one study included individuals with severe depressive symptoms, and therefore conclusions cannot be made regarding the utility of virtual therapy for those with more severe presentations. Interventions ranged in duration from 6 to 24 weeks. We evaluated immediate effects of the intervention on depression outcomes; however, due to variability in follow-up assessments, we did not examine long-term outcomes. Patient adherence to the intervention was not defined consistently across studies, and intervention fidelity was not assessed in most studies. Thus, neither variable could be evaluated as part of our risk of bias assessment. The studies included in KQ2 evaluated heterogeneous treatment populations (eg, veterans, primary care patients), and different in-person (eg, group therapy, individual therapy) and virtual (eg, telephone therapy, video therapy, virtual CBT) treatments. Although sensitivity analyses suggested that the results were the same when eliminating heterogeneous studies, additional studies are needed to have strong confidence in the results.

Comparison With Prior Work

The results of this study were consistent with older meta-analyses establishing the efficacy of virtual CBT for depression and anxiety compared with no intervention [8-10], and with a recent meta-analysis examining the effectiveness [44] of open-label, nonrandomized virtual and other remote interventions for depression and anxiety, compared with control conditions. Similar to prior meta-analyses [10,45], we found that the effect size comparing virtual intervention with waitlist was larger than that for TAU. Our results were also similar to past meta-analyses showing that outcomes for virtual treatment were at least as good as outcomes for face-to-face therapy [45,46]. Our study extended these findings by including only individuals diagnosed with MDD and by examining not only depressive symptoms but also remission and response rates. To our knowledge, this was the first meta-analysis to compare virtual with face-to-face interventions for individuals with clinically confirmed diagnoses of depressive disorders with a focus on remission and response. Our review adds to the literature by (1) focusing on depressive disorders and not only depressive symptoms, which could be subthreshold, less severe, and less likely to show a difference between two interventions; and (2) including two noninferiority trials, which provides a stronger test of whether virtual therapy provides at least the same benefit as in-person therapy. With this more rigorous test, virtual interventions performed as well as face-to-face therapy.

Strengths

Strengths of this systematic review included a multidimensional approach to assessing risk of bias, based on both the Cochrane risk of bias tool [20] and guidelines for applying the Cochrane tool to psychotherapy trials [21]. Psychotherapy trials, by definition, do not allow for participant blinding in the same way as medication trials. Yet, the underlying principle of blinding is believability of or confidence in the intervention to a similar degree across both the active intervention and control conditions. Only one trial with an active control condition assessed or attempted to control for participants' confidence in the intervention. This absence represents a weakness in the psychotherapy literature that should be addressed by future trials. Other strengths of this systematic review were the inclusion of only RCTs, trials that required a depression diagnosis at baseline, use of a validated assessment of depression outcome, and those with low to medium risk of bias, which increased the strength of the conclusions.

Conclusions

These results carry implications for health systems and mental health clinicians, policymakers, and researchers. Mental health clinics with long waitlists for evidence-based interventions and primary care clinics offering TAU could improve patient outcomes, reduce wait times, and reserve face-to-face sessions with therapists for those with the most severe symptoms by providing virtual interventions. With the rates of depression reaching epidemic proportions during the COVID-19 pandemic, existing efficacious technological solutions can help reduce the burden on the health care system, increase access to mental health care, and reduce the risk of COVID-19 transmission in health care settings. Implementation research is needed to determine when and for whom virtual interventions work best and when they may serve as an alternative to face-to-face therapy. Studies examining the efficacy of virtual adaptations of other evidence-based interventions for depression (eg, BA, acceptance and commitment therapy), optimal amount of guidance for virtual interventions (eg, regularly scheduled or as-needed coaching), optimal format for provider involvement (eg, telephone or email), and degree of provider training (eg, peer support, trained coaches, or licensed mental health providers) are needed to guide clinical decision-making. Nevertheless, our results suggest that virtual interventions provide an efficacious mechanism for scaling-up depression interventions to meet the growing demands created by the COVID-19 pandemic.

Acknowledgments

We thank our colleague, Nate Sowa, MD, for raising the important questions of whether virtual interventions work, whether in-person interventions work better, and which virtual interventions work best.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information: Tables A1-A5; Figure A1.
[DOC File , 2331 KB - mental_v10i1e38955_app1.doc]

References

1. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. JAMA Psychiatry 2018 Apr 01;75(4):336-346 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.4602] [Medline: 29450462]
2. Olfson M, Blanco C, Marcus SC. Treatment of adult depression in the United States. JAMA Intern Med 2016 Oct 01;176(10):1482-1491. [doi: 10.1001/jamainternmed.2016.5057] [Medline: 27571438]
3. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
4. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
5. Zhu D, Paige SR, Slone H, Gutierrez A, Lutzky C, Hedriana H, et al. Exploring telemental health practice before, during, and after the COVID-19 pandemic. J Telemed Telecare 2021 Jul 09:1357633X211025943. [doi: 10.1177/1357633X211025943] [Medline: 34241545]
6. Andersson G, Topooco N, Havik O, Nordgreen T. Internet-supported versus face-to-face cognitive behavior therapy for depression. Expert Rev Neurother 2016 Dec 15;16(1):55-60. [doi: 10.1586/14737175.2015.1125783] [Medline: 26610160]
7. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019 Jul 16;64(7):465-470 [FREE Full text] [doi: 10.1177/0706743719839381] [Medline: 31096757]
8. Königbauer J, Letsch J, Doebler J, Ebert D, Baumeister H. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
9. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
10. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
11. Hui VKY, Wong C, Ma EKY, Ho FYY, Chan CS. Treating depression with a smartphone-delivered self-help cognitive behavioral therapy for insomnia: study protocol for a parallel group randomized controlled trial. Trials 2020 Oct 09;21(1):843 [FREE Full text] [doi: 10.1186/s13063-020-04778-1] [Medline: 33036655]
12. Dennis C, Grigoriadis S, Zupancic J, Kiss A, Ravitz P. Telephone-based nurse-delivered interpersonal psychotherapy for postpartum depression: nationwide randomised controlled trial. Br J Psychiatry 2020 Apr 07;216(4):189-196. [doi: 10.1192/bjp.2019.275] [Medline: 32029010]
13. Thase ME, Wright JH, Eells TD, Barrett MS, Wisniewski SR, Balasubramani G, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. Am J Psychiatry 2018 Mar 01;175(3):242-250 [FREE Full text] [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
14. Buckley DI, Ansari MT, Butler M, Soh C, Chang CS. The refinement of topics for systematic reviews: lessons and recommendations from the Effective Health Care Program. J Clin Epidemiol 2014 Apr;67(4):425-432. [doi: 10.1016/j.jclinepi.2013.10.023] [Medline: 24581296]
15. Higgins J, Thomas J. Cochrane Handbook for Systematic Reviews of Interventions version 6.2. Cochrane Training. URL: https://training.cochrane.org/handbook/archive/v6.2 [accessed 2022-01-17]
16. Berkman N, Lohr K, Ansari M, Balk EM, Kane R, McDonagh M, et al. Grading the strength of a body of evidence when assessing health care interventions: an EPC update. J Clin Epidemiol 2015 Nov;68(11):1312-1324. [doi: 10.1016/j.jclinepi.2014.11.023] [Medline: 25721570]

17. Lundegard L. Efficacy of virtual behavioral health interventions for major depressive disorders. OSF Registries. 2020 Oct 01. URL: https://osf.io/8a976 [accessed 2022-01-17]
18. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
19. Human Development Report 2020: The Next Frontier: Human Development and the Anthropocene. United Nations Development Programme. URL: https://hdr.undp.org/content/human-development-report-2020 [accessed 2022-01-2]
20. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928-d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
21. Munder T, Barth J. Cochrane's risk of bias tool in the context of psychotherapy outcome research. Psychother Res 2018 May 09;28(3):347-355. [doi: 10.1080/10503307.2017.1411628] [Medline: 29224503]
22. Schwarzer G, Carpenter JR, Rücker G. Meta-analysis with R. New York, NY: Springer International Publishing; 2015.
23. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011 Dec;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
24. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: a randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
25. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki Å, Vernmark K, et al. Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. J Affect Disord 2013 Jun;148(2-3):331-337. [doi: 10.1016/j.jad.2012.12.020] [Medline: 23357657]
26. Johansson O, Bjärehed J, Andersson G, Carlbring P, Lundh L. Effectiveness of guided internet-delivered cognitive behavior therapy for depression in routine psychiatry: A randomized controlled trial. Internet Interv 2019 Sep;17:100247 [FREE Full text] [doi: 10.1016/j.invent.2019.100247] [Medline: 31249791]
27. Kenter RMF, Cuijpers P, Beekman A, van Straten A. Effectiveness of a web-based guided self-help intervention for outpatients with a depressive disorder: short-term results from a randomized controlled trial. J Med Internet Res 2016 Mar 31;18(3):e80 [FREE Full text] [doi: 10.2196/jmir.4861] [Medline: 27032449]
28. Smith J, Newby JM, Burston N, Murphy MJ, Michael S, Mackenzie A, et al. Help from home for depression: a randomised controlled trial comparing internet-delivered cognitive behaviour therapy with bibliotherapy for depression. Internet Interv 2017 Sep;9:25-37 [FREE Full text] [doi: 10.1016/j.invent.2017.05.001] [Medline: 30135834]
29. Forsell E, Bendix M, Holländare F, Szymanska von Schultz B, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: a randomised controlled trial. J Affect Disord 2017 Oct 15;221:56-64 [FREE Full text] [doi: 10.1016/j.jad.2017.06.013] [Medline: 28628768]
30. Hallgren M, Helgadóttir B, Herring MP, Zeebari Z, Lindefors N, Kaldo V, et al. Exercise and internet-based cognitive-behavioural therapy for depression: multicentre randomised controlled trial with 12-month follow-up. Br J Psychiatry 2016 Nov 02;209(5):414-420 [FREE Full text] [doi: 10.1192/bjp.bp.115.177576] [Medline: 27609813]
31. Löbner M, Pabst A, Stein J, Dorow M, Matschinger H, Luppa M, et al. Computerized cognitive behavior therapy for patients with mild to moderately severe depression in primary care: A pragmatic cluster randomized controlled trial (@ktiv). J Affect Disord 2018 Oct 01;238:317-326. [doi: 10.1016/j.jad.2018.06.008] [Medline: 29902736]
32. Moreno FA, Chong J, Dumbauld J, Humke M, Byreddy S. Use of standard webcam and internet equipment for telepsychiatry treatment of depression among underserved Hispanics. Psychiatr Serv 2012 Dec;63(12):1213-1217. [doi: 10.1176/appi.ps.201100274] [Medline: 23026854]
33. Pfeiffer PN, Pope B, Houck M, Benn-Burton W, Zivin K, Ganoczy D, et al. Effectiveness of peer-supported computer-based CBT for depression among veterans in primary care. Psychiatr Serv 2020 Mar 01;71(3):256-262. [doi: 10.1176/appi.ps.201900283] [Medline: 31931686]
34. Raevuori A, Vahlberg T, Korhonen T, Hilgert O, Aittakumpu-Hyden R, Forman-Hoffman V. A therapist-guided smartphone app for major depression in young adults: a randomized clinical trial. J Affect Disord 2021 May 01;286:228-238 [FREE Full text] [doi: 10.1016/j.jad.2021.02.007] [Medline: 33743385]
35. Wozney L, Olthuis J, Lingley-Pottie P, McGrath PJ, Chaplin W, Elgar F, et al. Strongest Families™ Managing Our Mood (MOM): a randomized controlled trial of a distance intervention for women with postpartum depression. Arch Womens Ment Health 2017 Aug 7;20(4):525-537. [doi: 10.1007/s00737-017-0732-y] [Medline: 28593360]
36. Flygare A, Engström I, Hasselgren M, Jansson-Fröjmark M, Frejgrim R, Andersson G, et al. Internet-based CBT for patients with depressive disorders in primary and psychiatric care: is it effective and does comorbidity affect outcome? Internet Interv 2020 Mar;19:100303 [FREE Full text] [doi: 10.1016/j.invent.2019.100303] [Medline: 32055451]
37. Johansson R, Ekbladh S, Hebert A, Lindström M, Möller S, Petitt E, et al. Psychodynamic guided self-help for adult depression through the internet: a randomised controlled trial. PLoS One 2012 May 29;7(5):e38021 [FREE Full text] [doi: 10.1371/journal.pone.0038021] [Medline: 22741027]

38. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 09;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
39. Oehler C, Görges F, Rogalla M, Rummel-Kluge C, Hegerl U. Efficacy of a guided web-based self-management intervention for depression or dysthymia: randomized controlled trial with a 12-month follow-up using an active control condition. J Med Internet Res 2020 Jul 14;22(7):e15361 [FREE Full text] [doi: 10.2196/15361] [Medline: 32673233]
40. Reins JA, Boß L, Lehr D, Berking M, Ebert DD. The more I got, the less I need? Efficacy of Internet-based guided self-help compared to online psychoeducation for major depressive disorder. J Affect Disord 2019 Mar 01;246:695-705. [doi: 10.1016/j.jad.2018.12.065] [Medline: 30611913]
41. Andersson G, Hesser H, Veilord A, Svedling L, Andersson F, Sleman O, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord 2013 Dec;151(3):986-994. [doi: 10.1016/j.jad.2013.08.022] [Medline: 24035673]
42. Egede LE, Acierno R, Knapp RG, Lejuez C, Hernandez-Tejada M, Payne EH, et al. Psychotherapy for depression in older veterans via telemedicine: a randomised, open-label, non-inferiority trial. Lancet Psychiatry 2015 Aug;2(8):693-701. [doi: 10.1016/S2215-0366(15)00122-4] [Medline: 26249300]
43. Mohr DC, Ho J, Duffecy J, Reifler D, Sokol L, Burns MN, et al. Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients: a randomized trial. JAMA 2012 Jun 06;307(21):2278-2285 [FREE Full text] [doi: 10.1001/jama.2012.5588] [Medline: 22706833]
44. Etzelmueller A, Vis C, Karyotaki E, Baumeister H, Titov N, Berking M, et al. Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. J Med Internet Res 2020 Aug 31;22(8):e18100 [FREE Full text] [doi: 10.2196/18100] [Medline: 32865497]
45. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
46. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]

Abbreviations

AHRQ: Agency for Healthcare and Research Quality
BA: behavioral activation
CBT: cognitive behavioral therapy
KQ: key question
MDD: major depressive disorder
MeSH: Medical Subject Heading
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference
SOE: strength of evidence
TAU: treatment as usual

Edited by J Torous, G Eysenbach; submitted 25.04.22; peer-reviewed by P von Wedel; comments to author 08.06.22; revised version received 03.08.22; accepted 18.08.22; published 09.01.23.

Please cite as:
Schiller CE, Prim J, Bauer AE, Lux L, Lundegard LC, Kang M, Hellberg S, Thompson K, Webber T, Teklezghi A, Pettee N, Gaffney K, Hodgins G, Rahman F, Steinsiek JN, Modi A, Gaynes BN
Efficacy of Virtual Care for Depressive Disorders: Systematic Review and Meta-analysis
JMIR Ment Health 2023;10:e38955
URL: https://mental.jmir.org/2023/1/e38955
doi:10.2196/38955
PMID:36622747

Review

Methodological and Quality Flaws in the Use of Artificial Intelligence in Mental Health Research: Systematic Review

Roberto Tornero-Costa[1*], MSc; Antonio Martinez-Millana[1*], PhD; Natasha Azzopardi-Muscat[2*], PhD; Ledia Lazeri[2*], MD; Vicente Traver[1*], PhD; David Novillo-Ortiz[2*], PhD

[1]Instituto Universitario de Investigación de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas, Universitat Politècnica de València, Valencia, Spain

[2]Division of Country Health Policies and Systems, World Health Organization, Regional Office for Europe, Copenhagen, Denmark

[*]all authors contributed equally

Corresponding Author:
David Novillo-Ortiz, PhD
Division of Country Health Policies and Systems
World Health Organization, Regional Office for Europe
Marmorej 51
Copenhagen, 2100
Denmark
Phone: 45 45 33 7198
Email: dnovillo@who.int

Abstract

Background: Artificial intelligence (AI) is giving rise to a revolution in medicine and health care. Mental health conditions are highly prevalent in many countries, and the COVID-19 pandemic has increased the risk of further erosion of the mental well-being in the population. Therefore, it is relevant to assess the current status of the application of AI toward mental health research to inform about trends, gaps, opportunities, and challenges.

Objective: This study aims to perform a systematic overview of AI applications in mental health in terms of methodologies, data, outcomes, performance, and quality.

Methods: A systematic search in PubMed, Scopus, IEEE Xplore, and Cochrane databases was conducted to collect records of use cases of AI for mental health disorder studies from January 2016 to November 2021. Records were screened for eligibility if they were a practical implementation of AI in clinical trials involving mental health conditions. Records of AI study cases were evaluated and categorized by the International Classification of Diseases 11th Revision (ICD-11). Data related to trial settings, collection methodology, features, outcomes, and model development and evaluation were extracted following the CHARMS (Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies) guideline. Further, evaluation of risk of bias is provided.

Results: A total of 429 nonduplicated records were retrieved from the databases and 129 were included for a full assessment—18 of which were manually added. The distribution of AI applications in mental health was found unbalanced between ICD-11 mental health categories. Predominant categories were Depressive disorders (n=70) and Schizophrenia or other primary psychotic disorders (n=26). Most interventions were based on randomized controlled trials (n=62), followed by prospective cohorts (n=24) among observational studies. AI was typically applied to evaluate quality of treatments (n=44) or stratify patients into subgroups and clusters (n=31). Models usually applied a combination of questionnaires and scales to assess symptom severity using electronic health records (n=49) as well as medical images (n=33). Quality assessment revealed important flaws in the process of AI application and data preprocessing pipelines. One-third of the studies (n=56) did not report any preprocessing or data preparation. One-fifth of the models were developed by comparing several methods (n=35) without assessing their suitability in advance and a small proportion reported external validation (n=21). Only 1 paper reported a second assessment of a previous AI model. Risk of bias and transparent reporting yielded low scores due to a poor reporting of the strategy for adjusting hyperparameters, coefficients, and the explainability of the models. International collaboration was anecdotal (n=17) and data and developed models mostly remained private (n=126).

Conclusions: These significant shortcomings, alongside the lack of information to ensure reproducibility and transparency, are indicative of the challenges that AI in mental health needs to face before contributing to a solid base for knowledge generation and for being a support tool in mental health management.

(JMIR Ment Health 2023;10:e42045) doi:10.2196/42045

KEYWORDS

artificial intelligence; mental health; health research; review methodology; systematic review; research methodology; research quality; trial methodology

Introduction

Mental health represents a vital element of individual and collective well-being, but stressful or adverse living, working, or economic conditions and social inequalities, violence, and conflict can put it at risk. The COVID-19 pandemic has demonstrated how vulnerable mental health can be. Mental health conditions represent one of the leading causes of suffering and disability in the European Region. In 2021, over 150 million people in the WHO (World Health Organization) European Region lived with a mental health condition, and only 1 in 3 people with depression receive the care they need. To address these gaps in mental health services and support, many of which have been exacerbated by the pandemic, WHO/Europe launched a new Pan-European Mental Health Coalition [1]. Mental health is a top priority for the WHO and is a flagship initiative of the European Programme of Work 2020-2025 [2].

Artificial intelligence (AI) has been increasingly used to provide methods and tools for improved diagnosis and treatment of diseases since 2010. AI is defined as the reproducibility of human-like reasoning and pattern extraction to solve problems [3]. AI involves a variety of methods that expand traditional statistical techniques. AI can find patterns that support decision making and hypotheses validation. AI offers a new scope of powerful tools to automate tasks, support clinicians, and deepen understanding of the causes of complex disorders. AI's presence and potential in health care are rapidly increasing in recent years. AI models need to be fed with the adequate data to be integrated in the clinical workflow and ensuring data quality is crucial [4]. Digitized data in health care are available in a range of formats, including structured data such as electronic health records or medical images, and nonstructured schemas, such as clinical handwritten notes [5].

Because of the possibilities AI offers, policymakers may gain insight into more efficient strategies to promote health and into the current state of mental disorders. However, AI often involves a complex use of statistics, mathematical approaches, and high-dimensional data that could lead to bias, inaccurate interpretation of results, and overoptimism of AI performance if it is not adequately handled [6]. Further, several lacking areas cause concern: transparent reporting in AI models that undermine replicability, potential ethical concerns, validation of generalizability, and positive collaboration in the research community [7,8].

The goals of this review are to map the applications of AI techniques in mental health research, reveal the prominent mental health aspects in this framework, and to assess the methodological quality of the recent scientific literature and evolution of this field in the last 5 years. Systematic reviews and meta-analyses (PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] 2020 statement) [9] will be used to design the search strategy and to funnel selection in this systematic overview.

Methods

Search Strategy

A systematic literature search was conducted on clinical trials on mental health disorders involving AI techniques using 4 electronic databases: PubMed, Scopus, IEEE Xplore, and Cochrane (Table 1). Search string queries are detailed in Appendix S1 in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

We specified 3 inclusion criteria for screening. Records were included if they reported a clinical trial (either interventional or observational), were related to mental health disorders, and featured an application of AI. For the final eligibility assessment, exclusion criteria were defined to constrain the review: the reported AI case is not applied for a mental health outcome (ie, applying tools to improve image quality), the record was not published in English, or the report was not published in the last 5 years to review the specific application of these techniques in clinical mental health research. These criteria were designed to evaluate the researching lines in mental health disorders in the last few years, which include the democratization of frameworks and tools for AI application.

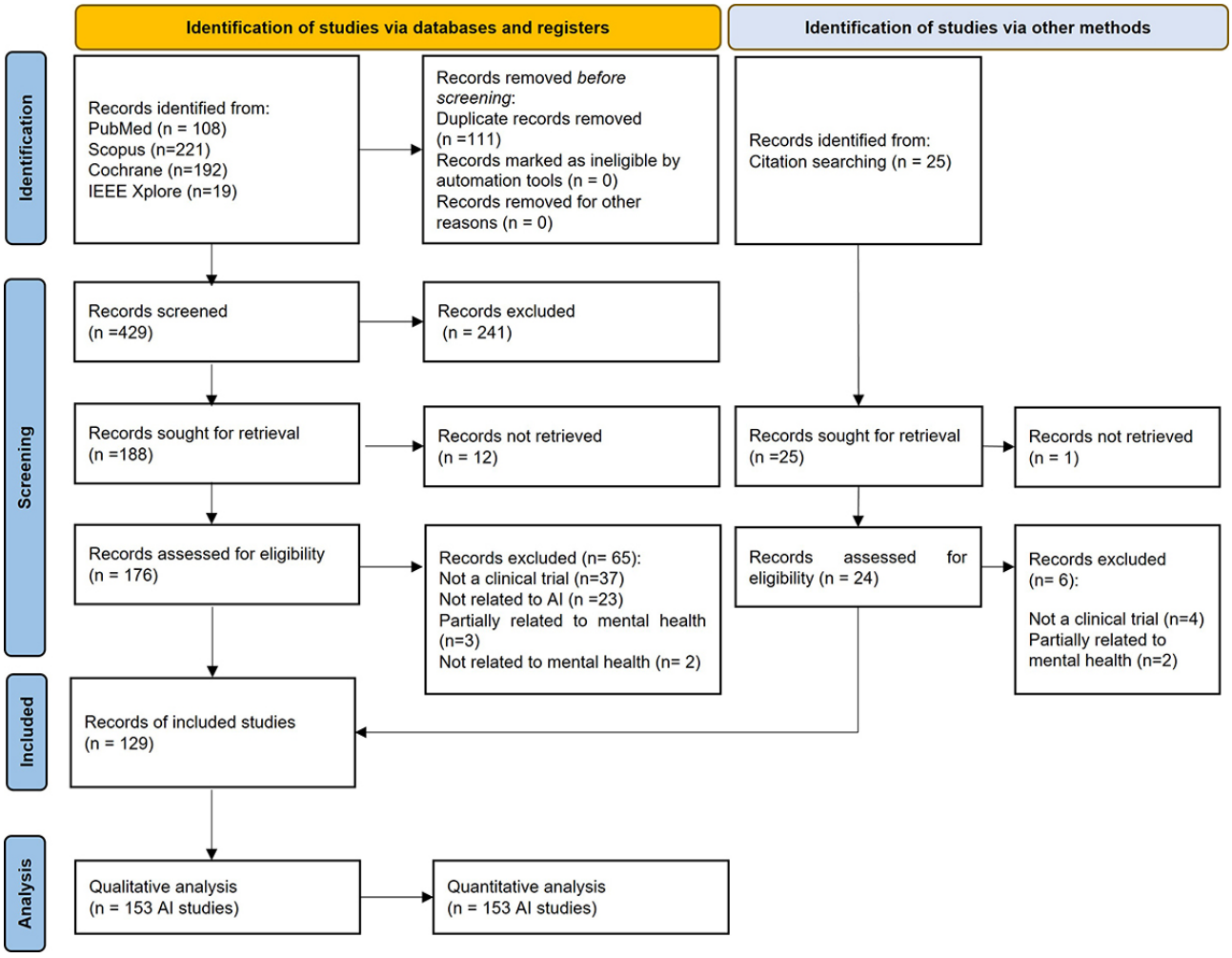
Table 1. Databases consulted and filters related to our search criteria applied in the search engines.

Database	Filters in search engine
PubMed	• Article type: Clinical trial • Language: English • Publication date: 5 years
Scopus	• Document type: Article • Language: English
IEEE Xplore	• Range years: 2016-2021
Cochrane	• Type: Trials • Publication date: 2016-2021 • Language: English

Selection Process

Figure 1 shows the flow diagram of the selection process. Records from the scientific literature were identified in the 4 databases defined in Table 1. The resulting data sets were combined in a Microsoft Excel spreadsheet (.xlsx), rearranged by DOI (digital object identifier), and checked for possible erroneous entries. Duplicated records were assessed by comparing DOI names and titles of the publication. A simple code in R 4.2.0 win32 (R Foundation for Statistical Computing) was used to find and tag records whose DOI name and title were already found in the database. The results were manually reviewed to correct minor errors due to misspellings of DOI or the title in the record database. The eligibility criteria for inclusion were then manually evaluated by the title and abstract of each record, and selected records were sought for retrieval. Retrieved records were fully screened and were dismissed if they did not meet the inclusion criteria or met the exclusion criteria. Finally, data and details were extracted for included AI studies.

Figure 1. Selection process: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram. AI: artificial intelligence.



Data Extraction

For the included AI applications, 11 categories and 35 data indicators are reported. These indicators were adapted from the CHARMS (Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies) guideline [10]. In addition, publication-related data such as author(s), title, journal, year of publication, and DOI were extracted for the analysis.

Categories were designed to evaluate the goal of the clinical trial and AI, the accessibility and quality of the development process of the data, how the AI model was designed and developed, results, and the reported discussion. Not only are the categories for data extraction designed for detailing the AI models, but also they evaluate the differences and linkages between trial design, data collection, and AI implementation. For further details, see Appendix S2 in Multimedia Appendix 1.

Quality Assessment

Risk of bias was evaluated by combining the Cochrane tool for randomized controlled trials [11] and PROBAST (Prediction Model Risk of Bias Assessment Tool) guidelines [12]. The Cochrane Handbook for Systematic Reviews of Interventions [13] accounted for the trial design and whether masking and blindness were applied or should have been. The PROBAST guidelines accounted for the suitability of the methodology for collecting the data, candidate predictors, and outcome definition for the AI model as well as how the AI model was applied and analyzed. Both guidelines were considered together to evaluate possible biased relations between trial design and AI applicability. Details of the methodology are provided in Appendix S3 in Multimedia Appendix 1.

Results

Articles Identified From the Database Searching

The search identified 540 records, all published in English. Excluding 111 duplicates, 429 articles were screened according to the eligibility criteria based on the title and abstract. The screening process concluded with 241 records excluded for not meeting the inclusion criteria. Of those, 188 records were sought for retrieval, with 12 found impossible to retrieve. Thus, 176 records were assessed for eligibility and 65 were excluded due to not being a clinical trial (n=37), not related to AI (n=23), partially related to mental health (n=3), or not related to mental health (n=2). Furthermore, limitations of search queries were minimized as much as possible by manually adding a selection of AI studies that were not found in the search (n=25). Records from this selection were also screened and sought for retrieval and eventually 18 studies were included. Ultimately, 129 records were included in the analysis. A record could involve 2 or more different cases of AI use for a different outcome, from now on referred to as an AI study. A total of 153 AI studies or AI applications were analyzed. Table 2 summarizes the most important information extracted for this systematic review. Details on the final analysis for each study can be found in Multimedia Appendix 2 (see also [14-142]).

Most used private data (n=142), and a small fraction used public data (n=10) or a mix (n=1). Most studies aimed to develop a new model (n=152); only 1 study aimed to validate a current model. No AI study was intended to update a previously developed AI model. Concerning mental health categories based on the International Classification of Diseases 11th Revision (ICD-11), nearly one-half of the studies (77/153, 50.3%) related to mood disorders, which combines the Depressive disorders (n=70) and Bipolar or related disorders (n=4) categories; 3 other studies used data from patients within both categories, labeled as “mood episodes.” The second most common category was Schizophrenia or other primary psychotic disorders (n=26), and the third was Disorders specifically associated with stress (n=12). Some studies included participants with different mental disorders (n=10).

Other categories were Anxiety or fear-related disorders (n=7); Secondary mental or behavioral syndromes associated with disorders or diseases classified elsewhere (n=5); Disorders due to substance use (n=5); Neurocognitive and dementia disorders (n=4); Neurodevelopmental disorders (n=1); Obsessive-compulsive or related disorders (n=1); Feeding or eating disorders (n=1); Bodily distress disorder (n=1); Personality disorders (n=1); and Mental or behavioral disorders associated with pregnancy, childbirth, or the puerperium, without psychotic symptoms (n=1).

Only 28.1% (43/153) of studies used original data collected within the study, while 71.9% (110/153) of studies retrieved data from databases or were a secondary analysis of clinical trials not designed for that purpose. The most common type of trial design was randomized clinical trial (n=62), followed by prospective cohort study designs (n=30) and nonrandomized clinical trial designs (n=15). Further, we found longitudinal naturalistic studies (n=15), cross-sectional designs (n=14), case-control designs (n=9), and case reports (n=2). Two reports of AI cases used a mix of trial designs and 4 did not report this or the references were unclear. Figure 2 shows the distribution of study design based on the prospective or retrospective collection of data.

Not all studies reported enough details to evaluate recruitment of participants (n=17). Almost one-half of the studies collected data from different locations (n=75), whereas the rest only reported 1 location (n=61). Of the multisite studies, only one-third used international collection (17/153, 11.1%). Only 13 of the 43 (30%) prospective collection studies followed a multisite collection method (n=13), and only 1 study was international.

Table 2. Key summary of artificial intelligence studies (N=153) analyzed (n=129 articles).

Mental health disorder (ICD-11[a]) section number: category	Artificial intelligence model family	Data type
6A0: Neurodevelopmental disorders	• Regression [14]	• Mixed[b] [14]
6A2: Schizophrenia or other primary psychotic disorders	• Competing models[c] [15-17] • Ensembled models [18,19] • Regression [20-23] • Statistical clustering [24] • SVM[d] [25-28] • Trees [29-31] • Regression and statistical clustering [32] • Regression and hierarchical clustering [33]	• Genomic data [29] • Medical image [15-19,21-23,27,28,33] • Mixed [20,26,30,32] • Questionnaires and scales [24,25,31]
6A6: Bipolar or related disorders	• Bayesian [34] • Manifold [35] • Regression [36] • SVM [37]	• Medical image [35,37] • Mixed [36] • Questionnaires and scales [34]
6A7: Depressive disorders	• Bayesian [38-40] • Competing models [41-53] • Deep learning [54] • Ensembled models [55-57] • Hierarchical clustering [58,59] • Markov model [60] • Mixture model [61] • Mixture model, regression, and trees [62] • Regression [63-74] • Relevance vector machine [75] • Statistical learning [76,77] • SVM [78-86] • Trees [87-93] • Trees and hierarchical clustering [94] • Trees and statistical learning [95]	• Audio recording [64] • Biomarkers [51,54,95] • Biosignal [41] • I[e] [80,81,87,93] • Genomic data [44,70,74] • Medical image [38,40,48,67-69,75,82,85] • Mixed [42,43,45,46,49,50,52,53,55-57,59,61,63,65,66,71-73,77,79,83,86,88-92,94] • Questionnaires and scales [47,58,60,62,76] • Text [39] • Video image [78,84]
6A6, 6A7: Mood episodes	• Competing models [96] • SVM [97] • Regression [98]	• Medical image [97,98] • Questionnaires and scales [96]
6B0: Anxiety or fear-related disorders	• Competing models [99-101] • Regression [102] • SVM [103] • Trees [104]	• Biomarkers [101] • Biosignal [100] • Mixed [99,103,104] • Text [102]
6B2: Obsessive-compulsive or related disorders	• Competing models [105]	• I [105]
6B4: Disorders specifically associated with stress	• Competing models [106] • Ensembled models [107,108] • Hierarchical clustering [109] • Mixture model and regression [110] • SVM [111-115] • Trees [116]	• Audio recording [116] • Biosignal [115] • Medical image [112-114] • Mixed [106-108,110,111] • Questionnaires and scales [109]
6B8: Feeding or eating disorders	• Competing models [117]	• Mixed [117]
6C2: Bodily distress disorder	• Regression [118]	• Mixed [118]
6C4: Disorders due to substance use	• Competing models [119] • Regression [120] • Trees [121,122]	• Medical image [120] • Mixed [119,121,122]

Mental health disorder (ICD-11[a]) section number: category	Artificial intelligence model family	Data type
6D1: Personality disorders	• Trees [123]	• Mixed [123]
6D7, 6D8: Neurocognitive disorders and dementia	• Competing models [124] • Ensembled models [125,126] • Trees [127]	• Mixed [124-127]
6E2: Mental or behavioral disorders associated with pregnancy, childbirth, or the puerperium, without psychotic symptoms	• Competing models [128]	• Mixed [128]
6E6: Secondary mental or behavioral syndromes associated with disorders or diseases classified elsewhere	• Bayesian [129] • Competing models [130] • Regression [131] • Trees [132,133]	• Mixed [130-133] • Questionnaires and scales [129]
Combination of some ICD-11 categories in mental health	• Bayesian [134] • Ensembled models [135-137] • Regression [138,139] • Regression and support vector machines [140] • Trees [141]	• Biosignal [140] • Mixed [135-137,139] • Questionnaires and scales [134,138,141]
Unspecified[f]	• Competing models [142]	• Mixed [142]

[a]ICD-11: International Classification of Diseases 11th Revision.

[b]Mixed: combination of type of data and predictors.

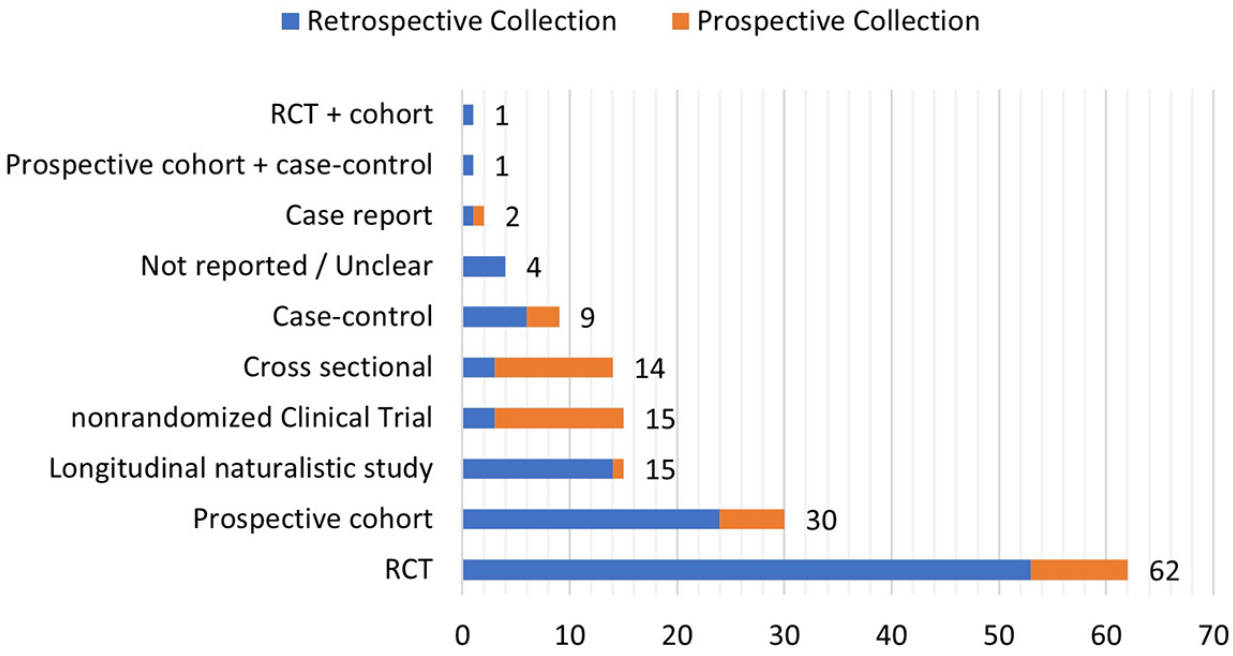
[c]Competing models: the study was designed for evaluate several types of artificial intelligence model families without assessing a priori adequacy.

[d]SVM: support vector machine.

[e]I: electronic health records.

[f]Unspecified: The outcome of the study is "mental health problems," therefore, it could not be classified in any specific category.

Figure 2. Count of trial designs where data were retrieved. Orange specifies only studies with their own designed trial. RCT: randomized clinical trial.



AI Applications

Studies were categorized according to the intended use of the AI models in the research. The most common category was studies for evaluating treatments, Treatment quality (n=44), followed by Subgroups/patterns identification (n=31), Predictor identification (n=28), Prognosis (n=23), Diagnose (n=20), and Forecasting symptoms (n=7).

Most Treatment quality applications used retrospective data from previous randomized clinical trials (n=28) and in these studies the clinical arms were treated as different cohorts to compare AI outcomes and performance (n=28). The same results can be found in Predictor identification, where close to one-half of the studies collected data from previous randomized clinical trials to compare different clinical arms (n=13). In the Subgroups/patterns identification category, studies collected data from a balanced mix of study designs, while in Prognosis the most common method was prospective cohort studies. In the Diagnose and Forecasting symptoms studies, none of the categories stood out in particular. More detailed results are presented in Appendix S4 in Multimedia Appendix 1.

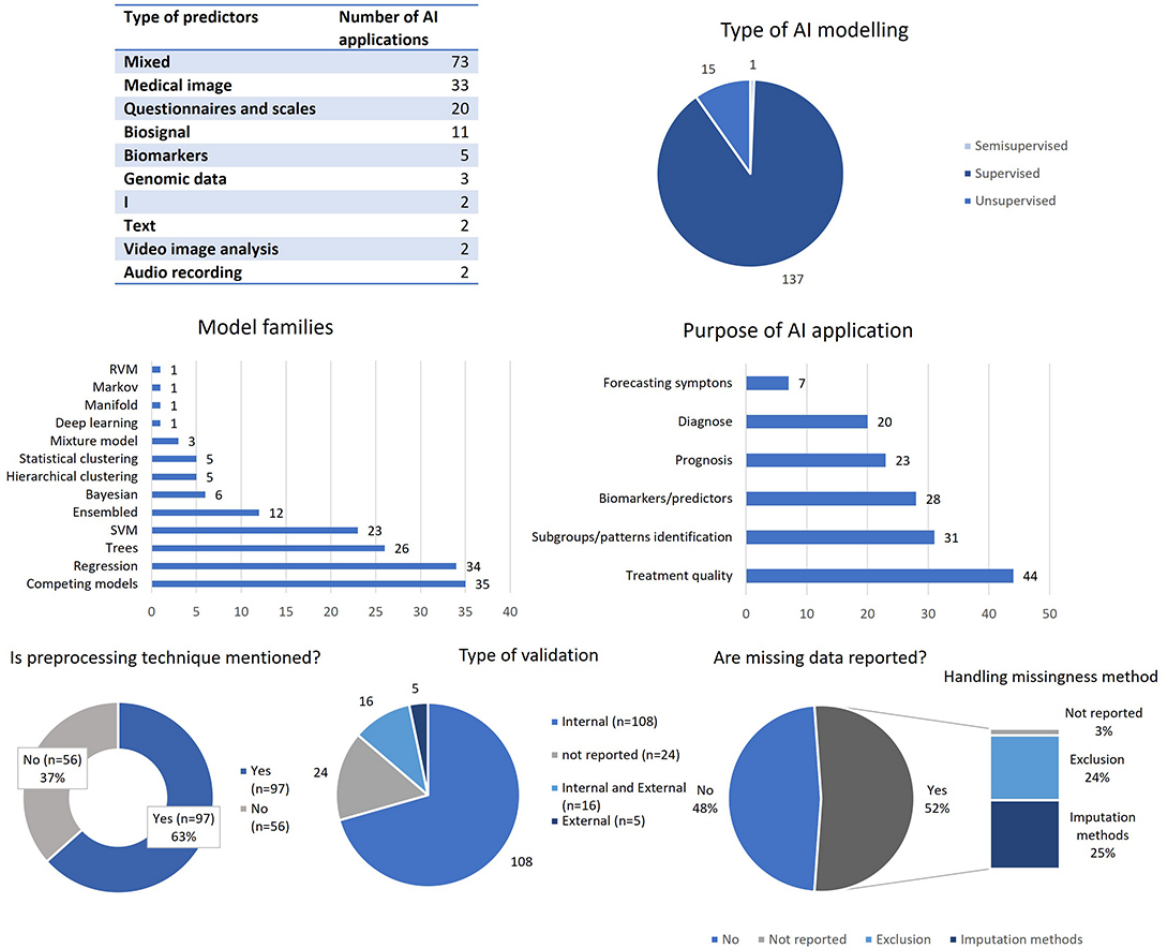
Figure 3 presents a dashboard that summarizes the AI model results regarding candidate predictors, preprocessing pipeline, AI techniques, and validation. For candidate predictors, many studies used a combination of data (n=73). The most individually common category was Medical image (n=33), which relates to medical imaging analysis (ie, region of interest or voxel-based morphometry), and the second was Questionnaires and scales (n=20), defined as any self-reported or interview-reported scale for symptom severity, conditions, or actual mood. The third was Biosignal collection (n=11), such as electroencephalography or electrocardiography and related analyses. Other data categories were Biomarkers (n=5), Genomic data (n=3), Electronic health records or I (n=2), Text (n=2), Video image analysis (n=2), and Audio recording (n=2). I refers to historical, demographic, and clinical information collected in hospitals and specialty care sites. Text refers to any data that are used for natural language processing analysis, such as written text or speech. Audio recording was introduced as the analysis of audio and voice features unrelated to language processing. The Mixed category (n=73 studies) combined data from I and different questionnaires and symptom scales (49/73, 67%); the remaining studies included other categories such as Biomarkers (n=7), Medical image (n=4), Genomic data (n=3), Biosignal (n=3), and Text (n=2). Medical image was also combined with Genomic data (n=1) and Biomarkers (n=1). Besides, Biomarkers were combined with Questionnaires and scales (n=2) and with Biosignals (n=1).

When evaluating data quality, methods to assess data suitability, and preprocessing pipelines, only 12/153 studies (7.8%) considered the statistical power of the sample size; 37.3% (57/153) of studies used a sample size of 150 or less to train the AI models. Only 13.7% (21/153) reported external validation (n=5) or reported both internal and external validation (n=16). The rest of the AI studies used only internal validation (n=108) or did not report the validation method (n=24). Only 38.6% (59/153) of studies reported a method to assess significance of their performance results, while the majority did not detail any (n=94).

AI studies used supervised learning, semisupervised learning, and unsupervised learning methods. No reinforcement learning algorithms were found. Regarding AI algorithms, the most popular family of techniques was regression (n=34), followed by trees (n=26) and support vector machine (n=23), which constitutes most AI studies. Other algorithms were Bayesian (n=6), statistical clustering (n=5), hierarchical clustering (n=5), mixture model (n=3), deep learning (n=1), manifold (n=1), Markov model (n=1), and relevance vector machine (n=1). In some cases, an Ensembled model was designed with the inclusion of different types of AI algorithms (n=12). Another category, Competing models (n=35), refers to the AI studies that did not predefine a specific AI technique or algorithm based on features of the data and instead applied different techniques with the intention to retain the model algorithm with the best performance to their outcome definition. These 35 studies used 144 AI techniques in total.

Regarding preprocessing methods, only 63.4% (97/153) of studies reported whether they applied any preprocessing technique to data or that preprocessing was not needed, while the rest did not report any (n=56). Regarding data gaps, only 52.3% (80/153) of studies reported or mentioned if there were some missing data in samples or not, while 47.7% (73/153) did not. Of the studies reporting missing data, 2.6% (4/153) did not report any method to handle missing data bias, whereas 24.2% (37/153) opted for excluding the samples and 25.5% (39/153) chose to impute the missing values from the data distribution by different imputation methods. Of these, only 2 studies detailed the type of missingness. The proportion of reporting missing data was similar for both retrospective and prospective data collection.

Figure 3. Dashboard and descriptive analytics on AI developing and preprocessing pipeline. AI: artificial intelligence; EHR: electronic health record; RVM: relevance vector machine; SVM: support vector machine.



Risk of Bias and Transparency

The risk of introducing bias is defined in Appendix S3 in Multimedia Appendix 1. Only 23 studies were found unlikely to introduce bias due to the trial design and evaluation of participants, whereas in the majority of studies the risk was high or unclear. In most cases, the risk of bias due to participants and the trial features was introduced by bias in the distribution of participants—that is, inclusion and exclusion criteria, loss of follow-up, and participants withdrawal—or the sample was not enough to be considered a good example of the target population. The definition and collection of the candidate predictors were mostly a low risk of introducing bias (n=16), with a few studies possibly introducing bias (n=21) or having an unclear risk of bias (n=16). Results for the outcome definition in the AI model are similar, with most studies evaluated as having a low risk of bias (n=101). Some studies were categorized as unclear (n=15) or high risk (n=37) due to unclear definitions of outcomes or combining data set of different populations whose outcomes were evaluated with different methods. The most important risk of bias was found when applying AI algorithms and their evaluation. Only a few studies were evaluated as unlikely to introduce some bias (n=5) and the vast majority of AI analysis introduced a high risk of bias (n=139), while 9 could not be assessed properly (n=9). The main issues for bias in the AI analysis were not appropriately preprocessing and arranging the data for the specifications of the applied AI model, a bad handling of missingness, or an insufficient validation of the performance to account for overfitting and optimism (Figure 4). Appendix S5 in Multimedia Appendix 1 shows a stratified analysis of the risk of bias based on disorders, study designs, and outcome.

Overall, only 1 AI study could be assessed as a low risk of being biased. The most contributing categories to the overall risk were Participants and AI analysis. Most studies were likely unbiased about the definition and collection of predictors and the outcome but they failed to apply these data later in the model—bad data engineering or bad validation of the models—or the trial design had some flaws from the beginning. It is worth mentioning that only 9 of the 153 models reported any hyperparameter tuning or coefficients of the models and most of them reported basic trees models coefficients and decision rules. Only 58 studies mentioned or reported predictor importance and less than one-half reported the ranking and evaluated the methodology to test it (Figure 5).

Figure 4. Analysis of the risk of bias following PROBAST (Prediction Model Risk of Bias Assessment Tool) categories as defined in Multimedia Appendix 1. AI: artificial intelligence.

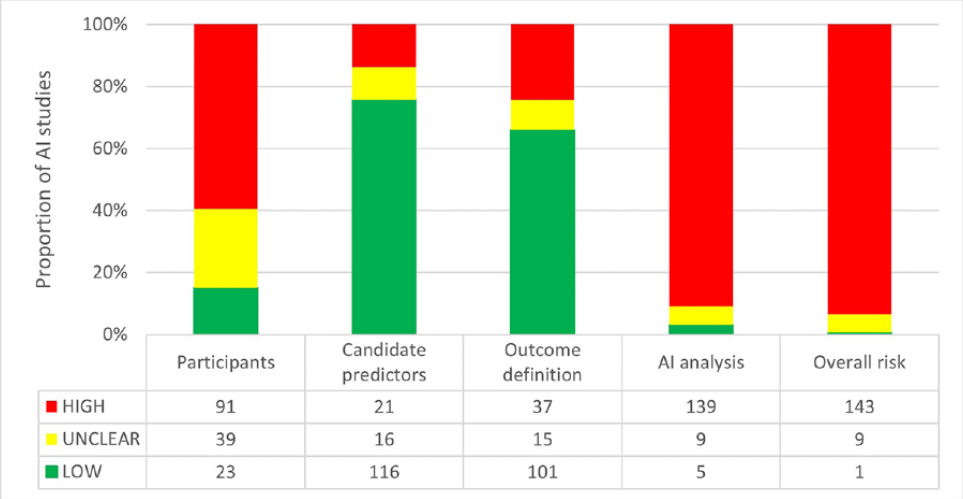
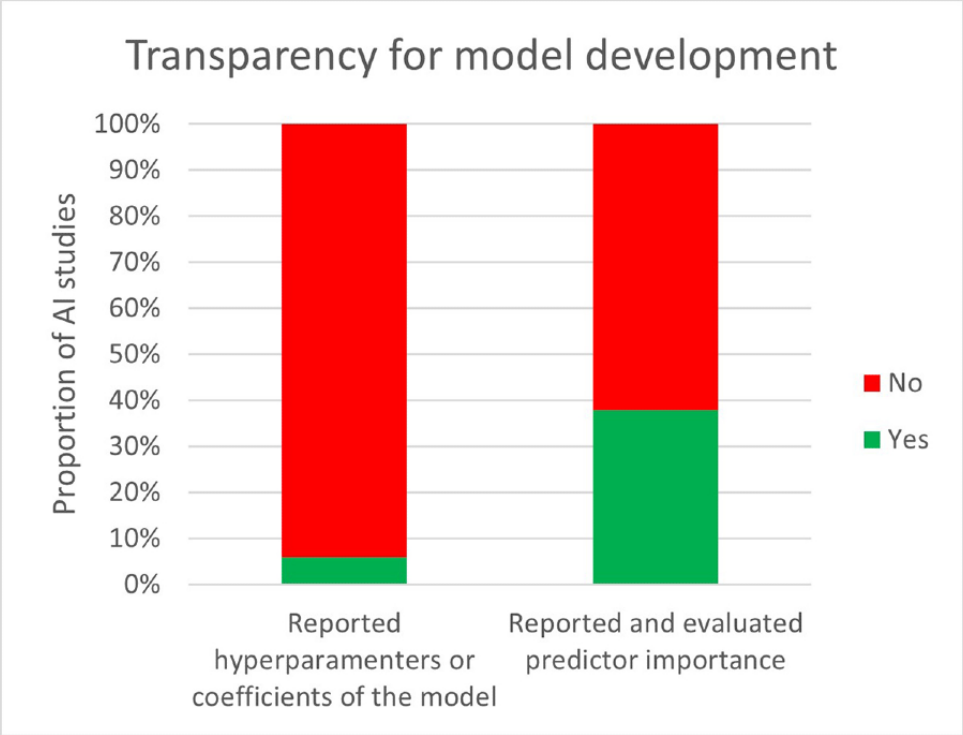


Figure 5. Reporting of candidate predictor importance as well as hyperparameters for model tuning and coefficients of models. AI: artificial intelligence.



Discussion

Principal Findings

This overview summarizes the development and status of AI applications in mental health care. The review focused on the period from 2016 to 2021 to understand the latest advancements of these applications in mental health research, including aspects related to the methodological quality, the risk of bias, and the transparency. Results may be limited by the keywords applied in the search queries. This analysis is only representative for the records retrieved in this search. However, the samples analyzed may be sufficient to judge the quality and current status of this field. Significant methodological flaws were found involving the statistical processes of AI applications and data preprocessing. Many studies overlook evaluating or reporting on data quality and how it fits the model. Some studies applied several AI techniques, here aggregated as "compelling models" studies, to select the most efficient technique without assessing their suitability for the problem they face, which may lead to overoptimism. Further, missing data management and the recruitment of participants are rarely reported, making it difficult to account for the risk of model overfitting. Preprocessing pipelines are not sufficiently reported, which hampers the reproducibility of the study or the adaptation of the AI techniques to the specific type of study. The use of reporting guidelines, such as the CONSORT-AI (AI version of the Consolidated Standards of Reporting Trials) for clinical trial reports involving AI [143], the SPIRIT-AI (AI version of the Standard Protocol Items: Recommendations for Interventional Trials) for clinical trial protocols involving AI [144], or the MI-CLAIM (Minimum Information About Clinical Artificial

Intelligence Modeling) checklist on minimum information for clinical AI modeling [145], would be very useful to ensure that the basic information about the study design and implementation is reported in these types of studies.

Some predictive models are being updated for AI reporting, such as TRIPOD-AI (AI version of the Transparent Reporting of a Multivariable Prediction Model of Individual Prognosis or Diagnosis) and PROBAST-AI (ie, the AI version of the PROBAST) [146]. However, these guidelines are rarely followed in the reviewed records. They lack transparent reporting on AI model features such as coefficients, hyperparameters, or predictor selection. Encouraging transparent reporting should be prioritized, as it would benefit second external validations and provide better accountability for reported models.

The incorporation of AI in mental health research is unbalanced between ICD-11 mental health disorders. Most research focusses on depressive disorders, where it combines severity questionnaires and scales with electronic health record and psychotic disorders using medical image data. External validation is very uncommon. Conducting suitable trial designs for the intended AI outcomes is understandably difficult in terms of money, time, and resources. Thus, it is common to apply data collected retrospectively. However, the original trial designs do not fit the specifications for AI development and most studies do not assess the appropriateness of these data. Notably, many authors may not understand the need to ensure an optimal preprocessing pipeline. In these cases, authors are aware of the poor performance of the models, but the proposed approach for improvement is suggested directly from a trial perspective rather than from assessing possible statistical bias or mistakes in model development, which could save cost and time over designing new studies.

Challenges

AI studies were analyzed to identify challenges and opportunities involving the use of AI in mental health. Typically, AI studies reported insufficient samples to ensure model generalizability [68,84,103]. Several authors reported bias because of the difficulty in adapting typical trial designs to an AI context. For example, some authors detail the constraining boundaries for selecting participants in randomized clinical trials as a limiting factor, which reduces the sample size and could overlook confounders [68,90]. Most randomized clinical trials noted possible variance between the collected data and the real-world data. However, observational studies can also introduce bias in AI models if the imbalance between cohorts is not adequately addressed [84,128]. In these studies, the variety in features such as prescribed medication could introduce confounders and bias that are difficult to manage [94]. Furthermore, in long-term studies, lack of follow-up or other conditions leading to a decrease of patients is an important limitation, mostly for prognosis studies or predictive evolution of condition severity [30,58]. These issues are worse for retrospective collection of data, where trial designs tend to diverge from the problematics of AI. Besides, some authors are aware of bias due to gaps, but most did not properly evaluate this risk.

A noticeable lack of internationalization was detected. Many studies focused only on local data, which contributes to small sample sizes and poor generalizability [115,127]. Encouraging partnerships and collaborations across countries and centers should be a priority, as it could facilitate external validation [71]. Some authors mention difficulties reconciling clinical practices with AI study requirements, usually due to ethics problems related to clinical practice in patients that can overlook confounders, that is, making it difficult to apply placebo controls in some interventions [82,115].

Another challenge is the explainability of complex AI models, which could make researchers reluctant to adopt techniques that map high-order interactions or “black-box” algorithms [81,122]. Researchers prefer simpler algorithms. The few studies that reported model coefficients and some explanation used decision trees. Another challenge is that contradictory findings could occur among studies [85].

Finally, some authors are aware of the opportunities that everyday devices and platforms such as phones and social networks offer but find it difficult to take advantage of these tools due to lack of standardization, which reduces the target population for defining a study [92].

Opportunities

Some studies introduced devices and platforms to improve the monitoring of patients. The application of everyday digital tools could reduce necessary resources and therefore facilitate data collection [99,127]. Promoting the use of frequently used devices combined with the application of AI models seems like a future trend that could improve the treatment of many conditions where the chance of treatment response decreases over time [126]. Further, it opens possibilities of internet-based treatments that could be conducted in real time with digital technology, easing the load on hospitals [99].

Data sharing and public databases should be encouraged to develop and implement more trustworthy AI models. AI models from clinical stage to clinical practice could be difficult but powerful tools to gain insights into predictor collection, human-based decisions, and AI biases while these techniques are being implemented in clinical world. Many studies report the high potential of AI in mental health for clinical support, computer-aided systems, and possibly preliminary screening [94,127].

Currently, many guidelines and initiatives exist to which researchers could adhere to in order to increase transparency and better use AI models. Currently, the EQUATOR (Enhancing the Quality and Transparency of Health Research) network initiative reports useful guidelines that could foster collaboration and implementation [147].

Conclusion

AI algorithms are increasingly being incorporated into mental health research; however, it is still uneven between ICD-11 categories. Collaboration is merely anecdotal, and data and developed models mostly remain private. Significant methodological flaws exist involving the statistical process of AI applications and data preprocessing pipelines. Only 1 study

was found reporting second validation, and 13.7% (21/153) reported external validation. The evaluation of the risk of bias and transparent reporting was discouraging. Model hyperparameters or trained coefficients are rarely reported, nor are insights about the explainability of the AI models. The lack of transparency and methodological flaws are concerning, as they delay the safe, practical implementation of AI. Furthermore, data engineering for AI models seems to be overlooked or misunderstood, and data are often not adequately managed. These significant shortcomings may indicate overly accelerated promotion of new AI models without pausing to assess their real-world viability.

Acknowledgments

This study was funded by the Division of Country Health Policies and Systems, Regional Office for Europe of the World Health Organization.

Disclaimer

DNO, NA-M and LL are staff members of the World Health Organization. The authors alone are responsible for the views expressed in this paper, and they do not necessarily represent the decisions, policies, or views of the World Health Organization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string queries; description of categories and indicators; evaluation of risk of bias; data collection for AI applications; and risk of bias stratified by disorders, study designs, and outcome. AI: artificial intelligence.
[DOCX File , 611 KB - mental_v10i1e42045_app1.docx]

Multimedia Appendix 2
Analysis results.
[XLSX File (Microsoft Excel File), 81 KB - mental_v10i1e42045_app2.xlsx]

References

1. The Pan-European Mental Health Coalition. WHO. URL: https://www.euro.who.int/en/health-topics/health-policy/european-programme-of-work/flagship-initiatives/the-pan-european-mental-health-coalition [accessed 2023-01-23]
2. WHO Regional Office for Europe. European Programme of Work, 2020–2025. WHO. 2020. URL: https://www.who.int/europe/about-us/our-work/european-programme-of-work [accessed 2023-01-23]
3. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med 2019 Jan;112(1):22-28 [FREE Full text] [doi: 10.1177/0141076818815510] [Medline: 30507284]
4. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
5. Borges do Nascimento IJ, Marcolino MS, Abdulazeem HM, Weerasekara I, Azzopardi-Muscat N, Gonçalves MA, et al. Impact of Big Data Analytics on People's Health: Overview of Systematic Reviews and Recommendations for Future Studies. J Med Internet Res 2021 Apr 13;23(4):e27275 [FREE Full text] [doi: 10.2196/27275] [Medline: 33847586]
6. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ 2021 Oct 20;375:n2281 [FREE Full text] [doi: 10.1136/bmj.n2281] [Medline: 34670780]
7. Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larrieux A. Towards Transparency by Design for Artificial Intelligence. Sci Eng Ethics 2020 Dec 16;26(6):3333-3361 [FREE Full text] [doi: 10.1007/s11948-020-00276-4] [Medline: 33196975]
8. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 2020 Mar 20;368:l6927 [FREE Full text] [doi: 10.1136/bmj.l6927] [Medline: 32198138]
9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Int J Surg 2021 Apr;88:105906. [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
10. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014 Oct 14;11(10):e1001744 [FREE Full text] [doi: 10.1371/journal.pmed.1001744] [Medline: 25314315]

11. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). London, UK: Cochrane; 2022.
12. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019 Jan 01;170(1):51. [doi: 10.7326/m18-1376]
13. Higgins J. Cochrane Handbook for Systematic Reviews of Interventions Version 5. The Cochrane Collaboration. London, UK: Cochrane Collaboration; 2011. URL: https://training.cochrane.org/handbook [accessed 2023-01-24]
14. Caye A, Agnew-Blais J, Arseneault L, Gonçalves H, Kieling C, Langley K, et al. A risk calculator to predict adult attention-deficit/hyperactivity disorder: generation and external validation in three birth cohorts and one clinical sample. Epidemiol Psychiatr Sci 2019 May 15;29:e37. [doi: 10.1017/s2045796019000283]
15. de Pierrefeu A, Löfstedt T, Laidi C, Hadj-Selem F, Bourgin J, Hajek T, et al. Identifying a neuroanatomical signature of schizophrenia, reproducible across sites and stages, using machine learning with structured sparsity. Acta Psychiatr Scand 2018 Dec 21;138(6):571-580. [doi: 10.1111/acps.12964] [Medline: 30242828]
16. Vieira S, Gong QY, Pinaya WHL, Scarpazza C, Tognin S, Crespo-Facorro B, et al. Using Machine Learning and Structural Neuroimaging to Detect First Episode Psychosis: Reconsidering the Evidence. Schizophr Bull 2020 Jan 04;46(1):17-26 [FREE Full text] [doi: 10.1093/schbul/sby189] [Medline: 30809667]
17. Bae Y, Kumarasamy K, Ali IM, Korfiatis P, Akkus Z, Erickson BJ. Differences Between Schizophrenic and Normal Subjects Using Network Properties from fMRI. J Digit Imaging 2018 Apr 18;31(2):252-261 [FREE Full text] [doi: 10.1007/s10278-017-0020-4] [Medline: 28924878]
18. Chand GB, Dwyer DB, Erus G, Sotiras A, Varol E, Srinivasan D, et al. Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. Brain 2020 Mar 01;143(3):1027-1038 [FREE Full text] [doi: 10.1093/brain/awaa025] [Medline: 32103250]
19. Koutsouleris N, Wobrock T, Guse B, Langguth B, Landgrebe M, Eichhammer P, et al. Predicting Response to Repetitive Transcranial Magnetic Stimulation in Patients With Schizophrenia Using Structural Magnetic Resonance Imaging: A Multisite Machine Learning Analysis. Schizophr Bull 2018 Aug 20;44(5):1021-1034 [FREE Full text] [doi: 10.1093/schbul/sbx114] [Medline: 28981875]
20. Anderson JP, Icten Z, Alas V, Benson C, Joshi K. Comparison and predictors of treatment adherence and remission among patients with schizophrenia treated with paliperidone palmitate or atypical oral antipsychotics in community behavioral health organizations. BMC Psychiatry 2017 Oct 18;17(1):346 [FREE Full text] [doi: 10.1186/s12888-017-1507-8] [Medline: 29047368]
21. Chung Y, Addington J, Bearden CE, Cadenhead K, Cornblatt B, Mathalon DH, North American Prodrome Longitudinal Study (NAPLS) Consortiumthe Pediatric Imaging, Neurocognition,Genetics (PING) Study Consortium. Use of Machine Learning to Determine Deviance in Neuroanatomical Maturity Associated With Future Psychosis in Youths at Clinically High Risk. JAMA Psychiatry 2018 Sep 01;75(9):960-968 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.1543] [Medline: 29971330]
22. Dillon K, Calhoun V, Wang Y. A robust sparse-modeling framework for estimating schizophrenia biomarkers from fMRI. J Neurosci Methods 2017 Jan 30;276:46-55 [FREE Full text] [doi: 10.1016/j.jneumeth.2016.11.005] [Medline: 27867012]
23. Schreiner M, Forsyth JK, Karlsgodt KH, Anderson AE, Hirsh N, Kushan L, et al. Intrinsic Connectivity Network-Based Classification and Detection of Psychotic Symptoms in Youth With 22q11.2 Deletions. Cereb Cortex 2017 Jun 01;27(6):3294-3306 [FREE Full text] [doi: 10.1093/cercor/bhx076] [Medline: 28383675]
24. Lefort-Besnard J, Varoquaux G, Derntl B, Gruber O, Aleman A, Jardri R, et al. Patterns of schizophrenia symptoms: hidden structure in the PANSS questionnaire. Transl Psychiatry 2018 Oct 30;8(1):237 [FREE Full text] [doi: 10.1038/s41398-018-0294-4] [Medline: 30375374]
25. Brodey B, Girgis R, Favorov O, Bearden C, Woods S, Addington J, et al. The Early Psychosis Screener for Internet (EPSI)-SR: Predicting 12 month psychotic conversion using machine learning. Schizophr Res 2019 Jun;208:390-396 [FREE Full text] [doi: 10.1016/j.schres.2019.01.015] [Medline: 30777603]
26. Koutsouleris N, Kahn RS, Chekroud AM, Leucht S, Falkai P, Wobrock T, et al. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. The Lancet Psychiatry 2016 Oct;3(10):935-946. [doi: 10.1016/s2215-0366(16)30171-7]
27. Nieuwenhuis M, Schnack HG, van Haren NE, Lappin J, Morgan C, Reinders AA, et al. Multi-center MRI prediction models: Predicting sex and illness course in first episode psychosis patients. Neuroimage 2017 Jan 15;145(Pt B):246-253 [FREE Full text] [doi: 10.1016/j.neuroimage.2016.07.027] [Medline: 27421184]
28. Rozycki M, Satterthwaite TD, Koutsouleris N, Erus G, Doshi J, Wolf DH, et al. Multisite Machine Learning Analysis Provides a Robust Structural Imaging Signature of Schizophrenia Detectable Across Diverse Patient Populations and Within Individuals. Schizophr Bull 2018 Aug 20;44(5):1035-1044 [FREE Full text] [doi: 10.1093/schbul/sbx137] [Medline: 29186619]
29. Chen J, Zang Z, Braun U, Schwarz K, Harneit A, Kremer T, et al. Association of a Reproducible Epigenetic Risk Profile for Schizophrenia With Brain Methylation and Function. JAMA Psychiatry 2020 Jun 01;77(6):628-636 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.4792] [Medline: 32049268]

30. Fond G, Bulzacka E, Boucekine M, Schürhoff F, Berna F, Godin O, FACE-SZ (FondaMental Academic Centers of Expertise for Schizophrenia) group, et al. Machine learning for predicting psychotic relapse at 2 years in schizophrenia in the national FACE-SZ cohort. Prog Neuropsychopharmacol Biol Psychiatry 2019 Jun 08;92:8-18. [doi: 10.1016/j.pnpbp.2018.12.005] [Medline: 30552914]
31. Gibbons RD, Chattopadhyay I, Meltzer HY, Kane JM, Guinart D. Development of a computerized adaptive diagnostic screening tool for psychosis. Schizophr Res 2022 Jul;245:116-121. [doi: 10.1016/j.schres.2021.03.020] [Medline: 33836922]
32. Martinuzzi E, Barbosa S, Daoudlarian D, Bel Haj Ali W, Gilet C, Fillatre L, OPTiMiSE Study Group. Stratification and prediction of remission in first-episode psychosis patients: the OPTiMiSE cohort study. Transl Psychiatry 2019 Jan 17;9(1):20 [FREE Full text] [doi: 10.1038/s41398-018-0366-5] [Medline: 30655509]
33. Skåtun KC, Kaufmann T, Doan NT, Alnæs D, Córdova-Palomera A, Jönsson EG, KaSP, et al. Consistent Functional Connectivity Alterations in Schizophrenia Spectrum Disorder: A Multisite Study. Schizophr Bull 2017 Jul 01;43(4):914-924 [FREE Full text] [doi: 10.1093/schbul/sbw145] [Medline: 27872268]
34. Busk J, Faurholt-Jepsen M, Frost M, Bardram JE, Vedel Kessing L, Winther O. Forecasting Mood in Bipolar Disorder From Smartphone Self-assessments: Hierarchical Bayesian Approach. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e15028 [FREE Full text] [doi: 10.2196/15028] [Medline: 32234702]
35. Sun ZY, Houenou J, Duclap D, Sarrazin S, Linke J, Daban C, et al. Shape analysis of the cingulum, uncinate and arcuate fasciculi in patients with bipolar disorder. J Psychiatry Neurosci 2017 Jan;42(1):27-36 [FREE Full text] [doi: 10.1503/jpn.150291] [Medline: 28234596]
36. Kim TT, Dufour S, Xu C, Cohen ZD, Sylvia L, Deckersbach T, et al. Predictive modeling for response to lithium and quetiapine in bipolar disorder. Bipolar Disord 2019 Aug;21(5):428-436. [doi: 10.1111/bdi.12752] [Medline: 30729637]
37. Lei D, Li W, Tallman MJ, Patino LR, McNamara RK, Strawn JR, et al. Changes in the brain structural connectome after a prospective randomized clinical trial of lithium and quetiapine treatment in youth with bipolar disorder. Neuropsychopharmacology 2021 Jun;46(7):1315-1323 [FREE Full text] [doi: 10.1038/s41386-021-00989-5] [Medline: 33753882]
38. Brown EC, Clark DL, Forkert ND, Molnar CP, Kiss ZHT, Ramasubbu R. Metabolic activity in subcallosal cingulate predicts response to deep brain stimulation for depression. Neuropsychopharmacology 2020 Sep;45(10):1681-1688 [FREE Full text] [doi: 10.1038/s41386-020-0745-5] [Medline: 32580207]
39. Carrillo F, Sigman M, Fernández Slezak D, Ashton P, Fitzgerald L, Stroud J, et al. Natural speech algorithm applied to baseline interview data can predict which patients will respond to psilocybin for treatment-resistant depression. J Affect Disord 2018 Apr 01;230:84-86. [doi: 10.1016/j.jad.2018.01.006] [Medline: 29407543]
40. Sikora M, Heffernan J, Avery ET, Mickey BJ, Zubieta J, Peciña M. Salience Network Functional Connectivity Predicts Placebo Effects in Major Depression. Biol Psychiatry Cogn Neurosci Neuroimaging 2016 Jan;1(1):68-76 [FREE Full text] [doi: 10.1016/j.bpsc.2015.10.002] [Medline: 26709390]
41. Al-Kaysi AM, Al-Ani A, Loo CK, Powell TY, Martin DM, Breakspear M, et al. Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification. J Affect Disord 2017 Jan 15;208:597-603. [doi: 10.1016/j.jad.2016.10.021] [Medline: 28029427]
42. Athreya A, Iyer R, Neavin D, Wang L, Weinshilboum R, Kaddurah-Daouk R, et al. Augmentation of Physician Assessments with Multi-Omics Enhances Predictability of Drug Response: A Case Study of Major Depressive Disorder. IEEE Comput Intell Mag 2018 Aug;13(3):20-31. [doi: 10.1109/mci.2018.2840660]
43. Bai R, Xiao L, Guo Y, Zhu X, Li N, Wang Y, et al. Tracking and Monitoring Mood Stability of Patients With Major Depressive Disorder by Machine Learning Models Using Passive Digital Data: Prospective Naturalistic Multicenter Study. JMIR Mhealth Uhealth 2021 Mar 08;9(3):e24365 [FREE Full text] [doi: 10.2196/24365] [Medline: 33683207]
44. Bao Z, Zhao X, Li J, Zhang G, Wu H, Ning Y, et al. Prediction of repeated-dose intravenous ketamine response in major depressive disorder using the GWAS-based machine learning approach. J Psychiatr Res 2021 Jun;138:284-290. [doi: 10.1016/j.jpsychires.2021.04.014] [Medline: 33878621]
45. Bartlett EA, DeLorenzo C, Sharma P, Yang J, Zhang M, Petkova E, et al. Pretreatment and early-treatment cortical thickness is associated with SSRI treatment response in major depressive disorder. Neuropsychopharmacology 2018 Oct 19;43(11):2221-2230 [FREE Full text] [doi: 10.1038/s41386-018-0122-9] [Medline: 29955151]
46. Bremer V, Becker D, Kolovos S, Funk B, van Breda W, Hoogendoorn M, et al. Predicting Therapy Success and Costs for Personalized Treatment Recommendations Using Baseline Characteristics: Data-Driven Analysis. J Med Internet Res 2018 Aug 21;20(8):e10275 [FREE Full text] [doi: 10.2196/10275] [Medline: 30131318]
47. Goerigk S, Hilbert S, Jobst A, Falkai P, Bühner M, Stachl C, et al. Predicting instructed simulation and dissimulation when screening for depressive symptoms. Eur Arch Psychiatry Clin Neurosci 2020 Mar 12;270(2):153-168. [doi: 10.1007/s00406-018-0967-2] [Medline: 30542818]
48. Hopman H, Chan S, Chu W, Lu H, Tse C, Chau S, et al. Personalized prediction of transcranial magnetic stimulation clinical response in patients with treatment-refractory depression using neuroimaging biomarkers and machine learning. J Affect Disord 2021 Jul 01;290:261-271. [doi: 10.1016/j.jad.2021.04.081] [Medline: 34010751]

49. Monaro M, Toncini A, Ferracuti S, Tessari G, Vaccaro MG, De Fazio P, et al. The Detection of Malingering: A New Tool to Identify Made-Up Depression. Front Psychiatry 2018 Jun 8;9:249 [FREE Full text] [doi: 10.3389/fpsyt.2018.00249] [Medline: 29937740]
50. Nie Z, Vairavan S, Narayan VA, Ye J, Li QS. Predictive modeling of treatment resistant depression using data from STAR*D and an independent clinical study. PLoS One 2018 Jun 7;13(6):e0197268 [FREE Full text] [doi: 10.1371/journal.pone.0197268] [Medline: 29879133]
51. Setoyama D, Kato TA, Hashimoto R, Kunugi H, Hattori K, Hayakawa K, et al. Plasma Metabolites Predict Severity of Depression and Suicidal Ideation in Psychiatric Patients-A Multicenter Pilot Analysis. PLoS One 2016 Dec 16;11(12):e0165267 [FREE Full text] [doi: 10.1371/journal.pone.0165267] [Medline: 27984586]
52. Shepherd-Banigan M, Smith VA, Lindquist JH, Cary MP, Miller KEM, Chapman JG, et al. Identifying treatment effects of an informal caregiver education intervention to increase days in the community and decrease caregiver distress: a machine-learning secondary analysis of subgroup effects in the HI-FIVES randomized clinical trial. Trials 2020 Feb 14;21(1):189 [FREE Full text] [doi: 10.1186/s13063-020-4113-x] [Medline: 32059687]
53. Zhao M, Feng Z. Machine Learning Methods to Evaluate the Depression Status of Chinese Recruits: A Diagnostic Study. NDT 2020 Nov;16:2743-2752. [doi: 10.2147/ndt.s275620]
54. Schultebraucks K, Choi KW, Galatzer-Levy IR, Bonanno GA. Discriminating Heterogeneous Trajectories of Resilience and Depression After Major Life Stressors Using Polygenic Scores. JAMA Psychiatry 2021 Jul 01;78(7):744-752 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.0228] [Medline: 33787853]
55. Castellani B, Griffiths F, Rajaram R, Gunn J. Exploring comorbid depression and physical health trajectories: A case-based computational modelling approach. J Eval Clin Pract 2018 Dec 02;24(6):1293-1309. [doi: 10.1111/jep.13042] [Medline: 30277297]
56. Jacobson NC, Nemesure MD. Using Artificial Intelligence to Predict Change in Depression and Anxiety Symptoms in a Digital Intervention: Evidence from a Transdiagnostic Randomized Controlled Trial. Psychiatry Res 2021 Jan;295:113618 [FREE Full text] [doi: 10.1016/j.psychres.2020.113618] [Medline: 33278743]
57. Kessler RC, Furukawa TA, Kato T, Luedtke A, Petukhova M, Sadikova E, et al. An individualized treatment rule to optimize probability of remission by continuation, switching, or combining antidepressant medications after failing a first-line antidepressant in a two-stage randomized trial. Psychol Med 2021 Mar 08;52(15):3371-3380. [doi: 10.1017/s0033291721000027]
58. Bondar J, Caye A, Chekroud AM, Kieling C. Symptom clusters in adolescent depression and differential response to treatment: a secondary analysis of the Treatment for Adolescents with Depression Study randomised trial. The Lancet Psychiatry 2020 Apr;7(4):337-343. [doi: 10.1016/s2215-0366(20)30060-2]
59. Kato M, Asami Y, Wajsbrot DB, Wang X, Boucher M, Prieto R, et al. Clustering patients by depression symptoms to predict venlafaxine ER antidepressant efficacy: Individual patient data analysis. J Psychiatr Res 2020 Oct;129:160-167 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.06.011] [Medline: 32912597]
60. Bueno ML, Hommersom A, Lucas PJ, Janzing J. A probabilistic framework for predicting disease dynamics: A case study of psychotic depression. J Biomed Inform 2019 Jul;95:103232 [FREE Full text] [doi: 10.1016/j.jbi.2019.103232] [Medline: 31201965]
61. Alexopoulos GS, Raue PJ, Banerjee S, Mauer E, Marino P, Soliman M, et al. Modifiable predictors of suicidal ideation during psychotherapy for late-life major depression. A machine learning approach. Transl Psychiatry 2021 Oct 18;11(1):536 [FREE Full text] [doi: 10.1038/s41398-021-01656-5] [Medline: 34663787]
62. Solomonov N, Lee J, Banerjee S, Flückiger C, Kanellopoulos D, Gunning FM, et al. Modifiable predictors of nonresponse to psychotherapies for late-life depression with executive dysfunction: a machine learning approach. Mol Psychiatry 2021 Sep 10;26(9):5190-5198 [FREE Full text] [doi: 10.1038/s41380-020-0836-z] [Medline: 32651477]
63. Bruijniks SJ, DeRubeis RJ, Lemmens LH, Peeters FP, Cuijpers P, Huibers MJ. The relation between therapy quality, therapy processes and outcomes and identifying for whom therapy quality matters in CBT and IPT for depression. Behav Res Ther 2021 Jan 28;139:103815. [doi: 10.1016/j.brat.2021.103815] [Medline: 33581481]
64. Di Y, Wang J, Li W, Zhu T. Using i-vectors from voice features to identify major depressive disorder. J Affect Disord 2021 Jun 01;288:161-166. [doi: 10.1016/j.jad.2021.04.004] [Medline: 33895418]
65. Gunlicks-Stoessel M, Klimes-Dougan B, VanZomeren A, Ma S. Developing a data-driven algorithm for guiding selection between cognitive behavioral therapy, fluoxetine, and combination treatment for adolescent depression. Transl Psychiatry 2020 Sep 21;10(1):321 [FREE Full text] [doi: 10.1038/s41398-020-01005-y] [Medline: 32958758]
66. Iniesta R, Malki K, Maier W, Rietschel M, Mors O, Hauser J, et al. Combining clinical variables to optimize prediction of antidepressant treatment outcomes. J Psychiatr Res 2016 Jul;78:94-102 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.03.016] [Medline: 27089522]
67. Liu Y, Admon R, Mellem MS, Belleau EL, Kaiser RH, Clegg R, et al. Machine Learning Identifies Large-Scale Reward-Related Activity Modulated by Dopaminergic Enhancement in Major Depression. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 2020 Feb;5(2):163-172. [doi: 10.1016/j.bpsc.2019.10.002]

68. Maglanoc LA, Kaufmann T, Jonassen R, Hilland E, Beck D, Landrø NI, et al. Multimodal fusion of structural and functional brain imaging in depression using linked independent component analysis. Hum Brain Mapp 2020 Jan;41(1):241-255 [FREE Full text] [doi: 10.1002/hbm.24802] [Medline: 31571370]
69. Mennen AC, Turk-Browne NB, Wallace G, Seok D, Jaganjac A, Stock J, et al. Cloud-Based Functional Magnetic Resonance Imaging Neurofeedback to Reduce the Negative Attentional Bias in Depression: A Proof-of-Concept Study. Biol Psychiatry Cogn Neurosci Neuroimaging 2021 Apr;6(4):490-497 [FREE Full text] [doi: 10.1016/j.bpsc.2020.10.006] [Medline: 33422469]
70. Salvetat N, Van der Laan S, Vire B, Chimienti F, Cleophax S, Bronowicki JP, et al. RNA editing blood biomarkers for predicting mood alterations in HCV patients. J Neurovirol 2019 Dec 22;25(6):825-836 [FREE Full text] [doi: 10.1007/s13365-019-00772-9] [Medline: 31332697]
71. van Bronswijk SC, DeRubeis RJ, Lemmens LHJM, Peeters FPML, Keefe JR, Cohen ZD, et al. Precision medicine for long-term depression outcomes using the Personalized Advantage Index approach: cognitive therapy or interpersonal psychotherapy? Psychol. Med 2019 Nov 22;51(2):279-289. [doi: 10.1017/s0033291719003192]
72. van Bronswijk SC, Bruijniks SJE, Lorenzo-Luaces L, Derubeis RJ, Lemmens LHJM, Peeters FPML, et al. Cross-trial prediction in psychotherapy: External validation of the Personalized Advantage Index using machine learning in two Dutch randomized trials comparing CBT versus IPT for depression. Psychother Res 2021 Jan 23;31(1):78-91. [doi: 10.1080/10503307.2020.1823029] [Medline: 32964809]
73. Webb CA, Trivedi MH, Cohen ZD, Dillon DG, Fournier JC, Goer F, et al. Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. Psychol Med 2018 Jul 2;49(07):1118-1127. [doi: 10.1017/s0033291718001708]
74. Wu W, Zhang Y, Jiang J, Lucas MV, Fonzo GA, Rolle CE, et al. An electroencephalographic signature predicts antidepressant response in major depression. Nat Biotechnol 2020 Apr 10;38(4):439-447 [FREE Full text] [doi: 10.1038/s41587-019-0397-3] [Medline: 32042166]
75. Fonzo GA, Etkin A, Zhang Y, Wu W, Cooper C, Chin-Fatt C, et al. Brain regulation of emotional conflict predicts antidepressant treatment response for depression. Nat Hum Behav 2019 Dec 23;3(12):1319-1331 [FREE Full text] [doi: 10.1038/s41562-019-0732-1] [Medline: 31548678]
76. Hartmann A, von Wietersheim J, Weiss H, Zeeck A. Patterns of symptom change in major depression: Classification and clustering of long term courses. Psychiatry Res 2018 Sep;267:480-489. [doi: 10.1016/j.psychres.2018.03.086] [Medline: 29980128]
77. Zilcha-Mano S, Brown PJ, Roose SP, Cappetta K, Rutherford BR. Optimizing patient expectancy in the pharmacologic treatment of major depressive disorder. Psychol Med 2018 Nov 13;49(14):2414-2420. [doi: 10.1017/s0033291718003343]
78. Kacem A, Hammal Z, Daoudi M, Cohn J. Detecting Depression Severity by Interpretable Representations of Motion Dynamics. Proc Int Conf Autom Face Gesture Recognit 2018 May;2018:739-745 [FREE Full text] [doi: 10.1109/FG.2018.00116] [Medline: 30271308]
79. Bailey N, Hoy K, Rogasch N, Thomson R, McQueen S, Elliot D, et al. Responders to rTMS for depression show increased fronto-midline theta and theta connectivity compared to non-responders. Brain Stimul 2018 Jan;11(1):190-203. [doi: 10.1016/j.brs.2017.10.015] [Medline: 29128490]
80. Browning M, Kingslake J, Dourish CT, Goodwin GM, Harmer CJ, Dawson GR. Predicting treatment response to antidepressant medication using early changes in emotional processing. Eur Neuropsychopharmacol 2019 Jan;29(1):66-75 [FREE Full text] [doi: 10.1016/j.euroneuro.2018.11.1102] [Medline: 30473402]
81. Byun S, Kim AY, Jang EH, Kim S, Choi KW, Yu HY, et al. Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol. Comput Biol Med 2019 Sep;112:103381 [FREE Full text] [doi: 10.1016/j.compbiomed.2019.103381] [Medline: 31404718]
82. Cash RFH, Cocchi L, Anderson R, Rogachov A, Kucyi A, Barnett AJ, et al. A multivariate neuroimaging biomarker of individual outcome to transcranial magnetic stimulation in depression. Hum Brain Mapp 2019 Nov 01;40(16):4618-4629 [FREE Full text] [doi: 10.1002/hbm.24725] [Medline: 31332903]
83. Furukawa TA, Debray TPA, Akechi T, Yamada M, Kato T, Seo M, et al. Can personalized treatment prediction improve the outcomes, compared with the group average approach, in a randomized trial? Developing and validating a multivariable prediction model in a pragmatic megatrial of acute treatment for major depression. Journal of Affective Disorders 2020 Sep;274:690-697. [doi: 10.1016/j.jad.2020.05.141]
84. Horigome T, Sumali B, Kitazawa M, Yoshimura M, Liang K, Tazawa Y, et al. Evaluating the severity of depressive symptoms using upper body motion captured by RGB-depth sensors and machine learning in a clinical interview setting: A preliminary study. Compr Psychiatry 2020 Feb 20;98:152169 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152169] [Medline: 32145559]
85. Redlich R, Opel N, Grotegerd D, Dohm K, Zaremba D, Bürger C, et al. Prediction of Individual Response to Electroconvulsive Therapy via Machine Learning on Structural Magnetic Resonance Imaging Data. JAMA Psychiatry 2016 Jun 01;73(6):557-564. [doi: 10.1001/jamapsychiatry.2016.0316] [Medline: 27145449]
86. Zhou X, Wang Y, Zeng D. Outcome-Weighted Learning for Personalized Medicine with Multiple Treatment Options. Proc Int Conf Data Sci Adv Anal 2018 Oct;2018:565-574 [FREE Full text] [doi: 10.1109/DSAA.2018.00072] [Medline: 30931437]

87. Barrett BW, Abraham AG, Dean LT, Plankey MW, Friedman MR, Jacobson LP, et al. Social inequalities contribute to racial/ethnic disparities in depressive symptomology among men who have sex with men. Soc Psychiatry Psychiatr Epidemiol 2021 Feb 11;56(2):259-272 [FREE Full text] [doi: 10.1007/s00127-020-01940-7] [Medline: 32780176]
88. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. The Lancet Psychiatry 2016 Mar;3(3):243-250. [doi: 10.1016/s2215-0366(15)00471-x]
89. Foster S, Mohler-Kuo M, Tay L, Hothorn T, Seibold H. Estimating patient-specific treatment advantages in the 'Treatment for Adolescents with Depression Study'. J Psychiatr Res 2019 May;112:61-70. [doi: 10.1016/j.jpsychires.2019.02.021] [Medline: 30856378]
90. Kambeitz J, Goerigk S, Gattaz W, Falkai P, Benseñor IM, Lotufo PA, et al. Clinical patterns differentially predict response to transcranial direct current stimulation (tDCS) and escitalopram in major depression: A machine learning analysis of the ELECT-TDCS study. J Affect Disord 2020 Mar 15;265:460-467. [doi: 10.1016/j.jad.2020.01.118] [Medline: 32090773]
91. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Cai T, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. Mol Psychiatry 2016 Oct;21(10):1366-1371 [FREE Full text] [doi: 10.1038/mp.2015.198] [Medline: 26728563]
92. Pratap A, Atkins DC, Renn BN, Tanana MJ, Mooney SD, Anguera JA, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety 2019 Jan 21;36(1):72-81 [FREE Full text] [doi: 10.1002/da.22822] [Medline: 30129691]
93. Rajpurkar P, Yang J, Dass N, Vale V, Keller AS, Irvin J, et al. Evaluation of a Machine Learning Model Based on Pretreatment Symptoms and Electroencephalographic Features to Predict Outcomes of Antidepressant Treatment in Adults With Depression: A Prespecified Secondary Analysis of a Randomized Clinical Trial. JAMA Netw Open 2020 Jun 01;3(6):e206653 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.6653] [Medline: 32568399]
94. Kautzky A, Möller HJ, Dold M, Bartova L, Seemüller F, Laux G, et al. Combining machine learning algorithms for prediction of antidepressant treatment response. Acta Psychiatr Scand 2021 Jan 27;143(1):36-49 [FREE Full text] [doi: 10.1111/acps.13250] [Medline: 33141944]
95. Lee Y, Mansur RB, Brietzke E, Kapogiannis D, Delgado-Peraza F, Boutilier JJ, et al. Peripheral inflammatory biomarkers define biotypes of bipolar depression. Mol Psychiatry 2021 Jul 03;26(7):3395-3406 [FREE Full text] [doi: 10.1038/s41380-021-01051-y] [Medline: 33658605]
96. Ma Y, Ji J, Huang Y, Gao H, Li Z, Dong W, et al. Implementing machine learning in bipolar diagnosis in China. Transl Psychiatry 2019 Nov 18;9(1):305 [FREE Full text] [doi: 10.1038/s41398-019-0638-8] [Medline: 31740657]
97. Deng F, Wang Y, Huang H, Niu M, Zhong S, Zhao L, et al. Abnormal segments of right uncinate fasciculus and left anterior thalamic radiation in major and bipolar depression. Prog Neuropsychopharmacol Biol Psychiatry 2018 Feb 02;81:340-349. [doi: 10.1016/j.pnpbp.2017.09.006] [Medline: 28912043]
98. Nielsen SFV, Madsen KH, Vinberg M, Kessing LV, Siebner HR, Miskowiak KW. Whole-Brain Exploratory Analysis of Functional Task Response Following Erythropoietin Treatment in Mood Disorders: A Supervised Machine Learning Approach. Front Neurosci 2019 Nov 20;13:1246 [FREE Full text] [doi: 10.3389/fnins.2019.01246] [Medline: 31824247]
99. Hoogendoorn M, Berger T, Schulz A, Stolz T, Szolovits P. Predicting Social Anxiety Treatment Outcome Based on Therapeutic Email Conversations. IEEE J. Biomed. Health Inform 2017 Sep;21(5):1449-1459. [doi: 10.1109/jbhi.2016.2601123]
100. Ihmig FR, Neurohr-Parakenings F, Schäfer SK, Lass-Hennemann J, Michael T. On-line anxiety level detection from biosignals: Machine learning based on a randomized controlled trial with spider-fearful individuals. PLoS One 2020 Jun 23;15(6):e0231517 [FREE Full text] [doi: 10.1371/journal.pone.0231517] [Medline: 32574167]
101. Sharma A, Verbeke WJMI. Understanding importance of clinical biomarkers for diagnosis of anxiety disorders using machine learning models. PLoS One 2021 May 10;16(5):e0251365 [FREE Full text] [doi: 10.1371/journal.pone.0251365] [Medline: 33970950]
102. Demiris G, Corey Magan KL, Parker Oliver D, Washington KT, Chadwick C, Voigt JD, et al. Spoken words as biomarkers: using machine learning to gain insight into communication as a predictor of anxiety. J Am Med Inform Assoc 2020 Jun 01;27(6):929-933 [FREE Full text] [doi: 10.1093/jamia/ocaa049] [Medline: 32374378]
103. Frick A, Engman J, Alaie I, Björkstrand J, Gingnell M, Larsson E, et al. Neuroimaging, genetic, clinical, and demographic predictors of treatment response in patients with social anxiety disorder. J Affect Disord 2020 Jan 15;261:230-237 [FREE Full text] [doi: 10.1016/j.jad.2019.10.027] [Medline: 31655378]
104. Lebowitz ER, Zilcha-Mano S, Orbach M, Shimshoni Y, Silverman WK. Moderators of response to child-based and parent-based child anxiety treatment: a machine learning-based analysis. J Child Psychol Psychiatry 2021 Oct 24;62(10):1175-1182. [doi: 10.1111/jcpp.13386] [Medline: 33624848]
105. Lenhard F, Sauer S, Andersson E, Månsson KN, Mataix-Cols D, Rück C, et al. Prediction of outcome in internet-delivered cognitive behaviour therapy for paediatric obsessive-compulsive disorder: A machine learning approach. Int J Methods Psychiatr Res 2018 Mar 28;27(1):e1576 [FREE Full text] [doi: 10.1002/mpr.1576] [Medline: 28752937]

106. Schultebraucks K, Qian M, Abu-Amara D, Dean K, Laska E, Siegel C, et al. Pre-deployment risk factors for PTSD in active-duty personnel deployed to Afghanistan: a machine-learning approach for analyzing multivariate predictors. Mol Psychiatry 2021 Sep 02;26(9):5011-5022 [FREE Full text] [doi: 10.1038/s41380-020-0789-2] [Medline: 32488126]
107. Rangaprakash D, Deshpande G, Daniel TA, Goodman AM, Robinson JL, Salibi N, et al. Compromised hippocampus-striatum pathway as a potential imaging biomarker of mild-traumatic brain injury and posttraumatic stress disorder. Hum Brain Mapp 2017 Jun 15;38(6):2843-2864 [FREE Full text] [doi: 10.1002/hbm.23551] [Medline: 28295837]
108. Schultebraucks K, Shalev AY, Michopoulos V, Grudzen CR, Shin S, Stevens JS, et al. A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor. Nat Med 2020 Jul 06;26(7):1084-1088. [doi: 10.1038/s41591-020-0951-z] [Medline: 32632194]
109. Grisanzio KA, Goldstein-Piekarski AN, Wang MY, Rashed Ahmed AP, Samara Z, Williams LM. Transdiagnostic Symptom Clusters and Associations With Brain, Behavior, and Daily Function in Mood, Anxiety, and Trauma Disorders. JAMA Psychiatry 2018 Feb 01;75(2):201-209 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.3951] [Medline: 29197929]
110. Hinrichs R, van Rooij SJH, Michopoulos V, Schultebraucks K, Winters S, Maples-Keller J, et al. Increased Skin Conductance Response in the Immediate Aftermath of Trauma Predicts PTSD Risk. Chronic Stress (Thousand Oaks) 2019 Apr 24;3:247054701984444 [FREE Full text] [doi: 10.1177/2470547019844441] [Medline: 31179413]
111. Breen MS, Thomas KG, Baldwin DS, Lipinska G. Modelling PTSD diagnosis using sleep, memory, and adrenergic metabolites: An exploratory machine-learning study. Hum Psychopharmacol 2019 Mar 22;34(2):e2691. [doi: 10.1002/hup.2691] [Medline: 30793802]
112. Jin C, Jia H, Lanka P, Rangaprakash D, Li L, Liu T, et al. Dynamic brain connectivity is a better predictor of PTSD than static connectivity. Hum Brain Mapp 2017 Sep 12;38(9):4479-4496 [FREE Full text] [doi: 10.1002/hbm.23676] [Medline: 28603919]
113. Salminen LE, Morey RA, Riedel BC, Jahanshad N, Dennis EL, Thompson PM. Adaptive Identification of Cortical and Subcortical Imaging Markers of Early Life Stress and Posttraumatic Stress Disorder. J Neuroimaging 2019 May;29(3):335-343 [FREE Full text] [doi: 10.1111/jon.12600] [Medline: 30714246]
114. Yuan M, Qiu C, Meng Y, Ren Z, Yuan C, Li Y, et al. Pre-treatment Resting-State Functional MR Imaging Predicts the Long-Term Clinical Outcome After Short-Term Paroxtine Treatment in Post-traumatic Stress Disorder. Front Psychiatry 2018 Oct 30;9:532 [FREE Full text] [doi: 10.3389/fpsyt.2018.00532] [Medline: 30425661]
115. Zandvakili A, Swearingen HR, Philip NS. Changes in functional connectivity after theta-burst transcranial magnetic stimulation for post-traumatic stress disorder: a machine-learning study. Eur Arch Psychiatry Clin Neurosci 2021 Feb 27;271(1):29-37. [doi: 10.1007/s00406-020-01172-5] [Medline: 32719969]
116. Marmar CR, Brown AD, Qian M, Laska E, Siegel C, Li M, et al. Speech-based markers for posttraumatic stress disorder in US veterans. Depress Anxiety 2019 Jul 22;36(7):607-616 [FREE Full text] [doi: 10.1002/da.22890] [Medline: 31006959]
117. Iceta S, Tardieu S, Nazare J, Dougkas A, Robert M, Disse E. An artificial intelligence-derived tool proposal to ease disordered eating screening in people with obesity. Eat Weight Disord 2021 Oct 02;26(7):2381-2385. [doi: 10.1007/s40519-020-01076-2] [Medline: 33387276]
118. Senger K, Schröder A, Kleinstäuber M, Rubel JA, Rief W, Heider J. Predicting optimal treatment outcomes using the Personalized Advantage Index for patients with persistent somatic symptoms. Psychother Res 2022 Feb 29;32(2):165-178. [doi: 10.1080/10503307.2021.1916120] [Medline: 33910487]
119. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42-47 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
120. Joseph JE, Vaughan BK, Camp CC, Baker NL, Sherman BJ, Moran-Santa Maria M, et al. Oxytocin-Induced Changes in Intrinsic Network Connectivity in Cocaine Use Disorder: Modulation by Gender, Childhood Trauma, and Years of Use. Front Psychiatry 2019 Jul 19;10:502 [FREE Full text] [doi: 10.3389/fpsyt.2019.00502] [Medline: 31379621]
121. Laska EM, Siegel CE, Lin Z, Bogenschutz M, Marmar CR. Gabapentin Enacarbil Extended-Release Versus Placebo: A Likely Responder Reanalysis of a Randomized Clinical Trial. Alcohol Clin Exp Res 2020 Sep 31;44(9):1875-1884 [FREE Full text] [doi: 10.1111/acer.14414] [Medline: 33460198]
122. Pan Y, Liu H, Metsch LR, Feaster DJ. Factors Associated with HIV Testing Among Participants from Substance Use Disorder Treatment Programs in the US: A Machine Learning Approach. AIDS Behav 2017 Feb 8;21(2):534-546 [FREE Full text] [doi: 10.1007/s10461-016-1628-y] [Medline: 27933461]
123. Schmitgen MM, Niedtfeld I, Schmitt R, Mancke F, Winter D, Schmahl C, et al. Individualized treatment response prediction of dialectical behavior therapy for borderline personality disorder using multimodal magnetic resonance imaging. Brain Behav 2019 Sep 14;9(9):e01384 [FREE Full text] [doi: 10.1002/brb3.1384] [Medline: 31414575]
124. Bang S, Son S, Roh H, Lee J, Bae S, Lee K, et al. Quad-phased data mining modeling for dementia diagnosis. BMC Med Inform Decis Mak 2017 May 18;17(Suppl 1):60-10 [FREE Full text] [doi: 10.1186/s12911-017-0451-3] [Medline: 28539115]
125. Yao T, Sweeney E, Nagorski J, Shulman JM, Allen GI. Quantifying cognitive resilience in Alzheimer's Disease: The Alzheimer's Disease Cognitive Resilience Score. PLoS One 2020 Nov 5;15(11):e0241707 [FREE Full text] [doi: 10.1371/journal.pone.0241707] [Medline: 33152028]

126. Grassi M, Rouleaux N, Caldirola D, Loewenstein D, Schruers K, Perna G, Alzheimer's Disease Neuroimaging Initiative. A Novel Ensemble-Based Machine Learning Algorithm to Predict the Conversion From Mild Cognitive Impairment to Alzheimer's Disease Using Socio-Demographic Characteristics, Clinical Information, and Neuropsychological Measures. Front Neurol 2019 Jul 16;10:756 [FREE Full text] [doi: 10.3389/fneur.2019.00756] [Medline: 31379711]
127. Li Q, Zhao Y, Chen Y, Yue J, Xiong Y. Developing a machine learning model to identify delirium risk in geriatric internal medicine inpatients. Eur Geriatr Med 2022 Feb 23;13(1):173-183. [doi: 10.1007/s41999-021-00562-9] [Medline: 34553310]
128. Andersson S, Bathula DR, Iliadis SI, Walter M, Skalkidou A. Predicting women with depressive symptoms postpartum with machine learning methods. Sci Rep 2021 Apr 12;11(1):7877 [FREE Full text] [doi: 10.1038/s41598-021-86368-y] [Medline: 33846362]
129. Li Y, Rosenfeld B, Pessin H, Breitbart W. Bayesian Nonparametric Clustering of Patients with Advanced Cancer on Anxiety and Depression. 2017 Presented at: 16th IEEE International Conference on Machine Learning and Applications (ICMLA); December 18-21, 2017; Cancun, Mexico p. 674-678. [doi: 10.1109/icmla.2017.00-83]
130. Zhou S, Fei S, Han H, Li J, Yang S, Zhao C. A Prediction Model for Cognitive Impairment Risk in Colorectal Cancer after Chemotherapy Treatment. Biomed Res Int 2021 Feb 20;2021:6666453-6666413 [FREE Full text] [doi: 10.1155/2021/6666453] [Medline: 33688501]
131. Vitinius F, Escherich S, Deter H, Hellmich M, Jünger J, Petrowski K, et al. Somatic and sociodemographic predictors of depression outcome among depressed patients with coronary artery disease - a secondary analysis of the SPIRR-CAD study. BMC Psychiatry 2019 Feb 04;19(1):57 [FREE Full text] [doi: 10.1186/s12888-019-2026-6] [Medline: 30717711]
132. Wallert J, Gustafson E, Held C, Madison G, Norlund F, von Essen L, et al. Predicting Adherence to Internet-Delivered Psychotherapy for Symptoms of Depression and Anxiety After Myocardial Infarction: Machine Learning Insights From the U-CARE Heart Randomized Controlled Trial. J Med Internet Res 2018 Oct 10;20(10):e10754 [FREE Full text] [doi: 10.2196/10754] [Medline: 30305255]
133. Gu S, Zhou J, Yuan C, Ye Q. Personalized prediction of depression in patients with newly diagnosed Parkinson's disease: A prospective cohort study. J Affect Disord 2020 May 01;268:118-126. [doi: 10.1016/j.jad.2020.02.046] [Medline: 32158001]
134. Montazeri F, de Bildt A, Dekker V, Anderson GM. Network Analysis of Behaviors in the Depression and Autism Realms: Inter-Relationships and Clinical Implications. J Autism Dev Disord 2020 May 18;50(5):1580-1595. [doi: 10.1007/s10803-019-03914-4] [Medline: 30778821]
135. Ekhtiari H, Kuplicki R, Yeh H, Paulus MP. Physical characteristics not psychological state or trait characteristics predict motion during resting state fMRI. Sci Rep 2019 Jan 23;9(1):419 [FREE Full text] [doi: 10.1038/s41598-018-36699-0] [Medline: 30674933]
136. Rosellini AJ, Stein MB, Benedek DM, Bliese PD, Chiu WT, Hwang I, et al. Predeployment predictors of psychiatric disorder-symptoms and interpersonal violence during combat deployment. Depress Anxiety 2018 Nov 13;35(11):1073-1080 [FREE Full text] [doi: 10.1002/da.22807] [Medline: 30102442]
137. Smith R, Feinstein JS, Kuplicki R, Forthman KL, Stewart JL, Paulus MP, Tulsa 1000 Investigators, et al. Perceptual insensitivity to the modulation of interoceptive signals in depression, anxiety, and substance use disorders. Sci Rep 2021 Jan 22;11(1):2108 [FREE Full text] [doi: 10.1038/s41598-021-81307-3] [Medline: 33483527]
138. Liang S, Vega R, Kong X, Deng W, Wang Q, Ma X, et al. Neurocognitive Graphs of First-Episode Schizophrenia and Major Depression Based on Cognitive Features. Neurosci Bull 2018 Apr 2;34(2):312-320 [FREE Full text] [doi: 10.1007/s12264-017-0190-6] [Medline: 29098645]
139. Popovic D, Ruef A, Dwyer DB, Antonucci LA, Eder J, Sanfelici R, PRONIA Consortium. Traces of Trauma: A Multivariate Pattern Analysis of Childhood Trauma, Brain Structure, and Clinical Phenotypes. Biol Psychiatry 2020 Dec 01;88(11):829-842. [doi: 10.1016/j.biopsych.2020.05.020] [Medline: 32782139]
140. Zandvakili A, Philip NS, Jones SR, Tyrka AR, Greenberg BD, Carpenter LL. Use of machine learning in predicting clinical response to transcranial magnetic stimulation in comorbid posttraumatic stress disorder and major depression: A resting state electroencephalography study. J Affect Disord 2019 Jun 01;252:47-54 [FREE Full text] [doi: 10.1016/j.jad.2019.03.077] [Medline: 30978624]
141. Perez Arribas I, Goodwin GM, Geddes JR, Lyons T, Saunders KEA. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. Transl Psychiatry 2018 Dec 13;8(1):274 [FREE Full text] [doi: 10.1038/s41398-018-0334-0] [Medline: 30546013]
142. Tate AE, McCabe RC, Larsson H, Lundström S, Lichtenstein P, Kuja-Halkola R. Predicting mental health problems in adolescence using machine learning techniques. PLoS One 2020 Apr 6;15(4):e0230389 [FREE Full text] [doi: 10.1371/journal.pone.0230389] [Medline: 32251439]
143. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020 Sep 09;370:m3164 [FREE Full text] [doi: 10.1136/bmj.m3164] [Medline: 32909959]
144. Rivera SC, Liu X, Chan A, Denniston AK, Calvert MJ, SPIRIT-AICONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. BMJ 2020 Sep 09;370:m3210 [FREE Full text] [doi: 10.1136/bmj.m3210] [Medline: 32907797]

145. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med 2020 Sep 09;26(9):1320-1324 [FREE Full text] [doi: 10.1038/s41591-020-1041-y] [Medline: 32908275]
146. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021 Jul 09;11(7):e048008 [FREE Full text] [doi: 10.1136/bmjopen-2020-048008] [Medline: 34244270]
147. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. J R Soc Med 2016 Feb;109(2):67-77 [FREE Full text] [doi: 10.1177/0141076815625599] [Medline: 26880653]

Abbreviations

AI: artificial intelligence
CONSORT: Consolidated Standards of Reporting Trials
CHARMS: Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies
DOI: digital object identifier
EQUATOR: Enhancing the Quality and Transparency of Health Research
ICD-11: International Classification of Diseases 11th Revision
IEEE: Institute of Electrical and Electronics Engineers
MI-CLAIM: Minimum Information About Clinical Artificial Intelligence Modeling
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Risk of Bias Assessment Tool
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials
SVM: support vector machine
TRIPOD: Transparent Reporting of a Multivariable Prediction Model of Individual Prognosis or Diagnosis

Edited by J Torous; submitted 19.08.22; peer-reviewed by D Shung, Z Dai, C Price; comments to author 09.09.22; revised version received 02.11.22; accepted 20.11.22; published 02.02.23.

Please cite as:
Tornero-Costa R, Martinez-Millana A, Azzopardi-Muscat N, Lazeri L, Traver V, Novillo-Ortiz D
Methodological and Quality Flaws in the Use of Artificial Intelligence in Mental Health Research: Systematic Review
JMIR Ment Health 2023;10:e42045
URL: https://mental.jmir.org/2023/1/e42045
doi:10.2196/42045
PMID:36729567

Review

The Use of Sensors to Detect Anxiety for In-the-Moment Intervention: Scoping Review

Rosie Dobson[1,2], PhD; Linwei Lily Li[3]; Katie Garner[1], MHealthPsych; Taria Tane[3], PGDipArts; Judith McCool[3], PhD; Robyn Whittaker[1,2], PhD

[1]National Institute for Health Innovation, University of Auckland, Auckland, New Zealand

[2]Institute for Innovation and Improvement, Te Whatu Ora Waitematā, Auckland, New Zealand

[3]School of Population Health, University of Auckland, Auckland, New Zealand

Corresponding Author:
Rosie Dobson, PhD
National Institute for Health Innovation
University of Auckland
Building 507, Grafton Campus
22-30 Park Avenue, Grafton
Auckland, 1023
New Zealand
Phone: 64 9 373 7599
Fax: 64 9 373 1710
Email: r.dobson@auckland.ac.nz

Abstract

Background: With anxiety a growing issue and barriers to accessing support services, there is a need for innovative solutions to provide early intervention. In-the-moment interventions support individuals to recognize early signs of distress and use coping mechanisms to prevent or manage this distress. There is potential for wearable sensors linked to an individual's mobile phone to provide in-the-moment support tailored to individual needs and physiological responses.

Objective: The aim of this scoping review is to examine the role of sensors in detecting the physiological signs of anxiety to initiate and direct interventions for its management.

Methods: Relevant studies were identified through searches conducted in Embase, MEDLINE, APA PsycINFO, ProQuest, and Scopus. Studies were identified if they were conducted with people with stress or anxiety or at risk of anxiety and included a wearable sensor providing real-time data for in-the-moment management of anxiety.

Results: Of the 1087 studies identified, 11 studies were included in the review, including 5 randomized controlled trials and 6 pilot or pretesting studies. The results showed that most studies successfully demonstrated improvements in their target variables. This included overall anxiety and stress levels, and the implementation of in-the-moment stress and anxiety management techniques such as diaphragmatic breathing. There was wide variation in the types of sensors used, physiological measures, and sensor-linked interventions.

Conclusions: This review indicates that sensors are potentially a useful tool in detecting anxiety and facilitating the implementation of a known control mechanism to reduce anxiety and improve mood, but further work is needed to understand the acceptability and effectiveness of this type of intervention.

(JMIR Ment Health 2023;10:e42611) doi:10.2196/42611

KEYWORDS

anxiety; wearables; sensors; mental health; digital mental health; digital health intervention; wearable device

Introduction

Anxiety is a critical issue internationally [1], with COVID-19 and its associated lockdowns further increasing global psychological morbidity [2-4]. Stress and anxiety increase the risk of several harmful behaviors including problem drinking, drug use, self-harm, and suicide [5], and the annual cost of the burden of serious mental illness, addiction, and suicide is considerable [6,7].

Access to treatments for anxiety and other mental health conditions are a significant challenge [8]. The current system provides accessible treatment for those with diagnosed serious mental illness but fails to support what is commonly referred to as the *missing middle*, people with mild to moderate mental health needs that may not fit diagnostic criteria. For example, young people in particular generally only access mental health services when their condition has started to significantly affect their daily life, with many continuing to suffer from extended periods of mild to moderate anxiety-related morbidity alone [9,10]. The stigma associated with mental health conditions combined with a perceived lack of accessible support can exacerbate the experience of anxiety for some [11,12]. Even with the growth in web-based/phone services offering immediate support for people with mental health concerns, uptake is varied [13]. Concerns around privacy and being too intrusive are known to prevent people from accessing digital mental health services [14]. Further, facilitators to engagement with these types of interventions include personalization and that they enhance feelings of control over the person's health [15]. Self-help–focused in-the-moment digital solutions that help individuals to control their own mood and emotions may have the most potential [13].

Wearable devices, which are worn on the body or clothing, provide an innovative tool for the detection, diagnosis, and management of health conditions through the noninvasive measurement of physiological information in real time [8]. There is growing interest in and accessibility of wearable sensors, with rapid uptake in the wider population [16]. Newer wearable sensors enable additional data collection that can be linked to an individual's mobile phone to tailor support to their needs and physiological responses. The use of digital biofeedback techniques in anxiety is growing [17], presenting the potential for simple personal mobile interventions designed to target anxiety. For example, by alerting individuals early to the signs of their increasing stress or anxiety, there is the potential to circumvent an episode before it worsens. Early evidence has shown the potential for the use of wearable sensors for this purpose [18-22]. Although commercially available sensors are more commonly used for fitness tracking rather than mental health, with their wide uptake, they provide a potential tool for accessible population-based mental health interventions.

Despite the subjective nature of assessing anxiety, the measurement of the physiological changes common in anxiety can be used as indicators of a person's stress response and experience of anxiety. A number of biomarkers have been found to provide real-time objective measurement of the physiological impacts of anxiety and exposure to stressors [23,24], including cardiac changes (eg, heart rate variability [HRV], pulse photoplethysmography [PPG]), changes in respiratory rate, changes in electrical activity in the brain (eg, electroencephalogram [EEG]), changes in body temperature, and galvanic skin response (eg, electrodermal activity [EDA]). Many of these biomarkers can be measured in real time using commercially available sensors potentially making the detection of anxiety in real-world settings feasible.

This paper provides the findings of a scoping review on the role of sensors in the management of anxiety with the goal of informing the development of digital interventions to treat mild to moderate anxiety.

Methods

Design

The aim of this scoping review was to examine the role of sensors in detecting the physiological signs of anxiety to initiate and direct interventions for its management. The review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines for the reporting of scoping reviews [25] (see Multimedia Appendix 1), and the protocol was not published.

Eligibility Criteria

Eligible studies included human participants of any age with anxiety or stress, or at risk of anxiety. Due to the limited number of studies, a decision was made to include studies of participants with stress, not just anxiety. Comorbidity of anxiety with other psychiatric conditions (eg, depression) were also considered since anxiety may often coexist with or be the result of other psychiatric conditions. Eligible studies needed to include a wearable sensor that provided real-time data/information for in-the-moment management of anxiety. As the focus of the review was on the use of sensors for the identification of early signs of anxiety, before the individual is likely to be consciously aware of the anxiety episode, only sensors that collected physiological data in a passive manner (not requiring active engagement from the user) were included; sensors that required manual input of data only were excluded. Sensors had be worn on the body, but no other restrictions on the type were applied (eg, commercial clinical grade). Studies including multifaceted interventions in which sensors were just one component were included in the review.

All study designs were included with the exception of protocols and reviews. There was no limit on year of publication because of the recency of the study subject. No restrictions on comparator or outcome measures were included. The review was restricted to full-text articles published in peer-reviewed journals. Studies were excluded if published in languages other than English or were published only in the form of conference abstracts.

Study Search Strategy

Searches were conducted from inception through to December 8, 2021, using Embase, MEDLINE, APA PsycINFO, ProQuest, and Scopus. Searches were limited to papers published in the English language. The initial search strategy conducted included words associated with anxiety, such as "stress," to form search strings. However, these produced results that were too broad, and the term was excluded. An example of the revised search strategy can be seen in Textbox 1. Reference lists of relevant previous reviews and included studies were searched for additional papers.

Textbox 1. Example search strategy.

1. exp wearable devices/ or wearable*.mp.
2. sensor.mp.
3. 1 or 2
4. exp anxiety/ or anxiety.mp.
5. (psychological stress or psychological distress).mp.
6. 4 or 5
7. 3 and 6
8. limit 7 to English language

Study Selection

Assessment of study eligibility was performed on the Rayyan platform. The first assessment was screened independently by LLL, identifying duplicates and filtering based on the title and abstract. The second screening of full-text papers was then undertaken by three authors LLL, RW, and RD, and disagreements between reviewers were resolved by discussion.

Data Extraction

Data were extracted using structured forms including study design (design, duration, setting), population characteristics, intervention type, sensor characteristics (type of sensor, testing conditions, role of sensor, sensor measures, sensor-linked intervention), and outcomes/findings. A narrative synthesis methodology was used to synthesize the data extracted.

Results

Study Selection

The total number of studies identified was 1139 from database searches, in which 51 were duplicates. A further 4 studies were identified through reference lists. This resulted in 1092 studies screened by title and abstract, with 911 being considered ineligible. The full texts of the remaining 181 full-text papers were reviewed according to the inclusion criteria. At the end of the screening process, 11 articles were included in the scoping systematic review. Figure 1 shows the study selection process, and Table 1 presents a summary of the included studies.

Figure 1. Flow diagram of study selection.

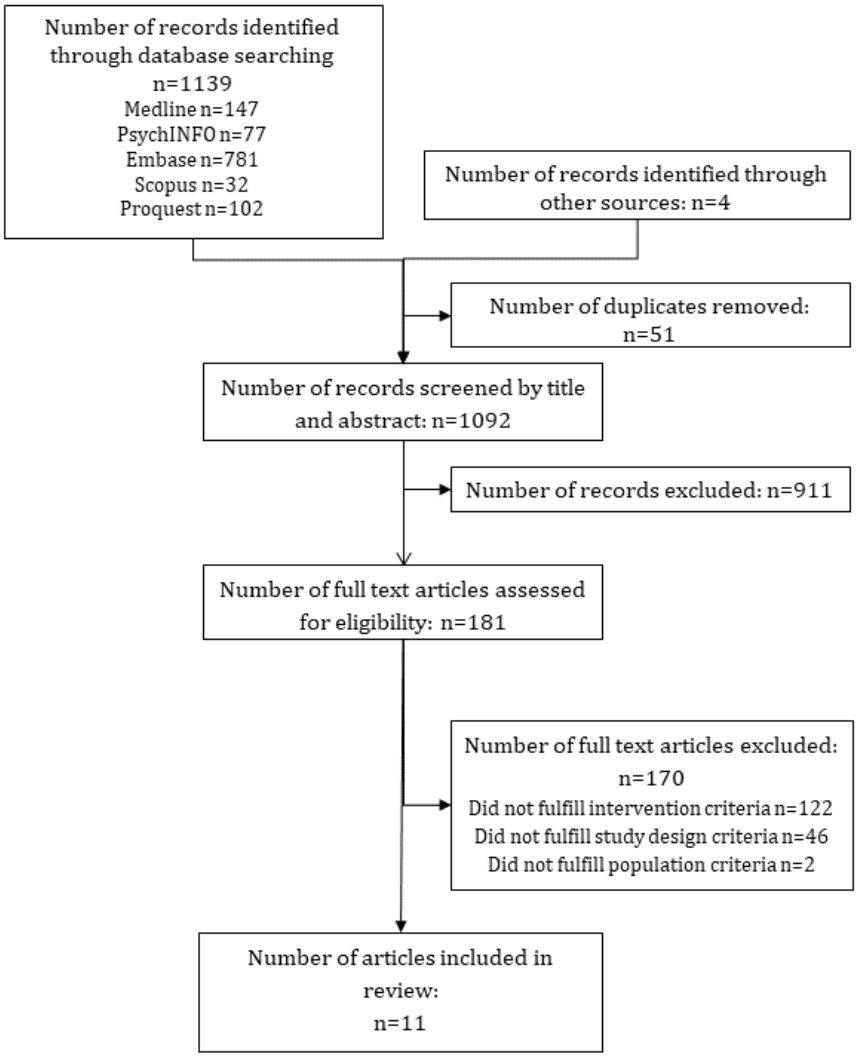


Table 1. Characteristics of included studies.

Study	Study design	Duration	Country	Population	Intervention type	Key findings
Chung et al [26], 2021	Pilot study	8 weeks	United States	Adults with a GAD-2[a] score ≥3 and PHQ-2[b] score ≤4 aged 24-47 years (N=14)	Assessed the feasibility of an HRVB[c] wearable device and remote stress management coach to reduce anxiety	Anxiety scores and depression scores decreased over the 8-week intervention**
Crivelli et al [27], 2018	Pretesting study	2 weeks	Italy	Adult professionals with no history of psychiatric or neurological disease (N=16), mean age 44.4 years	To test the training of Vipasyana meditation and technology-mediated mental training for stress management in people who were at risk of stress (which involved real-time acoustic feedback via an app based on changes in the physiological signature of the participant's mindset)	Significant decrease to perceived stress scores*, situational anxiety**, and anger and fatigue**
Jaramillo-Quintanar et al [28], 2020	Pretesting study	1 session	Mexico	Children with high stress levels (N=29), mean age 8.7 years	To test the feasibility of i-CARE, which measures HR[d] and blood oxygenation, and provides visual and auditory biofeedback to learn to regulate symptoms of anxiety	Results demonstrate that i-CARE is effective at inducing relaxation in children with high stress (no *P* value reported)
Kizakevich et al [29], 2019	Pilot study	1 year	United States	Adults with a military background (N=328)	To test the effectiveness of 4 different resilience training techniques each with or without HRV[e] biofeedback for people with a risk of stress; continuous acquisition of HRV data enables analysis of physiological response to stress and breathing training	No conclusions on the results of effectiveness could be drawn from the data
Millings et al [30], 2015	RCT[f]	4 weeks	United Kingdom	University students with >14 on the PSS-10[g] and <19 on the Beck Depression Inventory (N=92), mean age 23.7 years	Stress management program and a prototype wearable sensor kit comprising of and ECG[h] and EEG[i] sensor; compared stress management program alone, program and sensor, or no intervention groups	Significant reductions in levels of stress in those in the program alone condition*; however, sensors impeded the effectiveness of the program potentially due to technical issues
Nguyen et al [31], 2021	RCT	4 training sessions	Canada	Children and adolescents with ASD[j] aged 8-18 years (N=28)	Use of anxiety meter and breathing techniques; on the fourth visit participants were randomized to receive feedback on anxiety level or no feedback while completing a stress-eliciting task	The anxiety meter improved awareness of anxiety states, which lead to increased likelihood of initiating calming strategies under stress**
Ponzo et al [32], 2020	RCT	4 weeks	United Kingdom	University students aged 18-25 years with score >14 (stress) or >7 (anxiety) on DASS[k] (N=262)	Biobase program, mobile app comprising psychoeducational content, mood tracking via EMA[l], and in-the-moment exercises for stress and anxiety (eg, relaxation); real-time sensor data presented to user via app dashboard	Well-being increased and anxiety decreased during intervention and was sustained 2 weeks after the intervention**
Shruthi et al [33], 2021	Pilot study	2 months	India	Students (N=50), mean age 19 years	Wristband to treat anxiety that provided acupressure to the H7 point on the wrist crease when completing a stressful task vs group wearing Fitbit-like band	Intervention group had lower levels of anxiety compared to the active control group (note, no *P* values reported)

Study	Study design	Duration	Country	Population	Intervention type	Key findings
Serino et al [34], 2014	Pilot study	120 seconds or more on the app exercises	Italy	App users (N=68)	App that teaches guided relaxation, 3D biofeedback training, and stress self-tracking to help control respiration rate and therefore HR	The stress management exercises along with the app led to a significant decrease in perceived psychological stress**
Smith et al [35], 2020	RCT	4 weeks	United States	Adults who work in knowledgeable occupations (N=215), mean age 33.2 years	App for stress and anxiety that delivers mindfulness-based breathing from MBSR[m] and Spire Stone to measure subjective emotional state changes; both provide biofeedback about physiological state; the app provided push notifications	Participants in the treatment group experienced 15.8% fewer negative instances of stress**, 13% fewer instances of distressing symptoms, and 28.2%* fewer days feeling anxious compared to waitlist control**
Winslow et al [36], 2016	RCT	8-10 weeks	United States	Veterans (N=16), mean age 39.8 years; self-reported anger and stress	In-person CBT[n] + sensor and mobile app; alerted the user through the app when stress was detected and presented with prompts or reminders to engage with stress mitigation techniques	High attrition; significant reduction in anxiety and stress observed between intervention and control*; no difference in depression

[a]GAD-2: Generalized Anxiety Disorder 2-item.

[b]PHQ-2: Patient Health Questionnaire–2.

[c]HRVB: heart rate variability biofeedback.

[d]HR: heart rate.

[e]HRV: heart rate variability.

[f]RCT: randomized controlled trial.

[g]PSS-10: Perceived Stress Scale.

[h]ECG: electrocardiogram.

[i]EEG: electroencephalogram.

[j]ASD: autism spectrum disorder.

[k]DASS: Depression Anxiety Stress Scales.

[l]EMA: ecological momentary assessment.

[m]MBSR: mindfulness-based stress reduction.

[n]CBT: cognitive behavioral therapy.

*P<.05, **P<.01

Study Design

Of the studies included in this scoping review, 5 were randomized controlled trials (RCTs), and 6 were pilot studies or pretesting studies. Sample sizes ranged from 14 to 328 participants. The duration of the studies varied, with the shortest being one session [28,34] and the longest 8-10 weeks [36]. The conditions in which the sensor was worn or used also varied. Two studies were performed in lab conditions [28,31]. Six studies had participants wear the sensor continuously in everyday life [26,29,30,32,33,35], 1 study included combined conditions where participants wore sensors and attended weekly cognitive behavioral therapy (CBT) sessions [36], 1 study required participants to wear the sensor during meditation practice at home [27], and 1 study did not clearly define their testing conditions [34].

Participant Characteristics

Studies were conducted in the United Kingdom [30,32], the United States [26,29,35,36], Italy [27,34], Canada [31], Mexico [28], and India [33]. All studies were conducted in adult participants except 2 studies that were conducted with children [28,31]. Two studies included participants with anxiety [26,33] (1 diagnosed using a Generalized Anxiety Disorder 2-item score ≥3 and a Patient Health Questionnaire–2 score ≤4; the other had no specified measure), 4 studies were conducted in participants with no reported stress but may have had risk of stress (eg, had a high stress job or had downloaded a stress app) [27,29,34,35], 1 study included participants with at least mild levels of stress [32], 3 studies included participants experiencing high degrees of stress [28,30,36] (but only 1 [36] study had a clinical diagnosis), and 1 study included participants diagnosed with autism spectrum disorder (ASD) [31].

Sensor Characteristics

Table 2 presents a summary of the sensor characteristics of the included studies. Of the 11 studies, 3 included upper body sensors (ie, on chest or torso) [26,29,35], 3 used sensors that were wrist worn [32,33,36], 2 used sensors attached to the finger or thumb [28,36], and 1 study had a sensor in a pair of eyeglasses [27]; in 2 studies, there was no detail provided [30,34]. Cardiac activity was the main sensor measure in 8

studies [26,28-34]. Within these studies, heart rate (HR) was obtained by all 8 studies and HRV was calculated in 2 studies [26,29,30]. Other sensor measures included accelerometer data [31,36], EEG activity [27], blood oxygenation [28,36], infrared measurements of facial temperature [28], physical activity [30,35], sleep [32,33], step count [32], water intake [33], respiratory effort [35], PPG, EDA, and body temperature [36].

All sensors had some form of capacity to display their sensor data back to the participant. HRV or HR was visually displayed back to participants in 6 studies [28-30,32-34]. Other studies visually displayed PPG, EDA [32], and respiratory amplitude [35]. During the intervention exercises, participants were able to see their biofeedback as raw numbers, or the data was portrayed in another mode (eg, change in color to correspond to their psychological state) [30,31,35]. For example, in the study with children with ASD, anxiety state was also displayed as a color, where green was calm, yellow was rising anxiety, and red was anxious [31]. Two studies that measured EEG activity provided real-time auditory feedback based on their mindset [27] and alpha waves relative to beta waves [30], which both corresponded to their state of relaxation.

In 3 studies [26,31,35], sensors provided physical feedback to the participants. Two studies used a vibration to alert users of their anxious state [31,35], and 2 studies used vibrations to help guide the user in their breathing in conjunction with visual biofeedback [26,35].

Most studies (10/11, 91%) included sensors that were linked to some form of breathing techniques such as diaphragmatic breathing to help manage stress or anxiety. Eight studies used guided diaphragmatic breathing [26,28-32,34,36], and 2 studies included mindfulness-based breathing techniques [27,35]. Only 1 study did not include any breathing techniques but used acupressure on the wrist to help reduce anxiety [33].

Some of the studies had interventions that included additional information or therapy to supplement stress or anxiety management. Three studies included CBT-based techniques as part of their intervention [26,30,36], and 1 study had participants attend CBT once per week [36]. Two studies included psychoeducation on stress management [30,32]. In terms of whether participants were receiving treatment for anxiety outside the intervention, 1 study excluded participants who were receiving other forms of treatment [26], and the remaining studies did not specify this information.

Table 2. Sensor characteristics.

Study	Type of sensor	Testing conditions	Role of sensor	Sensor measures	Sensor-linked intervention	Sensor usability and acceptability
Chung et al [26], 2021	Lief Smart Patch; worn on torso under clothing	Continuous wear in everyday life	Reduce anxiety	HR[a], HRV[b], and accelerometer data through continuous monitoring	HRVB[c] wearable device and remote stress management coach	86% of participants wore the patch and completed ≥1- to 3-min HRVB exercise on at least 40 of 56 days; only 43% completed 3 or more 3-min HRVB exercises on at least 40 of 56 days
Crivelli et al [27], 2018	Lowdown Focus brain-sensing eyeglasses	Wore glasses when participating in meditation practice	Stress management	EEG[d] activity	Vipasyana meditation and technology-mediated mental training for stress management (which involved real-time acoustic feedback via an app based on changes in physiological signature of the participants' mindset)	Not reported
Jaramillo-Quintanar et al [28], 2020	i-CARE; sensor on child's finger	Lab administered	Regulate symptoms of anxiety	HR, blood oxygenation, and infrared measurements of facial temperature	i-CARE, which measures HR and blood oxygenation, and provides visual and auditory biofeedback	Not reported
Kizakevich et al [29], 2019	BART; chest belt sensor	Continuous wear in everyday life	Manage anxiety	HRV	4 different resilience training techniques each with or without HRV biofeedback; continuous acquisition of HRV data enables analysis of physiological response to stress and breathing training	Participants in the 6-week training regimen completed; 600 sessions during the first week; however, over the next several weeks, training compliance fell by almost one-third and to about one-half after 1 month
Millings et al [30], 2015	Prototype wearable sensor kit	Continuous wear in everyday life	Stress management	HRV, HR, and physical activity	Stress management program and a prototype wearable sensor kit comprising of an ECG[e] and EEG sensor	Qualitative results found that many participants experienced technical issues that caused frustration
Nguyen et al [31], 2021	Wearable Shimmer2 unit	Lab administered	Manage anxiety	HR, anxiety-level feedback on tablet (green: calm; yellow: rising anxiety; red: anxious)	Use of anxiety meter and breathing techniques	Not reported
Ponzo et al [32], 2020	Biobeam; wrist worn; continuous wear	Continuous wear in everyday life	Manage stress and anxiety	Sleep duration and quality, HR, step count	Biobase program, mobile app comprising psychoeducational content, mood tracking via EMA[f] and in-the-moment exercises (eg, relaxation); real-time sensor data presented to user via app dashboard	Not reported
Shruthi et al [33], 2021	Prototype wrist band	Continuous wear in everyday life	Manage anxiety	Oxygen levels, water intake, sleep, HR	Wrist band that provided acupressure to the H7 point on the wrist crease when completing a stressful task vs group wearing Fitbit-like band	Not reported
Serino et al [34], 2014	App called Positive Technology	Worn when using the app (not well defined)	Manage stress	HR	App that teaches guided relaxation, 3D biofeedback training, and stress self-tracking	Not reported

Study	Type of sensor	Testing conditions	Role of sensor	Sensor measures	Sensor-linked intervention	Sensor usability and acceptability
Smith et al [35], 2020	Device called Spire Stone: a clothing-attached device	Continuous wear in everyday life	Manage stress and anxiety	Respiratory effort and physical activity	App that delivers mindfulness-based breathing from MBSR[g] and Spire Stone to measure subjective emotional state changes; both provide biofeedback about physiological state; the app provided push notifications	Participants wore the device 52% of the days during the intervention period; 75% completed at least one educational guided breathing session and only 19% completed all 5 sessions
Winslow et al [36], 2016	E3 band, wrist worn	Continuous wear in everyday life	Manage stress	PPG[h], EDA[i], body temperature, accelerometer, HR	In-person CBT[j] + sensor and mobile app; alerted the user through the app when stress was detected and presented with prompts or reminders to engage with stress mitigation techniques	Individuals in the experimental group completed a significantly greater number of therapy sessions compared to the control group; 1 participant in the experimental group did not use the app but completed CBT

[a]HR: heart rate.

[b]HRV: heart rate variability.

[c]HRVB: heart rate variability biofeedback.

[d]EEG: electroencephalogram.

[e]ECG: electrocardiogram.

[f]EMA: ecological momentary assessment.

[g]MBSR: mindfulness-based stress reduction.

[h]PPG: photoplethysmography.

[i]EDA: electrodermal activity.

[j]CBT: cognitive behavioral therapy.

Study Outcomes

Of the 5 RCT studies included in this review, 4 found significant improvements to measures of stress or anxiety. One study assessed the effectiveness of an anxiety meter to improve awareness of anxiety symptoms and, therefore, engage in relaxation techniques in children with ASD [31]. Researchers found after four visits, the anxiety meter improved awareness of anxiety states, which then led to increased likelihood of engaging in calming strategies under stress. In another study, researchers tested the efficacy of a mobile app called BioBase paired with a wearable device on reducing anxiety and improving well-being [32]. Participants were university students with elevated levels of anxiety and stress. Results showed that well-being improved and anxiety decreased during the intervention compared to waitlist controls, and this was sustained 2 weeks post intervention. Smith et al [35] looked at whether an app that delivers mindfulness-based stress reduction and a sensor could improve mental health outcomes. Participants were employees who were randomized to the intervention or waitlist control. Researchers found that those in the intervention group experienced 15.8% fewer negative instances of stress, 13% fewer instances of distressing symptoms, and 28.2% fewer days feeling anxious post intervention compared to the controls. Lastly, a study assessed the effectiveness of a sensor paired with in-person CBT for stress management in a sample of veterans who were experiencing high levels of stress, finding a significant reduction in stress and anxiety between the intervention group and controls at follow-up [36].

Interestingly, 1 study found contrasting findings [30]. Researchers assessed whether the effectiveness of an online intervention for stress could be enhanced by using a sensor. Students who were stressed were randomized to either the stress management program alone, the program and the sensor, or no intervention. After 4 weeks, the results showed significant reductions in stress in the program alone condition and similar but weaker reductions in the program and sensor condition compared to the controls. Therefore, researchers suggested that it was possible that the sensor impeded the efficacy of the stress management program.

Three pilot or pretesting studies also found promising improvements to either stress or anxiety. One study assessed the effectiveness of an HRV biofeedback wearable device and stress management coach to reduce symptoms of anxiety [26]. Researchers found that after the 8-week duration, the intervention led to a clinically significant decrease in anxiety and depression scores [26]. In another pilot study, researchers tested the effectiveness of an app that teaches guided relaxation, 3D biofeedback training, and stress self-tracking for stress and anxiety. For participants who engaged in 120 seconds or more on the exercises, there were significant decreases in perceived psychological stress and HR; however, the change in HR was not significant [34]. Lastly, Crivelli et al [27] looked at the effectiveness of a training protocol that used brain-sensing eyeglasses to help participants become more aware of their mindset and develop greater stress coping. Researchers found that, post training, there was a significant decrease in stress scores, situational anxiety, anger and fatigue.

Discussion

This review explored the use of wearable sensors to detect, manage, and treat anxiety. The results showed that most studies successfully demonstrated improvements in their target variables. This included overall anxiety and stress levels, and the implementation of in-the-moment stress and anxiety management techniques such as diaphragmatic breathing. Where applicable, most studies also noted a reduction in secondary symptoms such as depression, anger, and negative emotions. The results therefore indicate that sensors are potentially a useful tool in detecting stress and anxiety and facilitating the implementation of a known control mechanism to reduce anxiety and improve mood.

The utility of sensors was shown to extend beyond sensing physiological signs of anxiety and stress. Sensors have the potential not only to alert the user to their change in emotional state but also to connect with other systems and prompt interventions that can treat the detected changes and deliver those interventions through the sensor itself. For example, the results of this review showed that when a user is alerted to their change in anxiety, the screen on some sensors could be used to guide the user through the subsequent recommended breathing exercise. Exploration of other in-the-moment control mechanism or therapeutic techniques could be explored to diversify and maximize the treatment options available through sensors.

This review also identified that sensors can be used by a wide variety of people in a range of contexts. Studies ranged in age and clinical severity, from mild to moderate levels of stress, depression, or anxiety to diagnosed clinical disorders including anxiety, posttraumatic stress disorder, and ASD. Studies also varied in length and nature, from wearing sensors and learning coping mechanisms in brief laboratory sessions to wearing the sensors for multiple weeks of everyday life. The variety of populations and contexts included in the literature thus far supports the idea that wearable sensors are a broadly applicable tool with flexible utility that may offer benefit to most populations.

This is an initial scoping review to determine the body of existing knowledge on this topic. The limitations of this review are that there are only a small number of RCTs; several studies were only performed in laboratory conditions; we did not limit the review to just anxiety (the inclusion of stress); and the sensors and conditions of use (eg, time, how they were worn) varied considerably. This indicates that there is not yet sufficient evidence to show that these interventions are feasible in real-world situations, although there appears to be sufficient potential for further investigation. With the growing use of consumer-wearable devices it may be possible in the future to use population-based, real-world data in this field, but there is much to be worked out before this can proceed in relation to ethics, privacy, and ownership of data.

There is also substantial further research required on the digital interventions that may be acceptable and effective in the moment for different cohorts, for example, the question of whether the sensor data should be made visible to the individual at the time or just the intervention. There is a risk that the availability of such continuous data could heighten anxiety in some people or could become the focus of their anxiety. It is also not clear whether an intervention based around identifying the physiological signs of anxiety would need to be long-term or could potentially *train* an individual to be more aware of their own symptoms earlier.

The acceptability and usability of the sensors is going to play a key role in their success, as was highlighted by Millings et al [30]. Technical issues can result in frustration and may not only result in disengagement with the intervention but also contribute to an individual's anxiety. Although there is potential for more work in this area, the authors are aware that the research and development of such interventions should be co-designed with potential end users. Without end-user involvement, interventions could miss the mark in terms of developing effective and acceptable digital health programs.

Acknowledgments

This review was conducted as part of a project funded by the Ministry of Business, Innovation & Employment (MBIE) National Science Challenge 'A Better Start' and Cure Kids project grant.

Authors' Contributions

RD, RW, TT, and JM designed the review. LLL and RD screened articles; LLL and KG performed the data extraction; and LLL, KG, and RD drafted the manuscript with input from all authors. Decisions regarding eligibility for inclusion and data extraction were verified independently by RW. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 109 KB - mental_v10i1e42611_app1.docx]

References

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
2. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
3. Al Omari O, Al Sabei S, Al Rawajfah O, Abu Sharour L, Aljohani K, Alomari K, et al. Prevalence and predictors of depression, anxiety, and stress among youth at the time of COVID-19: an online cross-sectional multicountry study. Depress Res Treat 2020;2020:8887727. [doi: 10.1155/2020/8887727] [Medline: 33062331]
4. Every-Palmer S, Jenkins M, Gendall P, Hoek J, Beaglehole B, Bell C, et al. Psychological distress, anxiety, family violence, suicidality, and wellbeing in New Zealand during the COVID-19 lockdown: A cross-sectional study. PLoS One 2020;15(11):e0241658 [FREE Full text] [doi: 10.1371/journal.pone.0241658] [Medline: 33147259]
5. Singhal A, Ross J, Seminog O, Hawton K, Goldacre MJ. Risk of self-harm and suicide in people with specific psychiatric and physical disorders: comparisons between disorders using English national record linkage. J R Soc Med 2014 May;107(5):194-204 [FREE Full text] [doi: 10.1177/0141076814522033] [Medline: 24526464]
6. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov 09;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]
7. The economic cost of serious mental illness and comorbidities in Australia and New Zealand. The Royal Australian and New Zealand College of Psychiatrists. 2016. URL: https://www.ranzcp.org/files/resources/reports/ranzcp-serious-mental-illness.aspx [accessed 2023-01-24]
8. Young AS, Klap R, Sherbourne CD, Wells KB. The quality of care for depressive and anxiety disorders in the United States. Arch Gen Psychiatry 2001 Jan;58(1):55-61. [doi: 10.1001/archpsyc.58.1.55] [Medline: 11146758]
9. Denny S, Farrant B, Cosgriff J, Harte M, Cameron T, Johnson R, et al. Forgone health care among secondary school students in New Zealand. J Prim Health Care 2013 Mar 01;5(1):11-18. [Medline: 23457690]
10. Mariu KR, Merry SN, Robinson EM, Watson PD. Seeking professional help for mental health problems, among New Zealand secondary school students. Clin Child Psychol Psychiatry 2012 Apr;17(2):284-297. [doi: 10.1177/1359104511404176] [Medline: 21852315]
11. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
12. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021 Feb;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
13. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
14. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
15. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
16. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443 [FREE Full text] [doi: 10.2196/22443] [Medline: 33064083]
17. Alneyadi M, Drissi N, Almeqbaali M, Ouhbi S. Biofeedback-based connected mental health interventions for anxiety: systematic literature review. JMIR Mhealth Uhealth 2021 Apr 22;9(4):e26038 [FREE Full text] [doi: 10.2196/26038] [Medline: 33792548]
18. Fukazawa Y, Ito T, Okimura T, Yamashita Y, Maeda T, Ota J. Predicting anxiety state using smartphone-based passive sensing. J Biomed Inform 2019 May;93:103151 [FREE Full text] [doi: 10.1016/j.jbi.2019.103151] [Medline: 30880254]
19. Hunkin H, King D, Zajac I. Wearable devices as adjuncts in the treatment of anxiety‐related symptoms: a narrative review of five device modalities and implications for clinical practice. Clin Psychol Sci Pract 2019 Sep;26(3):A. [doi: 10.1111/cpsp.12290]
20. Knight A, Bidargaddi N. Commonly available activity tracker apps and wearables as a mental health outcome indicator: a prospective observational cohort study among young adults with psychological distress. J Affect Disord 2018 Aug 15;236:31-36. [doi: 10.1016/j.jad.2018.04.099] [Medline: 29709718]

21. Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. Front Psychiatry 2021;12:625247 [FREE Full text] [doi: 10.3389/fpsyt.2021.625247] [Medline: 33584388]
22. Ryan J, Edney S, Maher C. Anxious or empowered? A cross-sectional study exploring how wearable activity trackers make their owners feel. BMC Psychol 2019 Jul 03;7(1):42 [FREE Full text] [doi: 10.1186/s40359-019-0315-y] [Medline: 31269972]
23. Hyde J, Ryan KM, Waters AM. Psychophysiological markers of fear and anxiety. Curr Psychiatry Rep 2019 Jun 04;21(7):56. [doi: 10.1007/s11920-019-1036-x] [Medline: 31161250]
24. Chesnut M, Harati S, Paredes P, Khan Y, Foudeh A, Kim J, et al. Stress markers for mental states and biotypes of depression and anxiety: a scoping review and preliminary illustrative analysis. Chronic Stress (Thousand Oaks) 2021;5:24705470211000338 [FREE Full text] [doi: 10.1177/24705470211000338] [Medline: 33997582]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Chung AH, Gevirtz RN, Gharbo RS, Thiam MA, Ginsberg JPJ. Pilot study on reducing symptoms of anxiety with a heart rate variability biofeedback wearable and remote stress management coach. Appl Psychophysiol Biofeedback 2021 Dec;46(4):347-358 [FREE Full text] [doi: 10.1007/s10484-021-09519-x] [Medline: 34308526]
27. Crivelli D, Fronda G, Venturella I, Balconi M. Stress and neurocognitive efficiency in managerial contexts. Int J Workplace Health Manage 2019 Apr 01;12(2):42-56. [doi: 10.1108/ijwhm-07-2018-0095]
28. Jaramillo-Quintanar D, Cruz-Albarran IA, Guzman-Sandoval VM, Morales-Hernandez LA. Smart sensor based on biofeedback to measure child relaxation in out-of-home care. Sensors (Basel) 2020 Jul 28;20(15):4194 [FREE Full text] [doi: 10.3390/s20154194] [Medline: 32731523]
29. Kizakevich PN, Eckhoff RP, Lewis GF, Davila MI, Hourani LL, Watkins R, et al. Biofeedback-assisted resilience training for traumatic and operational stress: preliminary analysis of a self-delivered digital health methodology. JMIR Mhealth Uhealth 2019 Sep 06;7(9):e12590 [FREE Full text] [doi: 10.2196/12590] [Medline: 31493325]
30. Millings A, Morris J, Rowe A, Easton S, Martin J, Majoe D, et al. Can the effectiveness of an online stress management program be augmented by wearable sensor technology? Internet Interventions 2015 Sep;2(3):330-339. [doi: 10.1016/j.invent.2015.04.005]
31. Nguyen J, Cardy RE, Anagnostou E, Brian J, Kushki A. Examining the effect of a wearable, anxiety detection technology on improving the awareness of anxiety signs in autism spectrum disorder: a pilot randomized controlled trial. Mol Autism 2021 Nov 14;12(1):72 [FREE Full text] [doi: 10.1186/s13229-021-00477-z] [Medline: 34775994]
32. Ponzo S, Morelli D, Kawadler JM, Hemmings NR, Bird G, Plans D. Efficacy of the digital therapeutic mobile app BioBase to reduce stress and improve mental well-being among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 06;8(4):e17767 [FREE Full text] [doi: 10.2196/17767] [Medline: 31926063]
33. Shruthi S, Priya D, Ranjith RB, Saronjini V, Ezhilarasi P, Ranjesh KS. Anxiety controlling wrist band. Smart Intelligent Computing Commun Technol 2021;38:145. [doi: 10.3233/apc210027]
34. Serino S, Cipresso P, Gaggioli A, Pallavicini F, Cipresso S, Campanaro D, et al. Smartphone for self-management of psychological stress: a preliminary evaluation of positive technology app. Revista Psicopatología Psicología Clínica 2014;19(3):253-260. [doi: 10.5944/rppc.vol.19.num.3.2014.13906]
35. Smith E, Santoro E, Moraveji N, Susi M, Crum A. Integrating wearables in stress management interventions: promising evidence from a randomized trial. Int J Stress Manage 2020 May;27(2):172-182. [doi: 10.1037/str0000137]
36. Winslow BD, Chadderdon GL, Dechmerowski SJ, Jones DL, Kalkstein S, Greene JL, et al. Development and clinical evaluation of an mHealth application for stress management. Front Psychiatry 2016;7:130 [FREE Full text] [doi: 10.3389/fpsyt.2016.00130] [Medline: 27507949]

Abbreviations

ASD: autism spectrum disorder
CBT: cognitive behavioral therapy
EDA: electrodermal activity
EEG: electroencephalogram
HR: heart rate
HRV: heart rate variability
PPG: pulse photoplethysmography
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial

Edited by J Torous, G Eysenbach; submitted 11.09.22; peer-reviewed by N Leonard, R Skews; comments to author 15.11.22; revised version received 20.12.22; accepted 12.01.23; published 02.02.23.

Please cite as:
Dobson R, Li LL, Garner K, Tane T, McCool J, Whittaker R
The Use of Sensors to Detect Anxiety for In-the-Moment Intervention: Scoping Review
JMIR Ment Health 2023;10:e42611
URL: https://mental.jmir.org/2023/1/e42611
doi:10.2196/42611
PMID:36729590

Review

Identifying and Categorizing Adverse Events in Trials of Digital Mental Health Interventions: Narrative Scoping Review of Trials in the International Standard Randomized Controlled Trial Number Registry

Aislinn D Gómez Bergin[1,2,3], PhD; Althea Z Valentine[1,3], PhD; Stefan Rennick-Egglestone[4], PhD; Mike Slade[4,5], PhD; Chris Hollis[1,2,3], PhD; Charlotte L Hall[1,2,3], PhD

[1]National Institute for Health and Care Research MindTech MedTech Co-operative, Institute of Mental Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[2]National Institute for Health and Care Research Nottingham Biomedical Research Centre, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[3]Mental Health and Clinical Neurosciences, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[4]School of Health Sciences, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[5]Faculty of Medicine and Health Sciences, Nord University, Namsos, Norway

Corresponding Author:
Charlotte L Hall, PhD
National Institute for Health and Care Research MindTech MedTech Co-operative
Institute of Mental Health, School of Medicine
University of Nottingham
University Park
Nottingham, NG7 2TU
United Kingdom
Phone: 44 0115 823 2438
Email: charlotte.hall@nottingham.ac.uk

Abstract

Background: To contextualize the benefits of an intervention, it is important that adverse events (AEs) are reported. This is potentially difficult in trials of digital mental health interventions, where delivery may be remote and the mechanisms of actions less understood.

Objective: We aimed to explore the reporting of AEs in randomized controlled trials of digital mental health interventions.

Methods: The International Standard Randomized Controlled Trial Number database was searched for trials registered before May 2022. Using advanced search filters, we identified 2546 trials in the category of mental and behavioral disorders. These trials were independently reviewed by 2 researchers against the eligibility criteria. Trials were included where digital mental health interventions for participants with a mental health disorder were evaluated through a completed randomized controlled trial (protocol and primary results publication published). Published protocols and primary results publications were then retrieved. Data were extracted independently by 3 researchers, with discussion to reach consensus when required.

Results: Twenty-three trials met the eligibility criteria, of which 16 (69%) included a statement on AEs within a publication, but only 6 (26%) reported AEs within their primary results publication. Seriousness was referred to by 6 trials, relatedness by 4, and expectedness by 2. More interventions delivered with human support (9/11, 82%) than those with only remote or no support (6/12, 50%) included a statement on AEs, but they did not report more AEs. Several reasons for participant dropout were identified by trials that did not report AEs, of which some were identifiable or related to AEs, including serious AEs.

Conclusions: There is significant variation in the reporting of AEs in trials of digital mental health interventions. This variation may reflect limited reporting processes and difficulty recognizing AEs related to digital mental health interventions. There is a need to develop guidelines specifically for these trials to improve future reporting.

(JMIR Ment Health 2023;10:e42501) doi:10.2196/42501

KEYWORDS

adverse events; harm; psychological interventions; clinical trials; review; digital; mobile phone

Introduction

Background

Digital mental health interventions provide information, support, and therapy for mental health problems via a technology or a digital platform [1]. These interventions come in various formats and may be delivered via an app, smartphone, computer, or tablet. These interventions may be self-guided, with little or no human interaction, thus offering the potential to treat large numbers of patients across geographically disperse and remote locations. An automated approach can reduce staff time and costs, allowing more control over the content and delivery of a mental health intervention. However, the increasing popularity and unique delivery of digital mental health interventions pose new challenges for the monitoring and reporting of harms in clinical trials. Several of these have been highlighted in a previous consensus statement [2]. Specifically, risks may arise in the absence of staff being present to support emotional processing, explain treatment content, or guide patients in how to engage with the technology, leading to a greater risk of misunderstanding or inadequate delivery of the intervention. Given the potential differences that need to be considered when delivering mental health interventions digitally, including the potential lack of direct human contact, it may be particularly difficult for researchers to understand how best to recognize, record, and categorize adverse events (AEs) [2]. This is in addition to the existing challenges recognized by mental health researchers in AE reporting, regardless of whether the intervention is digital. Reporting of AEs in clinical trials should conform to the Consolidated Standards Of Reporting Trials (CONSORT) harms statement [3]. There is also a specific CONSORT-eHealth for reporting trials of digital health interventions, which reflects the importance of recording harms [4]. To fully report AEs, researchers must have in place appropriate definitions for how to identify and measure AEs and how to categorize them to determine seriousness, expectedness, and relatedness to the intervention. The Good Clinical Practice guidelines of the International Council for Harmonization define an AE as an "untoward medical occurrence;" a serious AE (SAE) as "death/life threatening, inpatient hospitalization, disability, birth defect," or a "medically important event;" and "unexpected" when a reaction is not consistent with the outcomes expected [5]. However, these guidelines were developed for drug-related incidences. Applying guidance from drug trials does not always fit for mental health interventions, with risks often not relevant to the intervention, leading researchers to call for a broader and more appropriate interpretation to include areas such as additional treatment needs and life changes [6-8]. Defining what constitutes an AE and determining its relationship to the intervention usually requires a plausible sequence of events, whereby the intervention is directly linked to the AE. This is easier to define with pharmacological interventions, where the mechanisms of action are better understood [9,10].

Recent reviews of protocols for trials of mental health interventions found significant variations in how AEs were defined [7,8]. Considering the lack of consensus regarding AEs within mental health intervention research and the reliance of reporting guidelines such as CONSORT on trial teams to predefine AEs, this is concerning as it suggests the potential for underreporting. A 2014 review [11] found only 1 mental health intervention trial had been terminated on the grounds of unacceptable AEs, with very few considering harms relevant to psychological therapies such as distress or self-harm and no primary results publication reporting AEs. A review of protocols suggests an increase in the plans to record AEs [8], but it is unclear if this has translated to an increase in the actual reporting. AE reporting can be time-intensive for researchers and is arguably more straightforward when an intervention is delivered in-person, wherein AEs can be spontaneously reported or identified. However, digital mental health interventions often have reduced or no direct contact and there may be additional "unwanted events" [2] related to the digital nature of delivery (eg, technical issues). Research has previously indicated that monitoring reasons for withdrawal can highlight potential AEs, suggesting that this may be one way to identify unrecognized AEs [12]. The monitoring and reporting of AEs in trials for digital mental health interventions is a necessity to fully understand the risk-benefit ratio of such interventions. The current lack of clear guidelines for AEs precludes this from happening.

Objectives of This Review

This scoping review aims to explore how AEs are currently assessed in trials of digital mental health interventions according to their protocols and primary results publications. Further objectives of this review were to (1) identify how AEs are reported; (2) identify how AEs are defined and classified with regard to their seriousness, expectedness, and relatedness; (3) explore how AEs are monitored; (4) explore potential harms identified within reasons given for dropout and to better understand how researchers might employ certain strategies to mitigate potential risk, for example, through excluding those most at-risk; and (5) explore whether there was a link between what research teams set out to measure in their protocol and what they reported in their final report. This review focused on trials of digital mental health interventions; therefore, it was outside the scope of this study to review variations in trials that adopt different web-based or digital methods of data collection. Furthermore, this study did not aim to compare with trials of nondigital interventions.

Methods

Study Design

We hosted our protocol in the public domain [13]. We made 2 deviations from this protocol; we chose to limit only to published protocols, as this best fitted with understanding reporting trial procedures, and we chose to expand our review of AEs to look at the categorization of expectedness as well as

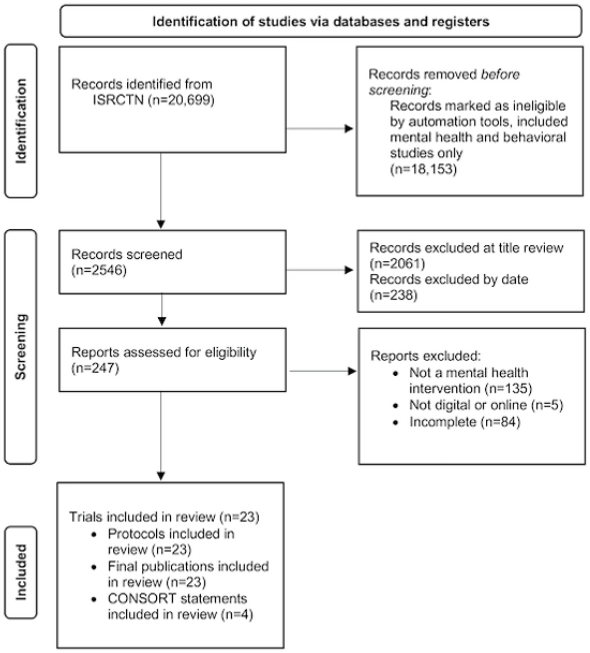
the monitoring of AEs and potential AEs to provide a more complete picture. This review was a scoping review [14] conducted in 1 database with systematic screening of records for eligibility.

Searches

The International Standard Randomized Controlled Trial Number (ISRCTN) registry was searched to identify relevant trials (May 2021). This registry was selected as the database to search, as it is the primary clinical trial registry recognized by the World Health Organization and the International Committee of Medical Journal Editors, and it provides a source for identifying clinical research studies across multiple sites, funders, settings, and conditions. The advanced search facility was used, wherein filters included trials only in the category of mental and behavioral disorders (see Figure 1 [15]). As such, there were no search terms, but all records that were brought up using this filter were searched.

Titles and ISRCTN abstracts were read by 2 authors (CLH and ADGB). Where it was not certain whether studies met the inclusion criteria, the plain English summary was additionally read to ascertain relevance. Reasons for exclusion were noted. A list of eligible trials was generated, and attempts were made to access (from the corresponding author and online) the published protocol and the main outcome paper(s) (ie, the paper(s) that present data from the primary outcome). The search was initially conducted on May 17, 2021, and updated on May 23, 2022, by identifying any trial that had been marked as incomplete during the previous searches and by checking for primary results publications.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram. CONSORT: Consolidated Standards Of Reporting Trials; ISRCTN: International Standard Randomized Controlled Trial Number.



Inclusion and Exclusion Criteria

Searches were restricted from 2004 (publication year of harm-reporting extension of the CONSORT statement) to May 17, 2021. In line with previous research, exclusions were made where there are additional physical health needs that may make AE reporting more common [12]. Interventions that act as an adjunct, where most of the intervention is delivered by a clinician, were excluded. Only complete trials were included, defined as those with both a published peer-reviewed protocol and primary results publication. An eligibility check on the full protocol(s) and paper(s) was conducted by 3 researchers independently with 99.2% agreement (ADGB, AZV, CLH). Any disagreement or discrepancy was resolved through discussion by all 3 where necessary. See Table 1 for the complete list of inclusion and exclusion criteria.

Table 1. Inclusion and exclusion criteria.

Criteria	Inclusion criteria	Exclusion criteria
Participants, population	People (any age and any gender) who fulfilled the diagnostic criteria for a mental health disorder according to the following International Classification of Diseases, tenth revision criteria. F20-29: schizophrenia, schizotypal and delusional disorders; F30-39: mood (affective) disorders, anxiety disorders (including obsessive-compulsive disorder); F40-48: neurotic, stress-related, and somatoform disorders; F50: eating disorders; F60-62: specific personality disorders (but excluding F63-69; F63: habit and impulse disorders, F64-69: gender-related disorders); F80-89: disorders of psychological development; F90-98: behavioral and emotional disorders with onset usually occurring in childhood and adolescence.	Not mental health patient, that is, research looks at health care providers, or parents/carers. People (any age and any gender) who fulfilled the diagnostic criteria for a mental health disorder according to the following International Classification of Diseases tenth revision criteria. F00-09 relate to organic mental disorders (eg, dementia, traumatic brain injury), F10-19: mental and behavioral disorders due to psychoactive substance use, F51-59: behavioral syndromes associated with physiological disturbances and physical factors (including sleep disorders), F70-79: mental retardation, F99: unspecified mental disorder.
Interventions, exposures	Any randomized controlled trial (RCT) (including definitive, pilot, feasibility, and exploratory RCTs) of psychological interventions (ie, non-Clinical Trial of an Investigational Medicinal Product Trial, including behavioral, cognitive, and psychosocial interventions) delivered online or digitally (ie, websites, apps, synchronous and asynchronous delivery, including both with and without human support) with a published protocol and final results.	No psychological therapeutic intervention, excludes activity-based interventions, interventions focused on vocational outcomes, and interventions using herbal/alternative medicines or pharmaceutical drug trials. Also excludes preventative trials on populations at risk. Excludes assessment, adjunct use, and medication management trials, or studies. Nonrandomized trials, observational studies, and expanded access studies. Studies published prior to 2004.
Comparators, control, context	Any control group and any setting will be considered, including health services, community or school settings, prison/forensic facilities. Research in all countries will be included.	N/A[a]
Outcomes	All outcomes related to adverse events were included: reporting, definitions, seriousness, expectedness, relatedness—along with reasons for dropout.	Studies recorded as “not yet recruiting” or “ongoing.”

[a]N/A: not applicable.

Data Extraction (Selection and Coding)

Relevant data were extracted independently by 2 researchers (ADGB and CLH) from all published trial documents, including the protocol, primary results publication, and any supplementary materials, including CONSORT checklists. Study quality was not assessed for risk of bias, as the purpose of this review was to identify information relevant to AEs rather than treatment outcomes. Categories of data for extraction are included in Multimedia Appendix 1.

Results

Overview

Our search is reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [15]. It identified 2546 unique citations (see Figure 1). We excluded 238 that did not meet the date criteria, 2061 after screening titles, and 224 after screening full text because they did not meet the inclusion criteria. Finally, 23 trials were identified (see Multimedia Appendix 2 for references) [16-38]. Their characteristics, including details on participant condition and intervention, are shown in Multimedia Appendix 3. Most interventions were delivered using the web/internet via either an app or through a web browser on a computer, laptop, tablet, or smartphone [17-24,26-34,36-38]. Three were delivered using virtual reality equipment [16,25,35]. Sixteen trials recruited adults older than 18 years [16-22,24,25,27-29,33,34,36,37], with a further 4 trials including those older than 16 years [26,31,35,38], and 3 trials with those younger than 18 years only [23,30,32].

Studies With AEs

Multimedia Appendix 4 outlines how AEs were reported, defined, monitored, and how seriousness, expectedness, and relatedness were addressed in the trials. Of the 23 trials, 7 did not mention AEs within any trial documentation [17-19,22,25,34,36]. Of the remaining 16 trials, 10 mentioned AEs in both the trial protocol and primary results publication [16,21,27-31,33,35,38], with 2 additionally mentioning AEs within a CONSORT checklist [31] and in the ISRCTN database [30]. One trial only referred to AEs within their primary results publication [32], 1 only within their protocol [23], and 3 only within a supplementary CONSORT checklist provided alongside their primary results publication [20,24,37]. The remaining trial mentioned AEs within their protocol and a CONSORT checklist [26]. Seven studies that included younger adults and children all mentioned AEs [23,26,30-32,35,38].

Reporting of AEs

Fifteen trials reported on the presence/absence of AEs within their primary results publication [16,20,21,24,26-33,35,37,38]. Seven of these trials reported no AEs [16,24,26,31-33,38]. Two trials reported that AEs were not applicable [20,37]. Four trials reported both AEs and SAEs [21,27,30,35] and 2 trials only reported AEs [28,29].

How AEs Were Defined

A definition of what classified as an AE was provided within the protocol of 8 trials, with 1 further trial doing so within their primary results publication [21,23,27-31,33,35]. However, there was a considerable variation. Three trials used a symptom checklist to define an AE, one of which did not then report AEs within their primary results publication [23,29,30]. Five trials identified fatal or life-threatening events but differed in the additional definitions [21,23,27,31,35]. One identified only in their primary results publication how AEs were defined, as deterioration in the primary outcome measure [28]. Two trials directly considered AEs specific to the device [31,32].

Seriousness of AEs

Six of the 14 trials that mentioned AEs referred to the seriousness of AEs, although one did not provide information as to how seriousness would be determined—only that it would be rated by severity [21,29-31,35,38]. Five provided a definition of seriousness, and of these, 4 stated that the event would be reviewed by a senior member of the research team. SAEs were defined as fatal (n=4), life-threatening (n=4), hospitalization (n=4), disability (n=3), congenital (n=2), and an event identified by a senior member of the trial team as serious (n=3).

Expectedness

One trial reported expected and unexpected AEs separately in their primary results publication [30]. An additional trial mentioned in their protocol that the investigators would determine if an AE was expected or unexplained; however, these were not reported in the primary results publication [27]. Although not explicitly referring to categorizing the event as it occurs, a further trial stated they would only monitor AEs inherent to the nature of the condition, which implies some form of expectedness [23]. One further trial stated they did not expect AEs [33].

Monitoring of AEs

Eight trials provided information on the monitoring and, in some cases, reviewing of AEs [21,23,27,29-31,35,38]. In 1 trial, AEs were monitored by a mental health worker in each session and throughout by reporting to a therapist, outcome assessor, or by self-report. One trial reviewed medical notes at the end of the trial [35]. Three used checklists [23,29,30], including the Short Mood and Feelings Questionnaire in every session [39], Generalized Anxiety Disorder [40] and Patient Health Questionnaire [41] that were provided frequently, a modified symptom checklist postintervention [42], and a modified side-effects scale used throughout and for 3 months after the intervention finished [43]. Potential AE/SAEs were also reported to and reviewed by senior team members (eg, the chief investigator) or external committees (n=8) [21,23,27,29-31,35,38]. One indicated that they were monitoring AEs using existing guidance [27] (Standard Protocol Items: Recommendations for Interventional Trials) [44]. One study indicated that it would be monitored through self-report [33].

Relatedness of AEs

Five trials addressed how related an event was to the intervention or trial procedures [27,29-31,35]. The process for identifying relatedness involved an assessment of the event by the research team in consideration of the participant's individual situation (eg, their condition, treatment), the temporal relatedness to the trial, discussion with parents/carers/participants, and an assessment by independent committees or trial investigators.

Potential AEs

To help us identify whether harms that may lead to AEs are not being recognized, recorded, or reported, we further scrutinized the data to highlight those areas identified in previous research as potentially related to AEs. This information is summarized in Multimedia Appendix 5 (studies that did not reference or report AEs, n=17) and Multimedia Appendix 6 (studies that reported AEs, n=6), in terms of support and training, dropouts, strategies to manage risk and eligibility criteria, alongside whether the trial identified any AEs.

Support

Providing support alongside an intervention offers more opportunities for contact with participants, potentially increasing the reporting of AEs. Direct support was defined as support provided in-person, for example, with someone delivering the intervention or existing services providing assistance. Brief support included support provided in-person but only to demonstrate or introduce the intervention. Remote support included delivery of support via digital methods, for example, by email. Eleven trials provided remote support [18-20,22,26,28,30,33,34,36,37], 7 trials provided direct support [16,23,25,27,31,32,35], and 4 trials provided brief support [17,21,24,38]. One did not provide support [29]. Trials that offered direct (6/7, 86%) and brief support (3/4, 75%) mentioned AEs more than those with remote support (5/11, 45%) although they did not report experiencing more AEs within the primary results publication. The trial that did not include support mentioned AEs [29].

Dropouts and Noncompleters

Five trials did not report on reasons for dropout and did not mention AEs [17-19,22,36]. One trial that stated there had been no AEs within the study [24] and another that indicated that AEs were not applicable [37] did not report on dropouts. Ten trials reported on reasons for dropout and did not mention or report AEs [16,20,23,25,26,31-34,38]. One trial reported on both AEs and dropouts [28]. Reasons for dropout are listed in Multimedia Appendices 5 and 6 with several identifiable as an AE/SAE within the definitions identified in our sample. Two trials that did not mention AEs reported significantly higher dropouts from the intervention arm versus those from a waitlist control [45,46]. Two that did mention AEs but stated that they were not applicable also had higher dropouts from the intervention arm compared to that from the website that was offered as a control [47,48]. One that also stated there were no AEs had a similar pattern of high dropouts in the intervention arm compared to that in the waitlist control [49].

Risk Mitigation Strategies

We looked at possible strategies that may have been used by studies not reporting AEs to manage risks in patients during the delivery of the intervention. Possible strategies identified

included reviewing all communication [18], preventing participants discussing certain topics (eg, suicide) [20], a user contract and system for flagging those considered higher need [22], asking clinicians to monitor patients [24], visible signposting within the intervention for support [34], and a panic button while in virtual reality [25]. Two virtual reality trials also included a cybersickness measure not used for AE reporting [16,25]. Eight trials excluded individuals with physical conditions that could impact their use of technologies (eg, visual problems) [16,21,24,25,27,29,32,35] or those with limited access/experience in technologies (n=9) [17, 18, 26, 28, 30, 31, 33, 34, 36]. Other common exclusion criteria included more serious and enduring illness (n=13) [18-25, 27-30, 35], substance misuse (n=7) [18, 21, 27, 28, 30, 35, 37], and suicidality (n=10) [18,19,21,22,24,28,30,35-37]. Thirteen trials screened participants for their language ability [16,17,24-26,30-32,34-38].

Discussion

To understand the reporting of AEs in trials of digital mental health interventions, we reviewed protocols and primary results publications of the completed trials listed in the ISRCTN registry, an international clinical trial database. Our findings indicated that although two-thirds of the trials mentioned AEs, only 26% (6/23) reported experiencing an AE or an SAE in their primary results publication. Additionally, there was a lack of reporting on the categorization of the seriousness, expectedness, and relatedness of AEs. This review has highlighted several challenges associated with recognizing and monitoring AEs in trials of digital mental health interventions.

Our findings indicate new challenges in monitoring AEs in trials of digital mental health interventions. How these interventions are delivered (eg, with no direct human contact) could lead to fewer AEs being identified within trials [2], but our findings suggested that although trials of interventions that offer direct or brief support are more likely to mention AEs in their protocol, they are not more likely to record their presence in the primary results publication. The aim of this study was to review the reporting of AEs within trials of digital mental health interventions; however, it also highlights the challenges within digital trial processes. Rather than the delivery of the intervention, it may be that certain aspects of the trial (eg, online-only data collection) make it particularly difficult to assess or record negative effects and thus lead to fewer recorded AEs. However, this was outside the scope of this review.

Our findings seemed to suggest that some populations are more likely to be monitored for AEs (eg, when they are younger). We are mindful of the additional burden that monitoring AEs can place on both participants and trial teams, but this must be carefully balanced with the need for methodological improvements. Our review suggested that the functionality of digital interventions could be utilized to support more standardized methods of AE reporting and recording through automating and streamlining the reporting of AEs for participants. Automated processes could be used to collect data on harms or flag for risk of harm (eg, if participant scores meet a predefined cutoff). The development of relevant symptom checklists deliverable in a web-based format would be an important step within this field. Some work has already started on the specific checklists for psychological interventions [50]; however, to the best of our knowledge, this has not been done digitally or online.

Alongside improved monitoring, it is also important to ensure that AEs within trials of digital mental health interventions are recognized and classified appropriately [2,4,8]. Our findings indicate difficulties in classifying AEs in digital mental health, which are likely to reflect the lack of relevant guidelines. As with other reviews, we identified several reasons for dropout indicative of AEs or SAEs, but these were not categorized as such by the studies. Additionally, several studies had higher dropouts associated with the digital intervention [45-49], indicating that identifying reasons for dropout may highlight unseen or unrecognized AEs or SAEs. Indeed, a trial excluded from this review due to a focus on relapse prevention found that 2 of those who withdrew had experienced an app-related AE [51]. This study also demonstrated the potential significance of AEs related to the technology itself, with 13 AEs reported as app-related, one of which was classified as serious. Most use the standard definition from the Good Clinical Practice guidelines of the International Council for Harmonization, and our findings suggest that important AEs linked to the technology may not be recognized, for example, cybersickness. Difficulties with, restricted, or no access to a digital intervention during a trial may mean that individuals experience heightened distress, misunderstand or misuse the intervention, or feel excluded. One paper in our review highlighted this as a potential harm where the automated eligibility check excluded those with more severe depression and contributed to feelings of “disappointment, frustration and a sense of exclusion” [52]. It is possible that researchers employ certain strategies to mitigate potential harms. Some examples from this review relate to the digital nature of the intervention, for example, excluding those with limited technical ability, while others relate to mental health, for example, providing human contact or support through which risk can be assessed. Research conducted by Papaioannou and colleagues [8] suggests that feelings of failure if the mental health intervention is not successful may be an unintended consequence definable as an AE within psychological therapies. Other researchers have called for a broader definition that includes, for example, deterioration of symptoms [7] and noted several potential AEs directly related to the remote delivery of interventions, including misuse [2].

Only 4 trials referred to how related an AE was to the intervention. The causal pathways in psychological interventions are less clear-cut than those in pharmacological interventions. A “dark logic” model has previously been proposed, whereby researchers set out a priori an intervention’s mechanism of action, which can be examined for where potential harms may occur [53]. The use of Trial Steering Committees to determine relatedness can provide some independence in this process. Our review suggests that AEs may not be that common (as indicated by only 6 of the 23 trials reporting AEs) and could be reviewed by a committee. However, it is perhaps more efficient to focus on SAEs or predefine events to be reviewed and to establish a blinded end point review committee to evaluate any bias in the categorization of relatedness. Furthermore, from our own

experiences conducting trials, reviewing temporal relationships between the intervention and the onset of the AE, whether the AE symptom had ever occurred before participation and asking the opinion of the participant were important in determining relatedness. However, these necessitate that direct contact is possible with participants. This can be done through the support provided, through the intervention itself, or embedded within trial processes.

Although preventative trials were outside the eligibility criteria of this review, it is likely by their very nature, they may have superior AE monitoring (ie, their primary aim is likely to be related to AEs) and thus, they may provide an opportunity for shared learning. For example, a feasibility randomized controlled trial that investigated a medical device (EMPOWER) to prevent relapse in patients with schizophrenia was particularly impressive in the AE monitoring [51]. EMPOWER included a smartphone for participants to actively monitor their symptoms, combined with peer support and clinical triage where needed. AEs were identified via contact with an unblinded peer support worker who monitored the participants during the trial via fortnightly telephone contact, and AEs were categorized according to their severity, intensity, and relatedness and included those that were app-related. By monitoring AEs routinely, the team was able to respond to the service users' arising needs. That study illustrates the potential advantage of blended (human/digital) interventions to facilitate AE reporting. However, for resource reasons or due to the nature of the intervention, this may not always be possible. It also may be particularly challenging to do blinded, when researchers are unaware of which treatment is being received. Of interest, this intervention was also classified as a medical device, and as such, adequate AE or SAE reporting is mandated in order to satisfy regulatory approval. Two included trials are also known to us to have interventions classified as medical devices, both of which demonstrated significantly more thorough AE monitoring and reporting [27,35]. It is possible that as guidelines on what classifies as a medical device (including software as a medical device) continue to evolve, there will be greater emphasis and requirement for more research teams to be appropriately monitoring AEs or SAEs.

Unlike other studies, we believe our research is strengthened by the inclusion of both protocols and primary results publications and is the first to explore AE reporting in trials of digital mental health interventions. This allowed us to compare between what research teams set out to measure and what they reported in the final papers. However, there would be merit in a future review including all studies, regardless of whether there was an available protocol and primary results publication. A limitation may also be that our search was constrained only to trials registered on ISRCTN with a published protocol and primary results publication, excluding other registries such as ClinicalTrials.gov. It is likely that registered trials with a published protocol are conducted with higher rigor than nonregistered trials; so, our findings may overrepresent the recording of AEs. Future reviews could explore if higher quality trials were more likely to record AEs. We did not explore whether teams had recorded processes related to AEs in additional standard operating procedures, as CONSORT defines these processes should be reported with the trial findings. It was also outside of the scope of this review to explore potential AEs in detail, but the review did highlight several areas unique to digital mental health interventions that merit further research to better understand their potential for harm. We identified that human support did not appear to lead to the reporting of more AEs; however, further research should consider comparing both web-based and offline trial processes, as more contact with researchers may lead to more opportunities for identifying AEs. Further research could also explore the nature of AEs in psychological trials, which was not the aim of this review. Additionally, it would be useful to directly compare AE reporting in studies conducted online and offline.

Following Papaioannou and colleagues' [8] recommendations for psychological trials, we provide further suggestions for AE reporting in trials of digital mental health interventions. These should support the development of specific guidance that captures their unique nature in relation to the processes of remote data collection and technological delivery.

1. Improving monitoring: It is essential that AEs are systematically monitored in trials for digital mental health interventions in a meaningful way, especially where the interventions will subsequently be offered more widely. Monitoring AEs more closely will enable researchers to identify negative effects that may not be clear otherwise.
2. Improving reporting: Trial teams need to balance the additional patient and staff burden resulting from monitoring AEs with the need to ensure accurate reporting to ensure the validity of the trial to assess the efficacy and safety of the intervention. For instance, digital mental health interventions that are delivered remotely could have clear reporting options embedded (eg, a link within a website or email) to monitor for AEs. More systematic processes unique to digital technologies could be utilized, for example, through the automation of reaching a predefined score on a risk item (ie, suicidal ideation) or an outcome measure. We endorse the need for patient and public involvement to improve understanding the burden from the patient perspective.
3. Improving classification: Mechanisms of action of the intervention should as far as possible be defined a priori to understand potential causality/relatedness of AEs, as well as exploring temporal relationships to engagement in the intervention and prior presence of the event before participation. A logic model to describe the mechanism of action could be developed upfront and explored further during the trial through a process evaluation. A process evaluation is a mixed methods approach to understand the quality of the implementation of a complex intervention as well as to understand the dose and reach of the intervention, analyze causal mechanisms, and identify contextual factors [54]. They can be used to support the interpretation of trial findings.
4. Independent review: Experts within digital mental health or at least the health-related factors associated with a technology could be consulted to support AE classification and included in independent oversight committees. This will help to predefine or identify SAEs that are unique or

related to the digital mental health intervention. These should consider the type of technology.

5. Understanding dropout: Future research and implementation can be supported through the recording and reviewing of reasons for dropout to better understand the seriousness/severity from the perspective of participants and any potential AEs unique to the technology.
6. Anonymized data sharing: Anonymized data sharing initiatives may also help inform research teams on the types of AEs likely to occur in digital mental health trials.

In summary, although most published protocols of studies are giving recognition to the identification of AEs in trials of digital mental health interventions, there appears to be a gap in the reporting of AEs in their primary results publications, with even fewer studies reporting on the categorization of AEs in relation to seriousness, relatedness, or expectedness. This highlights a potential lack of knowledge in identifying and classifying AEs in digital mental health. This gap also suggests potential methodological difficulties with gathering AE data, which may be more complex in remotely conducted studies, although this is yet to be addressed within a review. There is a need to develop guidelines for AE reporting specifically for these trial designs to improve reporting practices, conform to CONSORT guidelines, and fully understand the cost-benefit of digital mental interventions.

Acknowledgments

The authors ADGB, CLH, and CH acknowledge the support of the UK Research and Innovation Digital Youth Program award (MRC project reference MR/W002450/1), which is part of the Arts and Humanities Research Council/Economic and Social Research Council/Medical Research Council Adolescence, Mental Health, and the Developing Mind program. SR-E received salary support from the National Institute for Health and Care Research (NIHR) under its program grants for Applied Research Program (Personal experience as a recovery resource in psychosis: Narrative Experiences Online Program [RP-PG-0615-20016]). We would also like to thank Angela Summerfield for her continued administrative support and help. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This research was funded by the Medical Research Council Digital Youth Program, NIHR Nottingham Biomedical Research Centre, and NIHR MindTech MedTech Co-operative. The views expressed are those of the authors and not necessarily those of the National Health Service, Medical Research Council, the NIHR, or the Department of Health. These organizations had no role in study design, including collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication. The corresponding author (CLH) had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Data Availability

All relevant data are included in this publication. Trials included in this review are identified in the tables and multimedia appendices. All excluded studies are available on request to the corresponding author. Researchers who require further details may contact the corresponding author with a methodologically sound proposal.

Authors' Contributions

CLH conceived this study. The study design was refined by ADGB and AZV. AZV conducted the search, and ADGB updated it. Data extraction was conducted by ADGB, CLH, and AZV. ADGB wrote the original manuscript, with input from CLH. SR-E, MS, and CH critically reviewed the manuscript. All authors contributed to the interpretation of the data and redrafts. All authors had full access to the study data and take responsibility for the integrity and accuracy of the data and accept responsibility for publication. CLH is a guarantor.

Conflicts of Interest

All authors have completed the International Committee of Medical Journal Editors uniform disclosure form. AZV, MS, and SR-E declare no conflict of interests. ADGB, CLH, and CH contributed to at least one study [35] included in this review but declare no other conflicts of interest.

Multimedia Appendix 1
Data extraction strategies.
[PDF File (Adobe PDF File), 43 KB - mental_v10i1e42501_app1.pdf]

Multimedia Appendix 2
References for the reviewed trials.
[PDF File (Adobe PDF File), 133 KB - mental_v10i1e42501_app2.pdf]

Multimedia Appendix 3

Trial characteristics.
[PDF File (Adobe PDF File), 134 KB - mental_v10i1e42501_app3.pdf]

Multimedia Appendix 4
Adverse event reporting in trials.
[PDF File (Adobe PDF File), 111 KB - mental_v10i1e42501_app4.pdf]

Multimedia Appendix 5
Potential adverse events identified in trials that made no reference to adverse events or reported no adverse events.
[PDF File (Adobe PDF File), 169 KB - mental_v10i1e42501_app5.pdf]

Multimedia Appendix 6
Potential adverse events identified in trials that reported adverse events and serious adverse events.
[PDF File (Adobe PDF File), 109 KB - mental_v10i1e42501_app6.pdf]

References

1. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
2. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of Internet interventions. Internet Interventions 2014 Mar;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]
3. Ioannidis JP, Evans SJW, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004 Nov 16;141(10):781-788 [FREE Full text] [doi: 10.7326/0003-4819-141-10-200411160-00009] [Medline: 15545678]
4. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
5. E2D post-approval safety data management: definitions and standards for expedited reporting. USFDA. URL: https://tinyurl.com/39j9ber2 [accessed 2023-02-16]
6. Horigian VE, Robbins MS, Dominguez R, Ucha J, Rosa CL. Principles for defining adverse events in behavioral intervention research: lessons from a family-focused adolescent drug abuse trial. Clin Trials 2010 Feb;7(1):58-68 [FREE Full text] [doi: 10.1177/1740774509356575] [Medline: 20156957]
7. Klatte R, Strauss B, Flückiger C, Färber F, Rosendahl J. Defining and assessing adverse events and harmful effects in psychotherapy study protocols: A systematic review. Psychotherapy (Chic). Preprint posted online on January 20, 2022 2022. [doi: 10.1037/pst0000359] [Medline: 35049321]
8. Papaioannou D, Cooper C, Mooney C, Glover R, Coates E. Adverse event recording failed to reflect potential harms: a review of trial protocols of behavioral, lifestyle and psychological therapy interventions. J Clin Epidemiol 2021 Aug;136:64-76 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.03.002] [Medline: 33684508]
9. Mayo-Wilson E, Fusco N, Li T, Hong H, Canner JK, Dickersin K, MUDS investigators. Harms are assessed inconsistently and reported inadequately part 1: systematic adverse events. J Clin Epidemiol 2019 Sep;113:20-27 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.04.022] [Medline: 31055175]
10. Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. BMJ Open 2019 Mar 01;9(2):e024537 [FREE Full text] [doi: 10.1136/bmjopen-2018-024537] [Medline: 30826796]
11. Duggan C, Parry G, McMurran M, Davidson K, Dennis J. The recording of adverse events from psychological treatments in clinical trials: evidence from a review of NIHR-funded trials. Trials 2014 Aug 27;15:335 [FREE Full text] [doi: 10.1186/1745-6215-15-335] [Medline: 25158932]
12. Bottema-Beutel K, Crowley S, Sandbank M, Woynaroski TG. Adverse event reporting in intervention research for young autistic children. Autism 2021 Feb;25(2):322-335 [FREE Full text] [doi: 10.1177/1362361320965331] [Medline: 33076682]
13. Bergin A, Valentine AZ, Slade M, et al. How are adverse events identified and categorised in trials of digital mental health interventions? A systematic review protocol. Mindtech. URL: https://tinyurl.com/5n6h5mde [accessed 2022-06-14]
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
16. Virtual reality exposure therapy in psychosis (ISRCTN12929657). ISRCTN Registry. 2014. URL: https://www.isrctn.com/ISRCTN12929657 [accessed 2022-11-08]

17. Bipolar interactive psychoEDucation trial (ISRCTN81375447). ISRCTN Registry. 2008. URL: https://www.isrctn.com/ISRCTN81375447 [accessed 2022-11-08]
18. Internet delivered treatment for symptoms of depression in the community (ISRCTN03704676). International Standard Randomised Trial Number Registry. 2013. URL: https://www.isrctn.com/ISRCTN03704676 [accessed 2022-11-08]
19. Online depression in diabetes study: web-based cognitive behavioural therapy for diabetic adults with depression (ISRCTN24874457). International Standard Randomised Trial Number Registry. 2006. URL: https://www.isrctn.com/ISRCTN24874457 [accessed 2022-11-08]
20. The wellbeing project: internet interventions to improve mental health and wellbeing (ISRCTN65657330). International Standard Randomised Trial Number Registry. 2008. URL: https://www.isrctn.com/ISRCTN65657330 [accessed 2022-11-08]
21. Internet-based guided self-help for overweight and obese patients with binge eating disorder (ISRCTN40484777). International Standard Randomised Trial Number Registry. 2010. URL: https://www.isrctn.com/ISRCTN40484777 [accessed 2022-11-08]
22. Internet delivered treatment for generalized anxiety symptoms in students (ISRCTN16303842). International Standard Randomised Trial Number Registry. 2013. URL: https://www.isrctn.com/ISRCTN16303842 [accessed 2022-11-08]
23. A pilot trial of computerised cognitive behaviour therapy for depression in adolescents (ISRCTN31219579). International Standard Randomised Trial Number Registry. 2011. URL: https://www.isrctn.com/ISRCTN31219579 [accessed 2022-11-08]
24. Efficacy of low-intensity psychological intervention applied by (ICTs) for treatment of depression in primary care (ISRCTN82388279). International Standard Randomised Trial Number Registry. 2014. URL: https://www.isrctn.com/ISRCTN82388279 [accessed 2022-11-08]
25. The effects of virtual reality exposure therapy (VRET) to treat dental anxiety (ISRCTN25824611). International Standard Randomised Trial Number Registry. 2015. URL: https://www.isrctn.com/ISRCTN25824611 [accessed 2022-11-08]
26. Randomised controlled trial of the effectiveness of Big White Wall compared to other online support (ISRCTN12673428). International Standard Randomised Trial Number. 2016. URL: https://www.isrctn.com/ISRCTN12673428 [accessed 2022-11-08]
27. SlowMo trial: a digital therapy for people who fear harm from others 2017 (ISRCTN32448671). ISRCTN Registry. URL: https://www.isrctn.com/ISRCTN32448671 [accessed 2022-11-08]
28. Internet-delivered interventions for people with depression and anxiety in IAPT services (ISRCTN91967124). International Standard Randomised Trial Number Registry. 2017. URL: https://www.isrctn.com/ISRCTN91967124 [accessed 2022-11-08]
29. Digital cognitive behavioural therapy for anxiety (ISRCTN12765810). International Standard Randomised Trial Number Registry. URL: https://www.isrctn.com/ISRCTN12765810 [accessed 2022-11-08]
30. Online remote behavioural intervention for Tics (ORBIT) (ISRCTN70758207). International Standard Randomised Trial Number Registry. 2018. URL: https://www.isrctn.com/ISRCTN70758207 [accessed 2022-11-08]
31. ARIES: App to support recovery in early intervention services (ISRCTN10004994). International Standard Randomised Trial Number Registry. 2018. URL: https://www.isrctn.com/ISRCTN10004994 [accessed 2022-11-08]
32. Evaluating 'E-PLAYS' for children with social communication impairment (ISRCTN14818949). International Standard Randomised Trial Number Registry. 2018. URL: https://www.isrctn.com/ISRCTN14818949 [accessed 2022-11-08]
33. Self-help internet tool for social anxiety symptoms (ISRCTN15819951). International Standard Randomised Trial Number Registry. 2015. URL: https://www.isrctn.com/ISRCTN15819951 [accessed 2022-11-08]
34. Living with bipolar: a web-based self management intervention (ISRCTN64826171). International Standard Randomised Trial Number Registry. 2011. URL: https://www.isrctn.com/ISRCTN64826171 [accessed 2022-11-08]
35. The gameChange trial: automated virtual reality therapy to help patients with psychosis feel more confident in everyday situations (ISRCTN17308399). International Standard Randomised Trial Number Registry. 2019. URL: https://www.isrctn.com/ISRCTN17308399 [accessed 2022-11-08]
36. An online CBT life skills programme for low mood and anxiety (ISRCTN12890709). International Standard Randomised Trial Number Registry. 2015. URL: https://www.isrctn.com/ISRCTN12890709 [accessed 2022-11-08]
37. Obsessive compulsive treatment efficacy trial (OCTET) (ISRCTN73535163). International Standard Randomised Trial Number Registry. 2011. URL: https://www.isrctn.com/ISRCTN73535163 [accessed 2022-11-08]
38. Active assistance for psychological therapy (Actissist): Using mobile technology to deliver cognitive behaviour therapy for psychosis (ISRCTN34966555). International Standard Randomised Trial Number Registry. 2014. URL: https://www.isrctn.com/ISRCTN34966555 [accessed 2022-11-08]
39. Sharp C, Goodyer IM, Croudace TJ. The Short Mood and Feelings Questionnaire (SMFQ): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. J Abnorm Child Psychol 2006 Jun;34(3):379-391. [doi: 10.1007/s10802-006-9027-x] [Medline: 16649000]
40. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
41. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
42. Gu J, Miller CB, Henry AL, Espie CA, Davis ML, Stott R, et al. Efficacy of digital cognitive behavioural therapy for symptoms of generalised anxiety disorder: a study protocol for a randomised controlled trial. Trials 2020 Apr 23;21(1):357 [FREE Full text] [doi: 10.1186/s13063-020-4230-6] [Medline: 32326980]

43. Hollis C, Hall CL, Jones R, Marston L, Novere ML, Hunter R, et al. Therapist-supported online remote behavioural intervention for tics in children and adolescents in England (ORBIT): a multicentre, parallel group, single-blind, randomised controlled trial. Lancet Psychiatry 2021 Oct;8(10):871-882 [FREE Full text] [doi: 10.1016/S2215-0366(21)00235-2] [Medline: 34480868]
44. Chan A, Tetzlaff J, Altman D, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013 Feb 05;158(3):200-207 [FREE Full text] [doi: 10.7326/0003-4819-158-3-201302050-00583] [Medline: 23295957]
45. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: Its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31. [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]
46. van Bastelaar KMP, Pouwer F, Cuijpers P, Riper H, Snoek FJ. Web-based depression treatment for type 1 and type 2 diabetic patients: a randomized, controlled trial. Diabetes Care 2011 Feb;34(2):320-325 [FREE Full text] [doi: 10.2337/dc10-1248] [Medline: 21216855]
47. Griffiths KM, Mackinnon AJ, Crisp DA, Christensen H, Bennett K, Farrer L. The effectiveness of an online support group for members of the community with depression: a randomised controlled trial. PLoS One 2012;7(12):e53244 [FREE Full text] [doi: 10.1371/journal.pone.0053244] [Medline: 23285271]
48. Powell J, Williams V, Atherton H, Bennett K, Yang Y, Davoudianfar M, et al. Effectiveness and Cost-Effectiveness of a Self-Guided Internet Intervention for Social Anxiety Symptoms in a General Population Sample: Randomized Controlled Trial. J Med Internet Res 2020 Jan 10;22(1):e16804 [FREE Full text] [doi: 10.2196/16804] [Medline: 31821151]
49. Lovell K, Bower P, Gellatly J, Byford S, Bee P, McMillan D, et al. Low-intensity cognitive-behaviour therapy interventions for obsessive-compulsive disorder compared to waiting list for therapist-led cognitive-behaviour therapy: 3-arm randomised controlled trial of clinical effectiveness. PLoS Med 2017 Jun;14(6):e1002337 [FREE Full text] [doi: 10.1371/journal.pmed.1002337] [Medline: 28654682]
50. Rozental A, Kottorp A, Boettcher J, Andersson G, Carlbring P. Negative Effects of Psychological Treatments: An Exploratory Factor Analysis of the Negative Effects Questionnaire for Monitoring and Reporting Adverse and Unwanted Events. PLoS One 2016;11(6):e0157503 [FREE Full text] [doi: 10.1371/journal.pone.0157503] [Medline: 27331907]
51. Gumley AI, Bradstreet S, Ainsworth J, Allan S, Alvarez-Jimenez M, Birchwood M, et al. Digital smartphone intervention to recognise and manage early warning signs in schizophrenia to prevent relapse: the EMPOWER feasibility cluster RCT. Health Technol Assess 2022 May;26(27):1-174. [doi: 10.3310/hlze0479]
52. Morriss R, Kaylor-Hughes C, Rawsthorne M, Coulson N, Simpson S, Guo B, et al. A Direct-to-Public Peer Support Program (Big White Wall) Versus Web-Based Information to Aid the Self-management of Depression and Anxiety: Results and Challenges of an Automated Randomized Controlled Trial. J Med Internet Res 2021 Apr 23;23(4):e23487 [FREE Full text] [doi: 10.2196/23487] [Medline: 33890858]
53. Bonell C, Jamal F, Melendez-Torres GJ, Cummins S. 'Dark logic': theorising the harmful consequences of public health interventions. J Epidemiol Community Health 2015 Jan;69(1):95-98. [doi: 10.1136/jech-2014-204671] [Medline: 25403381]
54. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]

Abbreviations

AE: adverse event
CONSORT: Consolidated Standards Of Reporting Trials
ISRCTN: International Standard Randomized Controlled Trial Number
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SAE: serious adverse event

Edited by J Torous; submitted 06.09.22; peer-reviewed by E Eisner, S Allan; comments to author 26.10.22; revised version received 11.11.22; accepted 09.12.22; published 22.02.23.

Please cite as:
Gómez Bergin AD, Valentine AZ, Rennick-Egglestone S, Slade M, Hollis C, Hall CL
Identifying and Categorizing Adverse Events in Trials of Digital Mental Health Interventions: Narrative Scoping Review of Trials in the International Standard Randomized Controlled Trial Number Registry
JMIR Ment Health 2023;10:e42501
URL: https://mental.jmir.org/2023/1/e42501
doi:10.2196/42501
PMID:36811940

Review

Capturing the Dynamics of the Social Environment Through Experience Sampling Methods, Passive Sensing, and Egocentric Networks: Scoping Review

Anna M Langener[1,2,3], MSc; Gert Stulp[2], PhD; Martien J Kas[1], PhD; Laura F Bringmann[3,4], PhD

[1]Groningen Institute for Evolutionary Life Sciences, Groningen, Netherlands

[2]Department of Sociology, Faculty of Behavioural and Social Sciences, University of Groningen & Inter-University Center for Social Science Theory and Methodology, Groningen, Netherlands

[3]Department of Psychometrics and Statistics, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, Netherlands

[4]Interdisciplinary Center Psychopathology and Emotion Regulation, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Corresponding Author:
Anna M Langener, MSc
Groningen Institute for Evolutionary Life Sciences
Nijenborgh 7
Groningen, 9747 AG
Netherlands
Phone: 31 050 363 8
Email: langener95@gmail.com

Abstract

Background: Social interactions are important for well-being, and therefore, researchers are increasingly attempting to capture people's social environment. Many different disciplines have developed tools to measure the social environment, which can be highly variable over time. The experience sampling method (ESM) is often used in psychology to study the dynamics within a person and the social environment. In addition, passive sensing is often used to capture social behavior via sensors from smartphones or other wearable devices. Furthermore, sociologists use egocentric networks to track how social relationships are changing. Each of these methods is likely to tap into different but important parts of people's social environment. Thus far, the development and implementation of these methods have occurred mostly separately from each other.

Objective: Our aim was to synthesize the literature on how these methods are currently used to capture the changing social environment in relation to well-being and assess how to best combine these methods to study well-being.

Methods: We conducted a scoping review according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

Results: We included 275 studies. In total, 3 important points follow from our review. First, each method captures a different but important part of the social environment at a different temporal resolution. Second, measures are rarely validated (>70% of ESM studies and 50% of passive sensing studies were not validated), which undermines the robustness of the conclusions drawn. Third, a combination of methods is currently lacking (only 15/275, 5.5% of the studies combined ESM and passive sensing, and no studies combined all 3 methods) but is essential in understanding well-being.

Conclusions: We highlight that the practice of using poorly validated measures hampers progress in understanding the relationship between the changing social environment and well-being. We conclude that different methods should be combined more often to reduce the participants' burden and form a holistic perspective on the social environment.

(JMIR Ment Health 2023;10:e42646) doi:10.2196/42646

KEYWORDS

social context; experience sampling method; egocentric network; digital phenotyping; passive measures; ambulatory assessment; mobile phone

Introduction

Background

Humans are fundamentally social beings. The lack of social interactions and social relations is often associated with poor mental well-being [1-3]. Although positive social relationships can enhance mental well-being, negative relationships are found to be a risk factor for various psychological and neuropsychiatric disorders [4-7]. In addition, various psychological disorders are characterized by difficulties in engaging in social interactions [8]. Accordingly, clinical therapists and researchers are increasingly attempting to capture the social interactions of their patients to improve treatment outcomes [9,10].

A general question that psychology and other disciplines face in this quest is how to best capture people's social environment and its effect on an individual's well-being. This question has become more pertinent in recent years, in which research has recognized that important psychological processes are dynamic, implying that emotions and (social) behavior fluctuate over time, and static measures do not suffice in capturing these dynamics [11]. This insight has led to the development of several technologies in different disciplines that are able to capture the changing social environment, for example, day-to-day social interactions. In total, 3 methods are, to the best of our knowledge, currently most often used to assess the changing social environment occurring in daily life, namely, the experience sampling method (ESM) [12], passive sensing (including the electronically activated recorder [EAR]) [13,14], and egocentric networks [15]. The development and implementation of these methods have thus far occurred largely separately from each other.

In psychology, ESM is used as a standard tool to capture how psychological processes evolve over time. It offers great potential as it provides a picture of daily emotions and (social) behavior [16]. Participants receive a device to fill out brief daily questionnaires when a signal occurs. For example, participants fill out a questionnaire on an app after a push notification is sent.

Given that ESM can be burdensome for the participants, different disciplines have started to use passive sensing, which collects data via sensors from smartphones or other wearable devices (eg, smartwatches). This allows for monitoring participants constantly over time. For instance, smartphones can track an individual's whereabouts (through a GPS), social engagement (through the use of social media apps), and social interactions (by measuring people's conversations using the microphone [17,18]). A subgroup of studies that use sensors to passively collect data are EAR studies. In EAR studies, brief snippets of environmental sounds are recorded to investigate the activities or emotions of a person [14,19]. Emotions can be assessed by transcribing the audio snippet and using Linguistic Inquiry and Word Count [20]. Linguistic Inquiry and Word Count is a dictionary that calculates the percentages of included words for different categories. Some categories describe the emotional tone someone used while communicating, such as positive and negative affect [21].

Sociologists often use repeated egocentric networks to investigate how social connections change over time, for example, whether the occurrence of mental illness leads to network attrition, which means that fewer social contacts are included in the social network [22], or how the social network of a person changes during substance abuse disorder recovery [23]. To collect egocentric networks, researchers typically ask so-called name-generating questions to identify the important social contacts of a person, which are also called alters. For example, a name-generating question could be "Please list 25 names of individuals...with whom you have had contact in the last year" [24]. After alters are identified, further questions about the characteristics of those alters and the characteristics of the relationship to those alters are asked (eg, age of the alter, frequency of contact, and how close someone is to the alter) [15]. Overall, egocentric networks represent relationships (ties) between a specific individual (ego) and connected persons (alters). When relationships between alters are assessed, structural properties and composition effects of the ego's network can be calculated, such as density (the proportion of existing ties relative to all possible ties) [15], which has been shown to affect ego's health outcomes [25]. Similar to ESM, egocentric networks are burdensome to collect [26,27], for example, compared with passive measures.

These tools are likely to tap into different parts of people's changing social environment. In this paper, we broadly focus on all aspects of the social environment that differ for each person and vary over short time scales (eg, minutes, days, and weeks). We include the behavioral and psychological levels, also known as the structural and cognitive levels [28,29]. The behavioral level indicates both what people do and the structural aspects of the social network (eg, network size), whereas the psychological level shows, for instance, what people feel or think [28,29]. For example, a person can have frequent social contact (behavioral level) and still feel lonely (psychological level) [4]. Therefore, we consider both levels (psychological and behavioral) as important to capture the full changing social environment.

Objectives

So far, it remains unclear which parts of the changing social environment ESM, egocentric networks, and passive sensing attempt to capture in the literature and whether and where there is overlap. Therefore, the first question that we addressed is which aspects of the behavioral and psychological levels of the changing social environment are measured using the 3 different techniques. This question is important in light of the call for a better conceptualization of (psychological) constructs to advance theory and measurement [30-32]. We provided a way forward in how the changing social environment can be conceptualized by showing which period each method captures and which aspects of the social environment different methods tap into. This will also advance the understanding of mental health as it provides an overview of which aspects of the changing social environment are currently studied in relation to well-being.

For each of these methods, it is important to consider whether they measure the concept that they are intended to measure. Thus, the second question that we addressed is if and how the

measures are validated. This second question mirrors a recent debate that psychological research is facing, namely, the problems associated with the use of nonvalidated questionnaires [33]. We investigated whether this also applies to measures that are used to capture the social environment.

Finally, ESM, egocentric networks, and passive sensing come from different disciplines and, therefore, are likely to be used separately from each other. However, we believe that a holistic understanding of the dynamic relationship between mental health and the changing social environment requires insights from different aspects that are measured using the different tools. We do not know how often these tools have been combined in previous research and how they can be combined optimally. Consequently, the last question that we addressed is how the measures have been combined in previous research.

Methods

Protocol

We conducted a scoping review [34]. Although this was not a systematic review, we designed and wrote this study according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement. The PRISMA checklist can be found in Multimedia Appendix 1 [35]. This review was not registered before conducting it.

Literature Search Strategy

We aimed to identify studies that used methods that were able to capture the changing social environment in relation to well-being in an adult population. Therefore, we conducted a systematic search to identify studies that used ESM, passive sensing, or repeated egocentric networks to measure the changing social environment in relation to well-being or psychological disorders. The final database search was completed in July 2021. We conducted searches in Web of Knowledge, PsycINFO, and PubMed with no date limitation. We created 2 different search strings and ran each search string in each database to identify studies of interest. Both full search strings can be found in Multimedia Appendix 2 [36-38].

Eligibility Criteria

For inclusion, studies must have (1) used ESM, passive sensing, or repeated egocentric networks to measure the changing social environment (we included all studies that broadly focused on aspects of the social environment, which varied over short time scales [eg, minutes, days, and weeks], as well as studies that focused on behavioral aspects [such as social interactions] and psychological aspects [such as social support or relationship characteristics]); (2) measured the social environment in relation to well-being, psychological disorders, or psychiatric disorders; (3) measured the social environment in an adult population without nonpsychological medical conditions; (4) studied the social environment in the daily life of a participant; and (5) been published in English.

Consequently, we excluded studies that focused on children or older adults and medical treatment evaluations of nonpsychological or psychiatric disorders. Older adults were identified as such if the study explicitly said that it focused on, for example, “old age,” “older women,” or “older people” [39-41]. We also excluded studies that focused on medical conditions (eg, stroke). Furthermore, we excluded studies that only focused on the social environment in an occupational setting or on dyadic interactions as we were interested in methods that capture the broad social environment in daily life (see Multimedia Appendix 3 for a more extensive list).

The eligibility of the studies was assessed by 1 researcher (AML). A random subsample of studies (n=21) at the beginning of the coding process was independently assessed by 2 researchers (AML and LFB) to formulate eligibility criteria and discuss discrepancies with the research team.

Data Extraction

We extracted information about the study (year of publication and objective), sample characteristics (sample size, sample population, and country of residence), procedure (study length and assessment frequency), statistical analysis, validation, and which sensors were used in passive sensing studies.

We further extracted information on which questions were used to capture the social environment. The questions used in the selected ESM and egocentric network studies were diverse and, therefore, hard to analyze. Thus, we developed categories that summarized these questions. We developed those categories as an iterative process. After reading 15 studies, we summarized the measures of the social environment that were repeatedly used across the studies. While reading more papers, we revised these categories. In total, we identified 12 categories for ESM and 5 categories for egocentric networks, which are reported in the *Results* section. We created the category “Other” for items that did not fit in any of the other categories.

Data extraction was conducted by 1 researcher (AML). The categories were developed together, and specific cases were discussed within the research team to increase the clarity of the categorization of the studies.

Risk of Bias

Our primary aim was to identify studies that used measures to capture the changing social environment. Similarly to O’Donohue et al [42], we argue that the outcome and quality of the studies were less relevant and would not have affected the eligibility of the studies as no hypothesis was tested. Therefore, the risk of bias for each study was not assessed. However, the validity of the measures used was examined and is described in the *Question 2: How Well Are the Measures of the Social Environment Validated?* section.

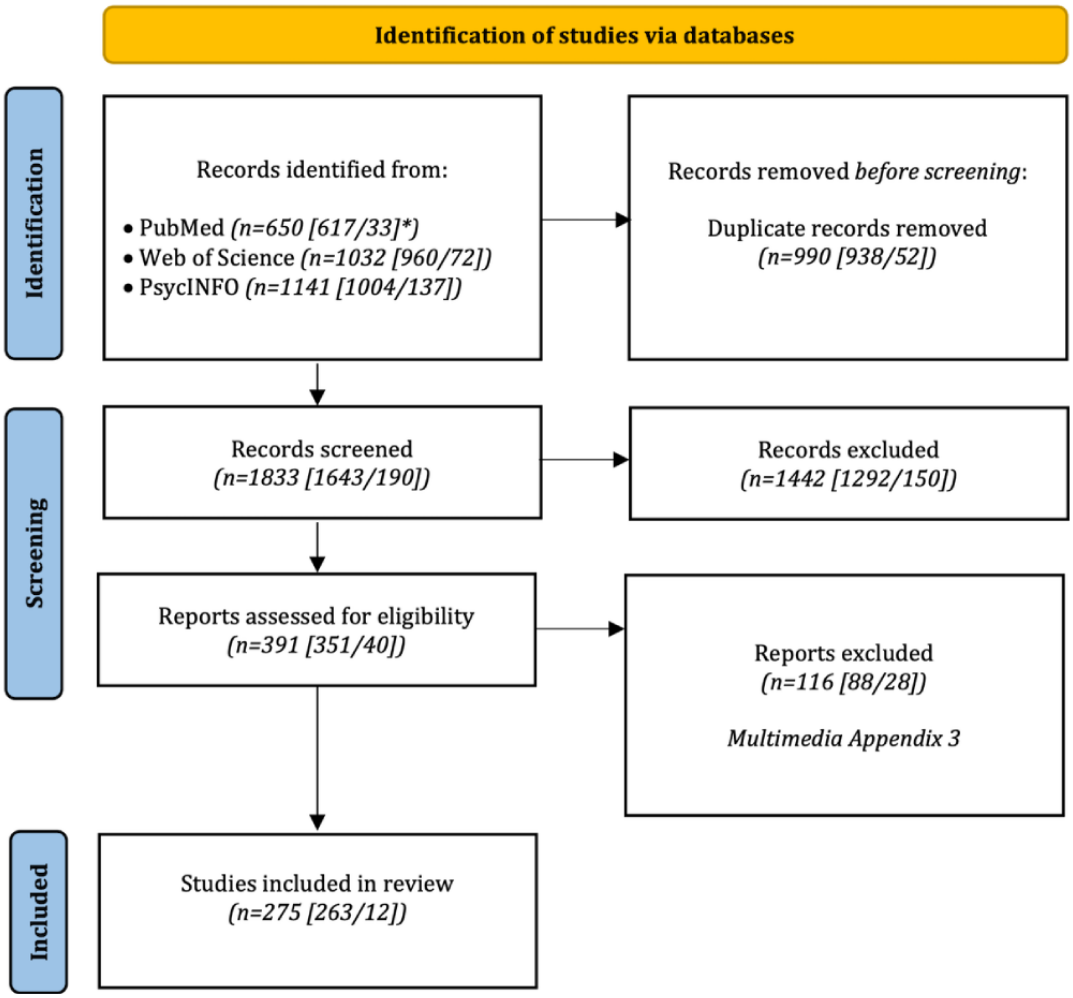
Results

Overview of the Selected Studies

We identified 1833 articles that measured the social environment using ESM, passive sensing, or repeated egocentric networks after removing duplicates (Figure 1). If we decided to exclude a study during the screening process, we indicated a reason for doing so. This overview can be found in Multimedia Appendix 3. We removed 78.67% (1442/1833) of the studies after title and abstract screening. In total, 29.7% (116/391) of the studies

were excluded after reviewing the full text. Thus, we included 275 studies in our scoping review.

Figure 1. Flow diagram of the included studies (adapted from Page et al [35]). *The first number indicates the records that were identified using the first search string, which aimed to extract studies that used the experience sampling method or passive sensing, whereas the second number indicates the records that were identified using the second search string, which aimed to extract studies that used repeated egocentric networks.



Most of the included studies only used 1 method, meaning that studies only used ESM (223/275, 81.1%), passive sensing (19/275, 6.9%), or egocentric networks (12/275, 4.4%). Note that some of the identified papers included multiple studies, which leads to a total of 238 studies that only used ESM, 20 studies that only used passive sensing, and 12 studies that only used egocentric networks.

In total, 5.5% (15/275) of the studies combined ESM and passive sensing to measure the social environment. Repeatedly assessed egocentric networks were not combined with other methods. However, 2.2% (6/275) of the studies combined egocentric networks that were assessed only once with ESM (4/6, 67%) or passive sensing (2/6, 33%).

Study Characteristics

Multimedia Appendix 3 provides an overview of the individual characteristics of each study. The studies were published between 1987 and 2021. The studies took place mainly in the United States (178/275, 64.7%) and Europe (80/275, 29.1%).

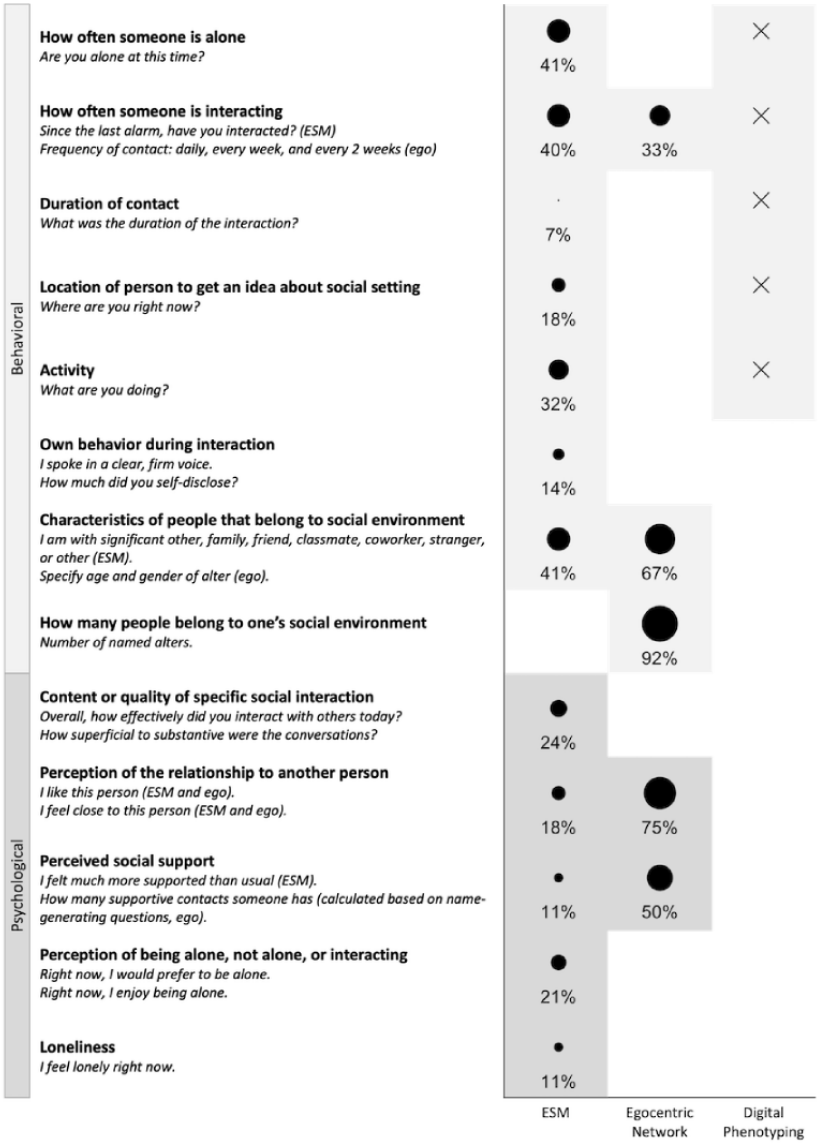
On average, 201 (SD 706; median 102) participants took part in the 275 selected studies. More than half of the studies included (healthy) students (105/275, 38.2%) or adults (63/275, 22.9%) as participants. In total, 2.2% (6/275) of the studies included participants with social anxiety disorder, and 2.5% (7/275) of the studies included participants with social anxiety disorder and healthy controls. In addition, of the 275 studies, 46 (16.7%) included participants with other psychological disorders, and 34 (12.4%) included participants with other psychological disorders and healthy controls. Next, 5.1% (14/275) of the studies included participants who used substances, for example, participants who were heavy drinkers, smoked, used cannabis, used other drugs, or were recovering from substance abuse disorder.

Question 1: Which Characteristics of the Social Environment Are Included and on Which Time Scale?

In this section, we describe which diverse aspects of the social environment and which temporal resolution (ie, assessment frequency and study duration) each method captured. ESM studies assessed the social environment using approximately 6 questionnaires per day, for a study length of 15 days. Questions that were asked in ESM studies were usually about what the participant was doing or feeling at the moment or since the last questionnaire. A known strength of ESM studies is that they tap into several aspects of the changing social environment by

capturing both behavioral and psychological characteristics directly through questionnaires (Figure 2). The behavioral level includes what people do and the structural aspects of the social network (eg, network size). In contrast, the psychological level shows, for instance, what people feel or think [28,29]. In total, 83.6% (199/238) of the selected ESM studies included behavioral aspects, and only 66.8% (159/238) of the ESM studies included psychological aspects.

Figure 2. Overview of which aspects each method captures. The example items are based on the items that were used in the selected experience sampling method (ESM) and egocentric network studies. We developed categories to summarize which aspects of the social environment were captured in the selected studies. The first column shows which aspects the ESM studies captured and how often (in percentage) those aspects were measured in the 238 selected ESM studies that only used 1 measurement. Subsequently, the second column shows how often (in percentage) different aspects were included in the 12 selected egocentric network studies. The last column shows which aspects can be potentially captured using passive sensing; as those aspects are only indirectly and often implicitly measured, we did not calculate any percentages.



Behavioral aspects that were frequently addressed were whether someone was alone (97/238, 40.8%) or interacting (94/238, 39.5%) [43] and the characteristics of social interaction partners (97/238, 40.8%). There was less consistency across ESM studies in which psychological aspects were captured. For example, studies occasionally included questions about the content or quality of an interaction (58/238, 24.4%) [43,44] or whether someone enjoyed being alone or not alone (49/238, 20.6%) [45]. In addition, questions about relationship characteristics were occasionally asked (42/238, 17.6%) [46]. Thus, ESM studies frequently involved whether someone was not alone or interacting, but such studies were less consistent in assessing the psychological level of the social environment and social behavior.

Similar to ESM studies, a strength of egocentric network studies is that they assess behavioral (12/12, 100% of egocentric network studies) and psychological (10/12, 83% of egocentric network studies) aspects of the social environment. ESM and egocentric networks capture similar aspects of the social environment. At the behavioral level, the egocentric network studies measured characteristics of social contacts (eg, age or gender; 8/12, 67%) and occasionally asked how often someone was interacting (4/12, 33%; Figure 2), for example, by asking about the frequency of contact in the past [47]. At a

psychological level, subjective relationship characteristics (9/12, 75%) and perceived social support (6/12, 50%) were often captured.

It is evident that, contrary to ESM studies, egocentric network studies focused more on the interaction partners than on the interaction itself. Studies were longer (approximately 4 years), with 3 rounds of measurements in these periods. Questions that were asked in egocentric network studies also referred to longer periods. For example, Francis [23] asked participants to name the people they had had contact with in the previous 6 months. Egocentric network studies asked a so-called name-generating question (eg, name people participants considered meaningful [47]) at the beginning of the study. This allowed them to give an estimate of the network size of a person. Furthermore, detailed information on interaction partners was available. In contrast, ESM often measures the characteristics of the social environment at a specific moment without linking them to a specific person. Even if a study specified with whom someone was interacting, often names or initials were missing and only broad categories were assessed (eg, significant other, family, friend, classmate, coworker, stranger, and other [46]).

In contrast to ESM and egocentric networks, a strength of studies that use passive sensing or EAR is that they can objectively and directly tap into behavioral aspects. These can be measured continuously using sensors during the day. The reviewed studies were, on average, 27 days long. Studies that used passive sensing or EAR mostly included the audio (microphone; 12/20, 60%) and sociability (calls or SMS text messages; 9/20, 45%) of a person. These measures can indirectly capture whether someone is alone or interacting [14,48]. In addition, mobility measures (GPS and accelerometer) were included in half (10/20, 50%) of the studies. Measures that were less often included were phone-related properties such as the total time spent on communication apps (3/20, 15%) [49].

Thus, ESM and passive sensing can both capture the location, activity, and contact frequency of a person (Figure 2). A weakness of passive sensing is that it can only indirectly capture the psychological level using its sensors (raw measures). For example, audio data must be manually coded to examine the emotions of a person [14], or passive measures must be combined with questionnaires to build models that can predict psychological variables such as loneliness [50,51]. In contrast, ESM can examine the psychological level directly by asking the person about their experiences and feelings.

Question 2: How Well Are the Measures of the Social Environment Validated?

In the previous section, we identified which elements of the social environment were captured by the different methods. In this section, we discuss the validity of these methods. In its broadest sense, validity refers to the extent to which a method measures what it is intended to measure [52,53]. This section discusses the strategies that the selected studies chose to provide evidence of validity.

For ESM (and all other questionnaire-based measures), it is essential that items that claim to measure a specific construct do so. Obviously, this is difficult as researchers may define their constructs differently and may have different opinions about what loneliness is, for example. A way to examine the validity of the selected studies is to investigate whether validated questionnaires were used. Of the selected ESM studies, only 27.7% (66/238) used a validated questionnaire. For example, the University of California, Los Angeles, Loneliness Scale was often used to assess loneliness [54,55]. In contrast, other studies used only 1 nonvalidated question, such as “At this moment, how lonely do you feel?” [56], to assess loneliness.

Some studies (45/238, 18.9%) reported that they included items that were already used in previous research. Using items that were used in previous research does not indicate validity. However, it does make the studies more comparable with each other. Slightly more than half (126/238, 52.9%) of the studies did not use validated scales or items from previous research. Nevertheless, some authors (25/238, 10.5%) provided arguments and details about their item selection [45,57]. Overall, 42.4% (101/238) of the ESM studies did not use validated items to measure the social environment and did not give a specific reason for including the items in their research.

Questions about the behavioral level of the social environment may require less validation or explanation than questions measuring the psychological level. For example, whether someone is alone or not is easier to measure than the feeling of loneliness. Thus, it is even more important to validate questions at the psychological level. To investigate the validity of questions at the psychological level, we looked at a subset of ESM studies that included at least one question at the psychological level of the social environment (159/238, 66.8%). This subset used slightly more validation strategies (103/159, 64.8%) compared with the full sample (137/238, 57.6%).

In contrast to ESM studies, egocentric network studies aim to capture a part of the social network of a person. Validity in this case is different from that in ESM studies as the overall goal is to obtain an accurate representation of the overall network compared with measuring specific constructs. A topic that hinders validity in this regard is the size of the assessed egocentric network [24]. Depending on the number of alters included, the validity of network characteristics might differ as including too few alters might bias the network composition and structure estimates [24,26,58]. McCarty et al [26] found that naming 25 alters can capture most of the structural patterns of a network. The number of alters that participants were allowed to name differed across the included studies. In half (6/12, 50%) of the studies, participants were allowed to name an unlimited number of alters. In other studies (5/12, 42%), the number of alters that participants could include was limited and, therefore, might have biased the network estimates. However, in all except 8% (1/12) of the studies, participants were allowed to name at least 25 alters and named, on average, fewer alters than they were allowed to [59,60]. Thus, it is likely that those studies still provided valid estimates of the network structure as the limit of alters that participants were allowed to enter was not reached and at least 25 alters were allowed to be included in the network. It is important to note that how many alters participants included in their network might also be influenced by their motivation. Low motivation can lead to a biased network estimate even if unlimited alters can be added [27].

Another aspect that might affect validity in egocentric network studies is whether an appropriate name generator was used and whether the questions asked to assess the relationship characteristics were valid. Using different name generators can affect the network size and average tie characteristics [61]. Therefore, using name generators that have been used and tested in previous research might increase the validity and comparability across studies. In most studies (10/12, 83% of the egocentric network studies), the name generator questions, as well as the content questions asked to assess the relationship characteristics, had been used in previous studies. A total of 8% (1/12) of the studies did not use questions based on previous studies but only asked whom a person had contacted in the last 6 months and several characteristics of the alter [23]. A total of 8% (1/12) of the studies also did not use questions from previous research but developed their own questions based on previous literature reviews and results from earlier assessed egocentric networks [62]. Thus, except for 8% (1/12) of the studies, egocentric network studies used strategies that increased the validity and comparability across studies.

For passive sensing, validation is rather different than for questionnaire-based items, as in ESM and egocentric networks. For passive sensing, it is important that the extracted features or algorithms applied indeed capture the parts of the social environment that they are designed to capture. First, we discuss passive sensing studies excluding EAR studies. In the selected studies, only half (6/12, 50%) of the studies that only used passive sensing measured how the extracted features or the built model with a combination of features related to the changing social environment. For example, Jacobson et al [63] aimed to detect how passively collected data (ie, accelerometers and incoming and outgoing calls and SMS text messages) can predict the level of anxiety in social situations (social anxiety). They found a correlation of r=0.7 between predicted social anxiety severity and social anxiety measured using a questionnaire. Thus, they concluded that passive sensing can be used to detect social anxiety. We discuss this issue further in the following section on the combination of methods. The other half (6/12, 50%) of the studies that used only passive sensing (excluding EAR studies) did not mention any techniques to show how their passive collected data related to the social environment. For instance, Tsapeli and Musolesi [64] used GPS and accelerometer data as implicit indicators of social interactions. However, it remains unclear how well GPS and accelerometer data can capture social interactions. In addition, Schuwerk et al [65] used the number of contacts as an indicator of social network size [65], but it is uncertain how accurately the number of contacts can describe the social network size as no research was done to investigate this.

In EAR studies, validity is about the coding of the audio files. For example, extracted audio features from the microphone should refer to the social interactions a person had. If the coding is performed manually, this should be done in a standardized way. Predefined coding schemes can help with this process. In total, 50% (4/8) of the EAR studies used such a predefined coding scheme [14] (eg, social environment coding of sound inventory [19]). A total of 25% (2/8) of the studies used interrater reliability, meaning that the extent of agreement across different coders was measured. In 12% (1/8) of the studies, 28% of the participants listened to parts of their audio to verify the given coding [66]. In total, 12% (1/8) of the studies only looked at the presence of speech or presence of others without any specific validation [67]. In this study, this was done automatically with the Google Cloud Speech-to-Text software [67]. If the software generated a transcript, the presence of speech was assumed without active verification by a human.

Overall, ESM, egocentric networks, and passive sensing use different methods to improve the validity of their collected data. Most of the egocentric network studies (10/12, 83%) and EAR studies (7/8, 88%) included procedures that improved the validity of their collected data. In contrast, this percentage was much lower in ESM and passive sensing studies, with only approximately 60% (137/238, 57.6%) of the ESM studies and 50% (6/12) of the passive sensing studies mentioning techniques that supported the validity of their collected data.

Question 3: How Were the Measures Combined in Previous Research?

Overview

To better understand how methods can be best combined, we identified 7.6% (21/275) of the studies that used more than one of our discussed methods. In total, 29% (6/21) of these studies used 1 method to assess the social environment, whereas the second method was used to capture something different (eg, mood) [68-71]. As we focus in this section on studies that used multiple methods to capture the social environment, we only discuss the remaining 71% (15/21) of the studies in more detail. We start by discussing studies that combined methods to measure several aspects of the social environment and their relation to a third variable (eg, well-being). These studies attempted to capture multiple aspects of the social environment through a combination of methods. Next, we discuss studies that examined the overlap in how different methods characterize the social environment. These studies relate to the validation of measures as different methods measured similar parts of the social environment and many studies investigated how those measures of the social environment agreed with each other.

Studies That Combined Different Methods to Cover a Wider Part of the Social Environment

In 20% (3/15) of the studies, different measures were combined to cover a wider part of the social environment [65,72,73]. A total of 67% (2/3) of these studies did this to investigate questions regarding different parts of the social environment. Schuwerk et al [65] investigated offline social interactions measured via ESM and interactions measured via smartphones (eg, number and duration of calls). Abel et al [72] aimed to capture social interactions objectively via EAR and, in addition, used daily questionnaires to capture subjective emotions during a social interaction [72]. Thus, a combination of methods helped answer several research questions.

Another study aimed to predict stress and mental health and highlighted that a combination of measures of the social environment (ie, passive sensing and ESM) provided better predictions of self-reported stress and poor mental health than each method on its own [73]. In this study, data were passively

collected using smartphones and wearable devices. In addition, participants had to fill out a diary. Overall, using passively collected data improved the prediction of stress and mental health, although the improvements were not large (from 72% to 82% for stress and from 85% to 87% for mental health).

Dynamically assessed egocentric networks were not combined with other methods, but one-time–assessed egocentric networks were used to collect more information on interaction partners or the total network size. These studies helped better understand how repeated egocentric networks could be combined with ESM or passive sensing in future research. First, 27% (4/15) of the studies combined one-time–assessed egocentric networks with ESM to identify important contacts before the ESM data collection started [74,75], determine the relationship characteristics of daily contacts [76], or have a measure for the network size of a person [77]. It was shown that the presence of people who provide emotional support, which was measured using a combination of ESM and egocentric networks, was associated with well-being measured via ESM [76]. In addition, controlling for network size was important for understanding stress and affect [77]. Second, 13% (2/15) of the studies used (measures similar to) one-time–assessed egocentric networks together with passive sensing [50,78] to assess network members (eg, close friends, family members, and friends on campus) and determine whether someone interacted with a close friend based on Bluetooth connections or phone numbers. Hence, combining egocentric networks with passive sensing can be useful to create meaningful variables based on the passive sensing measures, such as how many close friends someone had contact with instead of just counting Bluetooth connections.

Overall, few studies combined different techniques to measure the social environment. However, combining different techniques improved the prediction and understanding of the social environment in relation to mental well-being.

Studies That Examined the Overlap in How the Social Environment Is Characterized by Different Methods

We identified 27% (4/15) of studies that indicated that passively collected data correlated with parts of the social environment and social behavior assessed using ESM [79-82]. First, Abdullah et al [79] found promising results using passive measures such as nonsedentary duration based on the accelerometer and conversation frequency based on the microphone to predict how stable the social rhythm of a person is. The social rhythm of a person was measured through ESM questions on the daily routine, such as at what time someone would get out of bed and have their first social contact. In addition, other studies found a relationship between passive measures (such as total time traveled based on GPS) and a person's social context and activity (such as whether a person was alone, having a conversation, interacting, or in a location [80-82]). However, the strength of the association differed by study, measure, and participant group. For example, Fulford et al [81] showed that passive measures related to an individual's social behavior differed for individuals with schizophrenia and healthy controls. The association between distance traveled and number of interactions ranged from 0.07 for people with schizophrenia to 0.6 for healthy controls [81]. In general, they concluded that passively collected mobility features (GPS) relate moderately to ESM measures of social behavior for healthy individuals (ranging from a correlation of r=0.31 to r=0.7), whereas audio relates moderately to ESM measures of social behavior for people with schizophrenia (correlations of approximately 0.5). Overall, this 27% (4/15) of studies indicated that there are moderate associations between passive measures and social behavior but that it is difficult to draw any final conclusions because of the variation in methods and results across and within studies.

In the remainder of this section, we describe 13% (2/15) of studies that used EAR together with ESM. Interestingly, studies that used ESM together with EAR only showed moderate agreement between the assessed variables. Minor et al [21] investigated how EAR is implemented in students with schizotypy and how social engagement differs between students characterized by high or low schizotypy. Students had to wear EAR for 2 days and fill out ESM for 16 days. Surprisingly, even though the conclusions drawn from each measurement were similar, there was little overlap between EAR and ESM measures for positive and negative affect. The authors argue that both measures capture different facets of daily life. EAR captures affect without the interpretation of the participant, whereas ESM adds a subjective context to it, such as the quality of the relationship between interaction partners. Similarly, Sun et al [51] used EAR and ESM to measure the relationship between the quantity and quality of social interaction and well-being. They only found a moderate agreement between EAR and ESM regarding when participants were interacting (r=0.39), conversational depth (r=0.31), and self-disclosure during an interaction (r=0.31).

To conclude, our results indicate that passively collected data and ESM assess overlapping aspects of the social environment but that more validation studies are needed to investigate which aspects are robustly measured with passive sensing data across studies [79-81]. In addition, studies that used ESM together with EAR only showed moderate agreement between the assessed variables. This indicates that the convergent validity between the 2 measures is not high, meaning that different measures of the same construct do not agree much [83]. Nevertheless, the overall conclusions drawn from both measures (ESM and EAR) regarding the relationship to schizotypy and well-being were similar [21,51].

Discussion

Principal Findings

Our aim was to study how distinct methods originating from different disciplines were used to capture people's changing social environments, identify the strengths and weaknesses of each method, and detect opportunities to combine them. We focused on how ESM, passive sensing, and repeated egocentric networks are currently used to measure the social environment. We investigated how parts of the social environment were captured by each method and the validity of the measures used in each method. Furthermore, we examined how these methods have been combined in previous research.

In total, 3 major implications follow from our literature review, which we will discuss in detail in the following sections, and are summarized in Textbox 1. First, a combination of methods is essential to better capture the changing social environment as each method captures a different resolution (duration and frequency) of the social environment and assesses aspects of the social environment at different levels (psychological vs behavioral). However, a combination of methods has rarely been observed in previous research. Second, a combination of methods has tremendous potential to reduce both researcher and participant burden as there is overlap in the various constructs they capture. Third, measures of the social environment are rarely validated, which undermines the robustness of the conclusions drawn. We will provide suggestions in the following sections for future efforts to collect and analyze data.

Textbox 1. Recommendations for future studies measuring the changing social environment.

Combine methods to achieve a more comprehensive picture of the social environment

- Tap into more aspects of the social environment by using different methods to capture what people do and what people feel and perceive. Frequently assessed aspects in the included studies were the following:
  - What people do: how often someone is alone, how often someone is interacting, duration of an interaction, location, activity, own behavior during interaction (measured using the experience sampling method [ESM]), characteristics of persons that belong to the social environment, and how many people belong to one's social environment (network size, measured using egocentric networks)
  - What people think and perceive: content or quality of a specific interaction (measured using ESM), subjective perception of the relationship to another person, perceived social support, perception of being alone, not alone, or interacting (measured using ESM), and loneliness (measured using ESM)

Combine methods to measure aspects on a suitable frequency and reduce researcher and participant burden

- Choose a suitable time scale for capturing the constructs of interest, which means that fluctuating aspects should be measured using ESM and passive sensing and stable aspects can be captured using repeated egocentric networks (make sure to collect identifiers [nicknames or initials] in the ESM to collect the egocentric network). Stable characteristics might be the following:
  - Characteristics of a person that belongs to one's social network (eg, gender; age; and objective relationship characteristics such as friend, colleague, or coworker)
  - Subjective relationship characteristics that do not change daily (eg, how close someone is)
  - Perceived social support to specific persons
- Consider using passive measures to reduce the length of ESM questionnaires. For example, location or physical activity can potentially be assessed using passive sensing.

Better validation for ESM and passive sensing

- Use validated ESM items or other strategies that increase validity. Other strategies that were used in the studies in our review were the following:
  - Choosing questions based on previous research (eg, ESM Item Repository)
  - Providing arguments for why a specific item was included (especially if it was not validated)
  - Choosing ESM items that relate to validated cross-sectional scales
  - Explanation of the ESM questions to the participants
  - Doing a test round or pilot round
  - Doing a multiverse analysis
  - Choosing a suitable method for a psychometric evaluation of the items
- Make explicit what one aims to measure using passive measures (ie, which part of the changing social environment one aims to capture).
- To validate which part of the changing social environment the passive measures capture, consider using ESM in addition.

Combining Methods Will Achieve a More Comprehensive Picture of the Social Environment and Can Reduce Participant Burden (Compared With Only Using ESM)

We argue that future research would benefit from combining different methods to capture a wider part of the social environment. A combination will help measure aspects that would have been missed by a single method but are important for well-being. The results from Sano et al [73] substantiate this conclusion by showing that combining ESM and passive sensing provides better well-being predictions than each method on its own. However, our results show that a combination of methods has rarely been observed in previous research. Only 5.5% (15/275) of the studies combined ESM and passive sensing, and none of the selected studies combined all 3 methods.

Importantly, it is well known that ESM is suited to capture both the behavioral and psychological aspects of the social

environment and, therefore, is a great tool for capturing a wide part of the social environment. However, our results showed that, even though ESM can capture both levels, approximately one-third of the ESM studies (79/238, 33.2%) only captured the behavioral level (eg, whether someone was alone or interacting) without assessing the psychological level (eg, how the quality of the interaction was perceived). This does not align with previous research that showed that the psychological aspects of the social environment are important to understand well-being [7,46]. Thus, we would recommend that future studies include more aspects of the psychological level as ESM studies are especially suited to measure in-the-moment experiences.

The remainder of this section focuses on how these methods can be combined. First, combining ESM with egocentric networks would be beneficial to measure the variable of interest at a suitable frequency. On the one hand, it is known and supported by our results that egocentric networks are often only assessed infrequently (eg, twice a year). However, substantive fluctuations in mood and behavior occur over shorter time scales, and ESM is needed to capture those fluctuations [11]. In contrast, ESM measures the social environment several times a day, but some relationship characteristics might be more stable. Thus, similarly to Hopwood et al [84], we recommend explicitly thinking about the duration and frequency of change of a (social) construct and determining how often and frequently it should be measured based on a theoretical conceptualization [84]. We recommend capturing variables that do not change frequently using egocentric networks instead of using ESM. This reduces the length of daily ESM questionnaires and, thus, the participant burden [85].

For example, we know that whether a contact is close or not or the characteristics of an interaction partner (eg, gender) do not change frequently during the day. Nevertheless, our results show that these questions, if asked in ESM studies [46], are often asked multiple times a day. In the reviewed studies, often the identities of interaction partners were missing, and only general categories were assessed (eg, significant other, family, friend, classmate, coworker, stranger, and other [46]). Thus, our results indicate that the same questions on interaction partners may be asked multiple times per day. This is contrary to the purpose of ESM, which is designed to measure in-the-moment experiences rather than static characteristics. Hence, for future research, it would be beneficial to assess the names or initials of an interaction partner in ESM studies and measure more static relationship characteristics using an egocentric network at the beginning or end of a study (such as the nature of the relationship, closeness, and characteristics of the relationship with the interaction partner).

Some of the selected studies already did something similar by assessing the close contacts of a person using a questionnaire before passive sensing or ESM started [50,75,76]. This information helped cover more aspects of the social environment and shorten the ESM questionnaire. Moreover, Sun et al [51] proposed to assess relationship characteristics using a baseline questionnaire and only ask people during ESM who they had interacted with based on this questionnaire. Currently, Stadel [86] is working on combining ESM and (one-time–assessed) egocentric networks for clinical practice.

Passive sensing measures might be further useful to reduce the length of ESM questionnaires. Our review identified that some aspects of the social environment are captured via passive sensing and ESM, such as a person's location and activities [64,87]. Given the major burden for respondents to frequently report their activities and location through ESM, which reduces the validity of the data and response rate [88,89], passive sensing can be an important tool to assess the activity and location at a high resolution and for a long duration with low participant burden. There is an additional opportunity to replace ESM questions about the social environment (eg, number of interactions) with passive sensing measures (eg, distance traveled), which would further reduce participant burden [79-81,90]. However, it is too premature to make more specific recommendations as it is not yet clear how much overlap exists between passive measures and ESM and research suggests that it varies across studies and individuals. Furthermore, different methods might capture different underlying constructs even if they are supposed to measure the same ones [21,51]. Further validation is needed before particular questions can be replaced with passive sensing.

Improving Validation Techniques

Although the included studies attempted to characterize the social environment, we identified shortcomings in the use of validated measures to capture the social environment. We observed two main problems with respect to the validity of the data: (1) the use of nonvalidated items in ESM studies and (2) the use of features in passive sensing studies related to the social environment that were not validated.

Approximately 70% (172/238, 72.3%) of the ESM studies did not use items that were based on validated questionnaires. This is in line with recent articles emphasizing that psychological research has a major issue with validation and questionnaires without validation are applied frequently [33,91,92]. Using nonvalidated questionnaires can lead to problems in the interpretation of research results and reduce the robustness of the study conclusions [33]. It further makes it hard to compare results across studies.

A way forward to improve ESM measures would be to conduct more studies that validate measures for different aspects of the social environment and use strategies that improve the validity of the measures. For example, as described by Mestdagh and Dejonckheere [91], recent innovations such as the ESM Item Repository [93] or multiverse analysis [94] are a first step toward improving the reliability and validity of ESM item selection. Fortunately, different statistical methods have also become available that can be used to evaluate the reliability of ESM items [92,95-99].

In contrast to ESM, validity in egocentric network studies refers to obtaining an accurate estimate of network characteristics. This estimate highly depends on the number of alters included [24,26,58]. In addition, using an appropriate name generator and questions to assess the relationship characteristics is important. Half (6/12, 50%) of the selected studies allowed the

participants to name an unlimited number of alters. Even though the number of alters was limited in some studies, participants named, on average, less alters than they were allowed to. Thus, it is likely that these studies still provided a valid estimate of the network characteristics. The name generator and questions asked to assess the relationship characteristics were mostly used in previous research, which increases comparability (>80%).

Passive sensing suffers from a lack of studies that show that features based on passively collected data relate to the construct that is intended to be measured. Although some studies (6/12, 50%) used pre- and postassessment questionnaires [63] to investigate how passive measures related to the social environment, 50% (6/12) of the studies that only used passive sensing did not use any validation to show how their passive measures relate to the social behavior of a person (or other intended measures). For example, GPS and accelerometer data were used as implicit indicators of social interactions [64], and the number of contacts was used as an indicator of social network size [65]. It is entirely unclear how well these features map to what they are supposed to measure.

In addition, the studies were often vague about which social behavior they aimed to capture with the use of passive measures. They often divided their features into “mobility” and “social” features and did not explicitly indicate which part of the social environment passive sensors aimed to capture [48,100]. Depending on the research question, this might not be a problem, for example, if the aim is mainly to predict particular outcomes. However, it hinders the understanding of which social mechanism shapes well-being when it is unclear which behavior is captured using passive sensing. Thus, we propose clearly defining which constructs passive measures aim to capture and validating those measures.

Some of the studies that already combined different measures using ESM and passive sensing together showed that there is overlap between the features created through passive sensing and the variables assessed via ESM, which is another way to validate passive measures. However, these studies found that features differ across studies and individuals (eg, people with schizophrenia vs healthy controls [81]). Thus, we cannot draw any final conclusions about which ESM questionnaires are captured using passive sensing. For future research that aims to use passive sensing to capture the changing social environment, it is important to investigate whether passive sensing can be used to capture aspects of the social environment other than call and SMS text message behavior and, if so, to identify which features can be used.

Limitations

Although there are several strengths to this review, including its contribution to describing a way forward for how the social environment can be measured from a holistic perspective and the inclusion of many studies from multiple databases, some limitations warrant discussion. A limitation was that only 1 researcher screened the papers. Thus, especially the coding of which aspects of the social environment were captured using each method was not checked for interrater reliability. Nevertheless, the categories were created as an iterative process, and specific cases were discussed with all the researchers. We described each category and chose example items to increase the objective categorization of ESM questions, egocentric network questions, and passive sensors. Given that our interest lies in broadly describing the use of different methods to capture the social environment, minor changes in coding would not likely affect our general conclusions.

Regarding passive sensing, there were several studies that did not aim to measure the social environment but to investigate more generally how passive measures relate to well-being. Our review only included studies that explicitly aimed to use passive measures that related to the broader social environment of a person. Thus, studies that captured parts of the social environment using passive measures but did not explicitly mention this were not included in this review. These studies can be found in systematic reviews that cover the relationship between passive sensors and well-being more generally [36,37,101,102]. Recently, a new method has also evolved (ie, screenomics [103]) that captures the digital behavior of a person on their smartphone passively via multiple screenshots. As this method mainly focuses on the digital environment of a person, it was not included in our review, which aimed to capture the broad social environment of a person.

In our scoping review, we focused on studies that measured the social environment in relation to well-being in daily life. Thus, we excluded studies that measured the social environment only once or that primarily focused on medical conditions (eg, stroke) or only on dyadic relationships. We also excluded studies that only focused on children and adolescents or older people. For future research, it might be interesting to look at such medical studies and studies with children and adolescents or older people and investigate how they measured the social environment. In these groups, the social environment is similarly important for well-being.

Conclusions

Social interactions are important for well-being, which is widely recognized by researchers from different fields. Many different disciplines have developed measures that are able to capture changes in the social environment. In this study, we synthesized the literature on how different measures were used to capture the dynamic social environment in relation to well-being. Our results indicate that the combination of measures is currently lacking but that a combination is important to capture the social environment from a holistic perspective. Therefore, we propose combining these methods more often to reduce researcher and participant burden to improve data quality and cover more aspects of the changing social environment, which is needed to better predict a change in well-being. Finally, we call for more research that validates the measures used to capture aspects of the social environment.

Acknowledgments

The authors wish to thank Markus I Eronen, who helped us proofread the manuscript. GS was supported by the Netherlands Organisation for Scientific Research (Veni grant 451-15-034). LFB was supported by a Netherlands Organization for Scientific Research Veni grant (NWO-Veni 191G.037). The funding provider had no role in the study design; collection, analysis, or interpretation of the data; writing of the manuscript; or decision to submit the paper for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File , 32 KB - mental_v10i1e42646_app1.docx]

Multimedia Appendix 2
Literature search strategy.
[DOCX File , 20 KB - mental_v10i1e42646_app2.docx]

Multimedia Appendix 3
Overview of included studies.
[XLSX File (Microsoft Excel File), 1561 KB - mental_v10i1e42646_app3.xlsx]

References

1. Dodge R, Daly A, Huyton J, Sanders L. The challenge of defining wellbeing. Int J Wellbeing 2012 Aug 28;2(3):222-235 [FREE Full text] [doi: 10.5502/ijw.v2i3.4]
2. Huang M, Huang W. Innovative Approaches of Data Visualization and Visual Analytics. Pennsylvania, United States: IGI Global; 2014.
3. Kendler KS, Myers J, Prescott CA. Sex differences in the relationship between social support and risk for major depression: a longitudinal study of opposite-sex twin pairs. Am J Psychiatry 2005 Feb;162(2):250-256. [doi: 10.1176/appi.ajp.162.2.250] [Medline: 15677587]
4. Ge L, Yap CW, Ong R, Heng BH. Social isolation, loneliness and their relationships with depressive symptoms: a population-based study. PLoS One 2017 Aug 23;12(8):e0182145 [FREE Full text] [doi: 10.1371/journal.pone.0182145] [Medline: 28832594]
5. McElroy E, McIntyre JC, Bentall RP, Wilson T, Holt K, Kullu C, et al. Mental health, deprivation, and the neighborhood social environment: a network analysis. Clin Psychol Sci 2019 Mar 26;7(4):719-734. [doi: 10.1177/2167702619830640]
6. Sliedrecht W, de Waart R, Witkiewitz K, Roozen HG. Alcohol use disorder relapse factors: a systematic review. Psychiatry Res 2019 Aug;278:97-115. [doi: 10.1016/j.psychres.2019.05.038] [Medline: 31174033]
7. Teo AR, Choi H, Valenstein M. Social relationships and depression: ten-year follow-up from a nationally representative study. PLoS One 2013;8(4):e62396 [FREE Full text] [doi: 10.1371/journal.pone.0062396] [Medline: 23646128]
8. Krach S, Paulus FM, Bodden M, Kircher T. The rewarding nature of social interactions. Front Behav Neurosci 2010;4:22 [FREE Full text] [doi: 10.3389/fnbeh.2010.00022] [Medline: 20577590]
9. Bailey T, Shahabi L, Tarvainen M, Shapiro D, Ottaviani C. Moderating effects of the valence of social interaction on the dysfunctional consequences of perseverative cognition: an ecological study in major depression and social anxiety disorder. Anxiety Stress Coping 2019 Mar 22;32(2):179-195. [doi: 10.1080/10615806.2019.1570821] [Medline: 30667270]
10. Delespaul P, deVries M, van Os J. Determinants of occurrence and recovery from hallucinations in daily life. Soc Psychiatry Psychiatr Epidemiol 2002 Mar 8;37(3):97-104. [doi: 10.1007/s001270200000] [Medline: 11990012]
11. Kuppens P, Verduyn P. Emotion dynamics. Curr Opin Psychol 2017 Oct;17:22-26. [doi: 10.1016/j.copsyc.2017.06.004] [Medline: 28950968]
12. Kubey R, Larson R, Csikszentmihalyi M. Experience sampling method applications to communication research questions. J Commun 1996;46(2):99-120. [doi: 10.1111/j.1460-2466.1996.tb01476.x]
13. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
14. Mehl MR. The lay assessment of subclinical depression in daily life. Psychol Assess 2006 Sep;18(3):340-345. [doi: 10.1037/1040-3590.18.3.340] [Medline: 16953737]
15. Perry B, Pescosolido B, Borgatti S. Egocentric Network Analysis Foundations, Methods, and Models. Cambridge, Massachusetts, United States: Cambridge University Press; 2018.

16. Santangelo P, Ebner-Priemer U, Trull T. 11 experience sampling methods in clinical psychology. In: The Oxford Handbook of Research Strategies for Clinical Psychology. Oxford, United Kingdom: OUP USA; 2013.
17. Eskes P, Spruit M, Brinkkemper S, Vorstman J, Kas MJ. The sociability score: app-based social profiling from a healthcare perspective. Comput Human Behav 2016 Jun;59:39-48. [doi: 10.1016/j.chb.2016.01.024]
18. Jaques N, Taylor S, Azaria A, Ghandeharioun A, Sano A, Picard R. Predicting students' happiness from physiology, phone, mobility, and behavioral data. In: Proceedings of the International Conference on Affective Computing and Intelligent Interaction and Workshops. 2015 Presented at: International Conference on Affective Computing and Intelligent Interaction and Workshops; Sep 21 - 24, 2015; Xi'an, China. [doi: 10.1109/ACII.2015.7344575]
19. Mehl MR, Pennebaker JW. The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations. J Pers Soc Psychol 2003 Apr;84(4):857-870. [doi: 10.1037/0022-3514.84.4.857] [Medline: 12703653]
20. Pennebaker J, Booth R, Boyd R, Francis M. Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates; 2015.
21. Minor KS, Davis BJ, Marggraf MP, Luther L, Robbins ML. Words matter: implementing the electronically activated recorder in schizotypy. Personal Disord 2018 Mar;9(2):133-143. [doi: 10.1037/per0000266] [Medline: 29215902]
22. Perry BL, Pescosolido BA. Social network dynamics and biographical disruption: the case of “first-timers” with mental illness. Am J Sociol 2012 Jul;118(1):134-175. [doi: 10.1086/666377]
23. Francis MW. Transitions of women's substance use recovery networks and 12-month sobriety outcomes. Soc Networks 2020 Oct;63:1-10 [FREE Full text] [doi: 10.1016/j.socnet.2020.04.003] [Medline: 32675917]
24. Stadel M, Stulp G. Balancing bias and burden in personal network studies. Soc Networks 2022 Jul;70:16-24. [doi: 10.1016/j.socnet.2021.10.007]
25. Haines VA, Beggs JJ, Hurlbert JS. Contextualizing health outcomes: do effects of network structure differ for women and men? Sex Roles 2008 Apr 20;59(3-4):164-175. [doi: 10.1007/s11199-008-9441-3]
26. McCarty C, Killworth PD, Rennell J. Impact of methods for reducing respondent burden on personal network structural measures. Soc Networks 2007 May;29(2):300-315. [doi: 10.1016/j.socnet.2006.12.005]
27. Stulp G. Collecting large personal networks in a representative sample of Dutch women. Soc Networks 2021 Jan;64:63-71. [doi: 10.1016/j.socnet.2020.07.012]
28. Harpham T, Grant E, Thomas E. Measuring social capital within health surveys: key issues. Health Policy Plan 2002 Mar;17(1):106-111. [doi: 10.1093/heapol/17.1.106] [Medline: 11861592]
29. Phongsavan P, Chey T, Bauman A, Brooks R, Silove D. Social capital, socio-economic status and psychological distress among Australian adults. Soc Sci Med 2006 Nov;63(10):2546-2561. [doi: 10.1016/j.socscimed.2006.06.021] [Medline: 16914244]
30. Bringmann LF, Elmer T, Eronen MI. Back to basics: the importance of conceptual clarification in psychological science. Curr Dir Psychol Sci 2022 Jun 14;31(4):340-346. [doi: 10.1177/09637214221096485]
31. Eronen MI, Bringmann LF. The theory crisis in psychology: how to move forward. Perspect Psychol Sci 2021 Jul 29;16(4):779-788 [FREE Full text] [doi: 10.1177/1745691620970586] [Medline: 33513314]
32. Scheel AM, Tiokhin L, Isager PM, Lakens D. Why hypothesis testers should spend less time testing hypotheses. Perspect Psychol Sci 2021 Jul 16;16(4):744-755 [FREE Full text] [doi: 10.1177/1745691620966795] [Medline: 33326363]
33. Flake JK, Fried EI. Measurement schmeasurement: questionable measurement practices and how to avoid them. Advan Method Pract Psychol Sci 2020 Dec 02;3(4):456-465. [doi: 10.1177/2515245920952393]
34. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
35. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
36. Benoit J, Onyeaka H, Keshavan M, Torous J. Systematic review of digital phenotyping and machine learning in psychosis spectrum illnesses. Harv Rev Psychiatry 2020 Aug 12;28(5):296-304. [doi: 10.1097/hrp.0000000000000268]
37. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
38. Agampodi T, Agampodi SB, Glozier N, Siribaddana S. Measurement of social capital in relation to health in low and middle income countries (LMIC): a systematic review. Soc Sci Med 2015 Mar;128:95-104. [doi: 10.1016/j.socscimed.2015.01.005] [Medline: 25596373]
39. Fernandes A, Van Lenthe FJ, Vallée J, Sueur C, Chaix B. Linking physical and social environments with mental health in old age: a multisensor approach for continuous real-life ecological and emotional assessment. J Epidemiol Community Health 2021 May 04;75(5):477-483 [FREE Full text] [doi: 10.1136/jech-2020-214274] [Medline: 33148684]
40. Lawlor ER, Cupples ME, Donnelly M, Tully MA. Promoting physical activity among community groups of older women in socio-economically disadvantaged areas: randomised feasibility study. Trials 2019 Apr 25;20(1):234 [FREE Full text] [doi: 10.1186/s13063-019-3312-9] [Medline: 31023329]
41. Steptoe A, Leigh ES, Kumari M. Positive affect and distressed affect over the day in older people. Psychol Aging 2011 Dec;26(4):956-965. [doi: 10.1037/a0023303] [Medline: 21517182]

42. O'Donohue JS, Mesagno C, O'Brien B. How can stress resilience be monitored? A systematic review of measurement in humans. Curr Psychol 2019 Mar 14;40(6):2853-2876. [doi: 10.1007/s12144-019-00226-9]
43. Čolić J, Bassett TR, Latysheva A, Imboden C, Bader K, Hatzinger M, et al. Depersonalization and derealization in embarrassing social interactions: an experience sampling study in social phobia, major depression and controls. J Anxiety Disord 2020 Mar;70:102189. [doi: 10.1016/j.janxdis.2020.102189] [Medline: 32070861]
44. Mills J, Fuller-Tyszkiewicz M, Holmes M. State body dissatisfaction and social interactions. Psychol Women Q 2014 Feb 03;38(4):551-562. [doi: 10.1177/0361684314521139]
45. Schneider M, Reininghaus U, van Nierop M, Janssens M, Myin-Germeys I, GROUP Investigators. Does the social functioning scale reflect real-life social functioning? An experience sampling study in patients with a non-affective psychotic disorder and healthy control individuals. Psychol Med 2017 Dec 23;47(16):2777-2786. [doi: 10.1017/S0033291717001295] [Medline: 28534464]
46. Brown L, Strauman T, Barrantes-Vidal N, Silvia P, Kwapil T. An experience-sampling study of depressive symptoms and their social context. J Nerv Ment Dis 2011 Jun;199(6):403-409. [doi: 10.1097/NMD.0b013e31821cd24b] [Medline: 21629020]
47. ter Haar-Pomp L, Spreen M, Volker B, Bogaerts S. The impact of forced forensic psychiatric confinement on composition and structure in the personal networks of personality-disordered forensic psychiatric patients. J Forensic Psychiatry Psychol 2018 May 08;30(1):53-75. [doi: 10.1080/14789949.2018.1467947]
48. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666 [FREE Full text] [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]
49. Jagesar RR, Roozen MC, van der Heijden I, Ikani N, Tyborowska A, Penninx BW, et al. Digital phenotyping and the COVID-19 pandemic: capturing behavioral change in patients with psychiatric disorders. Eur Neuropsychopharmacol 2021 Jan;42:115-120. [doi: 10.1016/j.euroneuro.2020.11.012] [Medline: 33298386]
50. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, et al. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e13209 [FREE Full text] [doi: 10.2196/13209] [Medline: 31342903]
51. Sun J, Harris K, Vazire S. Is well-being associated with the quantity and quality of social interactions? J Pers Soc Psychol 2020 Dec;119(6):1478-1496. [doi: 10.1037/pspp0000272] [Medline: 31647273]
52. Borsboom D. Measuring the Mind Conceptual Issues in Contemporary Psychometrics. Cambridge, Massachusetts, United States: Cambridge University Press; 2005.
53. Bringmann LF, Eronen MI. Heating up the measurement debate: what psychologists can learn from the history of physics. Theory Psychol 2015 Nov 22;26(1):27-43. [doi: 10.1177/0959354315617253]
54. Lewis K. COVID-19: preliminary data on the impact of social distancing on loneliness and mental health. J Psychiatr Pract 2020 Sep;26(5):400-404. [doi: 10.1097/PRA.0000000000000488] [Medline: 32936586]
55. Neto F. Psychometric analysis of the short-form UCLA Loneliness Scale (ULS-6) in older adults. Eur J Ageing 2014 Dec 25;11(4):313-319 [FREE Full text] [doi: 10.1007/s10433-014-0312-1] [Medline: 28804337]
56. Merolla AJ, Hall JA, Bernhold Q. Perseverative cognition, distracted communication, and well‐being in everyday social interaction. Pers Relationship 2019 Aug 20;26(3):507-528. [doi: 10.1111/pere.12286]
57. Hepp J, Lane SP, Carpenter RW, Trull TJ. Linking daily-life interpersonal stressors and health problems via affective reactivity in borderline personality and depressive disorders. Psychosom Med 2020 Jan 8;82(1):90-98 [FREE Full text] [doi: 10.1097/PSY.0000000000000728] [Medline: 31592936]
58. Costenbader E, Valente TW. The stability of centrality measures when networks are sampled. Soc Networks 2003 Oct;25(4):283-307. [doi: 10.1016/S0378-8733(03)00012-1]
59. Marin A, Hampton K. Network instability in times of stability. Sociol Forum 2019 Feb 14;34(2):313-336 [FREE Full text] [doi: 10.1111/socf.12499]
60. Terhell EL, Broese van Groenou MI, van Tilburg T. Network contact changes in early and later postseparation years. Soc Networks 2007 Jan;29(1):11-24. [doi: 10.1016/j.socnet.2005.11.006]
61. Campbell KE, Lee BA. Name generators in surveys of personal networks. Soc Networks 1991 Sep;13(3):203-221. [doi: 10.1016/0378-8733(91)90006-F]
62. Pinfold V, Sweet D, Porter I, Quinn C, Byng R, Griffiths C, et al. Improving community health networks for people with severe mental illness: a case study investigation. In: Health Services and Delivery Research. Southampton (UK): NIHR Journals Library; Feb 2015.
63. Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. J Med Internet Res 2020 May 29;22(5):e16875 [FREE Full text] [doi: 10.2196/16875] [Medline: 32348284]
64. Tsapeli F, Musolesi M. Investigating causality in human behavior from smartphone sensor data: a quasi-experimental approach. EPJ Data Sci 2015 Dec 18;4(24). [doi: 10.1140/epjds/s13688-015-0061-1]

65. Schuwerk T, Kaltefleiter LJ, Au J, Hoesl A, Stachl C. Enter the wild: autistic traits and their relationship to mentalizing and social interaction in everyday life. J Autism Dev Disord 2019 Oct 4;49(10):4193-4208. [doi: 10.1007/s10803-019-04134-6] [Medline: 31273579]
66. Baddeley JL, Pennebaker JW, Beevers CG. Everyday social behavior during a major depressive episode. Soc Psychol Personal Sci 2012 Oct 08;4(4):445-452. [doi: 10.1177/1948550612461654]
67. Di Matteo D, Fotinos K, Lokuge S, Yu J, Sternat T, Katzman MA, et al. The relationship between smartphone-recorded environmental audio and symptomatology of anxiety and depression: exploratory study. JMIR Form Res 2020 Aug 13;4(8):e18751 [FREE Full text] [doi: 10.2196/18751] [Medline: 32788153]
68. Beltzer ML, Ameko MK, Daniel KE, Daros AR, Boukhechba M, Barnes LE, et al. Building an emotion regulation recommender algorithm for socially anxious individuals using contextual bandits. Br J Clin Psychol 2022 Jan 14;61 Suppl 1(S1):51-72. [doi: 10.1111/bjc.12282] [Medline: 33583059]
69. Dunton GF, Liao Y, Intille S, Huh J, Leventhal A. Momentary assessment of contextual influences on affective response during physical activity. Health Psychol 2015 Dec;34(12):1145-1153 [FREE Full text] [doi: 10.1037/hea0000223] [Medline: 26053885]
70. Määttänen I, Henttonen P, Väliaho J, Palomäki J, Thibault M, Kallio J, et al. Positive affect state is a good predictor of movement and stress: combining data from ESM/EMA, mobile HRV measurements and trait questionnaires. Heliyon 2021 Feb;7(2):e06243 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e06243] [Medline: 33681494]
71. McCormick B, Frey G, Lee C, Gajic T, Stamatovic-Gajic B, Maksimovic M. A pilot examination of social context and everyday physical activity among adults receiving Community Mental Health Services. Acta Psychiatr Scand 2009 Mar;119(3):243-247 [FREE Full text] [doi: 10.1111/j.1600-0447.2008.01331.x] [Medline: 19133878]
72. Abel DB, Salyers MP, Wu W, Monette MA, Minor KS. Quality versus quantity: determining real-world social functioning deficits in schizophrenia. Psychiatry Res 2021 Jul;301:113980 [FREE Full text] [doi: 10.1016/j.psychres.2021.113980] [Medline: 33979764]
73. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
74. Charles ST, Pasupathi M. Age-related patterns of variability in self-descriptions: implications for everyday affective experience. Psychol Aging 2003 Sep;18(3):524-536. [doi: 10.1037/0882-7974.18.3.524] [Medline: 14518813]
75. Pelt DH, van der Linden D, Dunkel CS, Born MP. The general factor of personality and daily social experiences: evidence for the social effectiveness hypothesis. Personal Indiv Diff 2020 Mar;155:109738. [doi: 10.1016/j.paid.2019.109738]
76. Harlow RE, Cantor N. To whom do people turn when things go poorly? Task orientation and functional social contacts. J Pers Soc Psychol 1995 Aug;69(2):329-340. [doi: 10.1037//0022-3514.69.2.329] [Medline: 7643308]
77. Lin XY, Lachman ME. Daily stress and affect across adulthood: the role of social interactions via different communication modes. Technol Mind Behav 2021 Feb 16;2(1) [FREE Full text] [doi: 10.1037/tmb0000026] [Medline: 35369392]
78. Oloritun RO, Ouarda TB, Moturu S, Madan A, Pentland A, Khayal I. Change in BMI accurately predicted by social exposure to acquaintances. PLoS One 2013 Nov 20;8(11):e79238 [FREE Full text] [doi: 10.1371/journal.pone.0079238] [Medline: 24278122]
79. Abdullah S, Matthews M, Frank E, Doherty G, Gay G, Choudhury T. Automatic detection of social rhythms in bipolar disorder. J Am Med Inform Assoc 2016 May;23(3):538-543. [doi: 10.1093/jamia/ocv200] [Medline: 26977102]
80. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
81. Fulford D, Mote J, Gonzalez R, Abplanalp S, Zhang Y, Luckenbaugh J, et al. Smartphone sensing of social interactions in people with and without schizophrenia. J Psychiatr Res 2021 May;137:613-620 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.11.002] [Medline: 33190842]
82. Stanislaus S, Vinberg M, Melbye S, Frost M, Busk J, Bardram JE, et al. Smartphone-based activity measurements in patients with newly diagnosed bipolar disorder, unaffected relatives and control individuals. Int J Bipolar Disord 2020 Nov 02;8(1):32 [FREE Full text] [doi: 10.1186/s40345-020-00195-0] [Medline: 33135120]
83. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 1959;56(2):81-105. [doi: 10.1037/h0046016]
84. Hopwood C, Bleidorn W, Wright A. Connecting theory to methods in longitudinal research. Perspect Psychol Sci 2022 May;17(3):884-894. [doi: 10.1177/17456916211008407] [Medline: 34739347]
85. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, et al. The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. Assessment 2022 Mar;29(2):136-151. [doi: 10.1177/1073191120957102] [Medline: 32909448]
86. Stadel M. Integrating personal networks into experience sampling. In L. F. Bringmann (Chair), Investigating the social context in daily life Symposium. 2021 Jul 01 Presented at: Society for Ambulatory Assessment Conference; July 1, 2021; Zurich, Switzerland.

87. Helbig-Lang S, Lang T, Petermann F, Hoyer J. Anticipatory anxiety as a function of panic attacks and panic-related self-efficacy: an ambulatory assessment study in panic disorder. Behav Cogn Psychother 2012 Oct 29;40(5):590-604. [doi: 10.1017/S1352465812000057] [Medline: 22373714]
88. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
89. Zinn S, Gnambs T. Analyzing nonresponse in longitudinal surveys using Bayesian additive regression trees: a nonparametric event history analysis. Soc Sci Comput Rev 2020 Jun 04;40(3):678-699. [doi: 10.1177/0894439320928242]
90. Jongs N, Jagesar R, van Haren NE, Penninx BW, Reus L, Visser PJ, et al. A framework for assessing neuropsychiatric phenotypes by using smartphone-based location data. Transl Psychiatry 2020 Jul 01;10(1):211 [FREE Full text] [doi: 10.1038/s41398-020-00893-4] [Medline: 32612118]
91. Mestdagh M, Dejonckheere E. Ambulatory assessment in psychopathology research: current achievements and future ambitions. Curr Opin Psychol 2021 Oct;41:1-8. [doi: 10.1016/j.copsyc.2021.01.004] [Medline: 33550191]
92. Wright AG, Zimmermann J. Applied ambulatory assessment: integrating idiographic and nomothetic principles of measurement. Psychol Assess 2019 Dec;31(12):1467-1480 [FREE Full text] [doi: 10.1037/pas0000685] [Medline: 30896209]
93. Kirtley O, Hiekkaranta A, Kunkels Y, Eisele G, Lüken M, Verhoeven D, et al. The Experience Sampling Method (ESM) item repository. OSF Home. 2018. URL: https://osf.io/kg376/ [accessed 2022-12-21]
94. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. Increasing transparency through a multiverse analysis. Perspect Psychol Sci 2016 Sep 29;11(5):702-712. [doi: 10.1177/1745691616658637] [Medline: 27694465]
95. Castro-Alvarez S, Tendeiro J, de Jonge P, Meijer R, Bringmann L. Mixed-effects trait-state-occasion model: studying the psychometric properties and the person–situation interactions of psychological dynamics. Structural Equation Model Multidisciplinary J 2021 Oct 12;29(3):438-451 [FREE Full text] [doi: 10.1080/10705511.2021.1961587]
96. Cranford JA, Shrout PE, Iida M, Rafaeli E, Yip T, Bolger N. A procedure for evaluating sensitivity to within-person change: can mood measures in diary studies detect change reliably? Pers Soc Psychol Bull 2006 Jul 02;32(7):917-929 [FREE Full text] [doi: 10.1177/0146167206287721] [Medline: 16738025]
97. Geldhof GJ, Preacher KJ, Zyphur MJ. Reliability estimation in a multilevel confirmatory factor analysis framework. Psychol Method 2014 Mar;19(1):72-91. [doi: 10.1037/a0032138] [Medline: 23646988]
98. Schuurman NK, Hamaker EL. Measurement error and person-specific reliability in multilevel autoregressive modeling. Psychol Methods 2019 Feb;24(1):70-91. [doi: 10.1037/met0000188] [Medline: 30188157]
99. Shrout P, Lane S. Psychometrics. In: Handbook of Research Methods for Studying Daily Life. New York, NY: The Guilford Press; 2012.
100. Buck B, Scherer E, Brian R, Wang R, Wang W, Campbell A, et al. Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. Schizophr Res 2019 Jun;208:167-172 [FREE Full text] [doi: 10.1016/j.schres.2019.03.014] [Medline: 30940400]
101. Antosik-Wójcińska AZ, Dominiak M, Chojnacka M, Kaczmarek-Majer K, Opara KR, Radziszewska W, et al. Smartphone as a monitoring tool for bipolar disorder: a systematic review including data analysis, machine learning algorithms and predictive modelling. Int J Med Inform 2020 Jun;138:104131. [doi: 10.1016/j.ijmedinf.2020.104131] [Medline: 32305023]
102. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. J Med Internet Res 2017 Jul 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
103. Ram N, Yang X, Cho M, Brinberg M, Muirhead F, Reeves B, et al. Screenomics: a new approach for observing and studying individuals' digital lives. J Adolesc Res 2020 Jan;35(1):16-50 [FREE Full text] [doi: 10.1177/0743558419883362] [Medline: 32161431]

Abbreviations

EAR: electronically activated recorder
ESM: experience sampling method
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by J Torous; submitted 12.09.22; peer-reviewed by N Thomas, H Jelinek, M Kapsetaki; comments to author 26.10.22; revised version received 21.12.22; accepted 02.01.23; published 17.03.23.

Please cite as:
Langener AM, Stulp G, Kas MJ, Bringmann LF
Capturing the Dynamics of the Social Environment Through Experience Sampling Methods, Passive Sensing, and Egocentric Networks: Scoping Review
JMIR Ment Health 2023;10:e42646
URL: https://mental.jmir.org/2023/1/e42646
doi:10.2196/42646
PMID:36930210

Review

Effectiveness of Digital Mental Health Tools to Reduce Depressive and Anxiety Symptoms in Low- and Middle-Income Countries: Systematic Review and Meta-analysis

Jiyeong Kim[1], BPharm, MPH; Lois M D Aryee[2], MPH; Heejung Bang[3], MSc, PhD; Steffi Prajogo[4], BSc; Yong K Choi[5], MPH, PhD; Jeffrey S Hoch[6], MA, PhD; Elizabeth L Prado[7], MSc, PhD

[1]Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, CA, United States

[2]Department of Nutrition and Food Science, University of Ghana, Accra, Ghana

[3]Division of Biostatistics, Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, CA, United States

[4]Johns Hopkins Bayview Medical Center, Baltimore, MD, United States

[5]Department of Health Information Management, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, United States

[6]Division of Health Policy and Management, Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, CA, United States

[7]Department of Nutrition, Institute for Global Nutrition, University of California, Davis, Davis, CA, United States

Corresponding Author:
Jiyeong Kim, BPharm, MPH
Department of Public Health Sciences
School of Medicine
University of California, Davis
1 Shields Avenue
Davis, CA, 95616
United States
Phone: 1 5307522793
Email: jiykim@ucdavis.edu

Abstract

Background: Depression and anxiety contribute to an estimated 74.6 million years of life with disability, and 80% of this burden occurs in low- and middle-income countries (LMICs), where there is a large gap in care.

Objective: We aimed to systematically synthesize available evidence and quantify the effectiveness of digital mental health interventions in reducing depression and anxiety in LMICs.

Methods: In this systematic review and meta-analysis, we searched PubMed, Embase, and Cochrane databases from the inception date to February 2022. We included randomized controlled trials conducted in LMICs that compared groups that received digital health interventions with controls (active control, treatment as usual, or no intervention) on depression or anxiety symptoms. Two reviewers independently extracted summary data reported in the papers and performed study quality assessments. The outcomes were postintervention measures of depression or anxiety symptoms (Hedges *g*). We calculated the pooled effect size weighted by inverse variance.

Results: Among 11,196 retrieved records, we included 80 studies in the meta-analysis (12,070 participants n=6052, 50.14% in the intervention group and n=6018, 49.85% in the control group) and 96 studies in the systematic review. The pooled effect sizes were −0.61 (95% CI −0.78 to −0.44; n=67 comparisons) for depression and −0.73 (95% CI −0.93 to −0.53; n=65 comparisons) for anxiety, indicating that digital health intervention groups had lower postintervention depression and anxiety symptoms compared with controls. Although heterogeneity was considerable (I^2=0.94 for depression and 0.95 for anxiety), we found notable sources of variability between the studies, including intervention content, depression or anxiety symptom severity, control type, and age. Grading of Recommendations, Assessments, Development, and Evaluation showed that the evidence quality was overall high.

Conclusions: Digital mental health tools are moderately to highly effective in reducing depression and anxiety symptoms in LMICs. Thus, they could be effective options to close the gap in depression and anxiety care in LMICs, where the usual mental health care is minimal.

Trial Registration: PROSPERO CRD42021289709; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=289709

(JMIR Ment Health 2023;10:e43066) doi:10.2196/43066

KEYWORDS

digital mental health; mHealth; mobile health; digital health; low- and middle-income country; depression; anxiety; mobile phone

Introduction

Background

Depressive and anxiety disorders are a leading cause of the global burden of disease [1]. Depression and anxiety contribute to an estimated 74.6 million years lived with disability, and 80% of this burden occurs in low- and middle-income countries (LMICs) [2]. However, the investment in mental health disorder prevention and treatment is substantially lower in LMICs than in high-income countries (HICs; $2 vs $50 per person in HICs) [3]. In LMICs, 80% to 95% of people with depression and anxiety do not receive the necessary mental health care, mainly because of the limited availability of service providers [4,5]. This large treatment gap could lead to detrimental consequences for the overall health of individuals with mental disorders, social and economic burden on their families, and large-scale societal loss in terms of decreased economic productivity owing to missed work (absenteeism) or reduced efficiency at work (presenteeism) [4,6].

In recent decades, the use of digital mental health (DMH) services to deliver mental health care has been increasing through the internet and other forms of technologies [7,8]. In this paper, we use the term DMH tools to refer to the following: either a digital platform as a tool to deliver mental health care (eg, cognitive behavioral therapy [CBT] provided by a service provider remotely via computer) or a digital platform itself as a main mental health intervention (eg, smartphone apps for depression). DMH tools have shown effectiveness among various populations in HICs, including youth, adults, older adults, and antenatal and postpartum women, in reducing conditions, such as mild to moderate or severe depression, social anxiety, panic disorder, suicidal ideation, posttraumatic stress disorder, attention-deficit/hyperactivity disorder, or insomnia [7,8]. A growing body of literature has started investigating barriers and facilitators for the successful implementation [9], usability, and acceptability to enhance user engagement with DMH tools [10,11]. Economic evaluations of DMH tools have reported that they are cost-effective [12,13]. These investigations informed us that well-designed DMH interventions could enhance access to quality mental health care with low-cost investment, which is ideal for LMICs. Indeed, the World Health Organization recognized the potential role of technology-supported mental health care tools in closing mental health treatment gaps in LMICs [14]. As the technology-enabling environment is expanding in LMICs (eg, 90% mobile phone penetration rate and 40% average internet connectivity) [15-17], diverse DMH interventions have been tested [18].

Prior Work

A few systematic reviews examined digital health interventions for mental disorders in LMICs and reported several limitations [19], including that the quality of studies was suboptimal [20] and most studies reported short-term follow-ups and a low retention rate [21]. Thus, the findings were inconclusive in determining the clinical impact of DMH interventions. A recent meta-analysis reported a moderate effect of digital psychological interventions on mental disorders in LMICs [22]. However, the results were not explicit for depression or anxiety, which are the most prevalent forms of mental health issues [1]. Of note, this study was restricted to adult populations. This is a notable limitation given that (1) adolescents and children have escalating mental health care needs and (2) technology-enabled mental health care delivery is a promising strategy for these age groups [23,24]. Moreover, under the unprecedented SARS/COVID-19 pandemic, technology-based mental health interventions have surged, and overall mental distress was elevated in LMICs [25]. Hence, we need to investigate the effectiveness of DMH tools on depression and anxiety in acquiring comprehensive knowledge of this promising strategy in depression and anxiety care in low-resource settings. Therefore, this study aimed to quantify the effectiveness of digital health tools in reducing depression and anxiety in adults, adolescents, and children in LMICs.

Methods

Search Strategy and Study Selection

We searched the PubMed, Embase, and Cochrane databases for papers without language restrictions from database inception to February 22, 2022. We adapted our search strategy from previous reviews and settings [22,26] and refined it further to tailor it to our review purpose. Search terms included text, keywords, and Medical Subject Headings for PubMed/Medline and Cochrane databases, and the expansion (/exp) function was applied in Embase in the following three main areas: (1) technology; (2) mental health (depression or anxiety); and (3) LMICs, based on the World Bank Country Classification of the year the study was conducted. The full search terms for each database are provided in Multimedia Appendix 1.

During the full-text review, a snowball search was applied to find relevant studies from the references of previous systematic reviews. In addition, papers suggested by the citation program (Mendeley) were screened. We used the Google Translate software for data extraction and quality assessment of papers published in non-English languages. We also screened papers on ClinicalTrials.gov and the International Clinical Trials Registry Platform via the Cochrane Library to find unpublished clinical trials.

Studies were included if they (1) were randomized controlled trials (RCTs), (2) used technology (eg, computer, tablet PC, internet, mobile app, telephone, texting, or video or audio files) either as a tool to deliver traditional mental health care or the technology itself was a main mental health intervention, (3) measured depression or anxiety as either a primary or secondary outcome, (4) targeted people with low to moderate severity of depression or anxiety (eg, from lightly symptomatic to moderate disorder), (5) were conducted in an LMIC, and (6) used any comparison group (eg, active control, treat as usual, no intervention, or waitlist control).

Widely used standardized programs were used for abstract screening (Covidence) and full-text review (Microsoft Excel). Two reviewers (JK and SP) independently screened the abstracts. JK extracted and double-checked the data. SP independently extracted the data from 20% of randomly selected papers. If there were disagreements during screening and full-text review for inclusion, a third reviewer (HB) resolved the conflict. We applied the same inclusion and exclusion criteria to both the systematic review and the meta-analysis. The primary reviewer (JK) imported the extracted data into the summary table, and another reviewer (LMDA) checked the data in the summary table for accuracy. All the coauthors reviewed the data in the summary table. The biostatistician (HB) checked a random sample of the final data for effect size calculation.

Data Analysis

The authors extracted the study population (demographics), study aims, settings, region, inclusion and exclusion criteria, interventions and controls (duration and details of procedures), outcome means, SDs, and measurement instruments. We contacted the corresponding authors to secure necessary data that were missing from published studies and excluded studies from the meta-analysis if the requested data were not provided. We evaluated an individual study's methodological quality and risk of bias using the Effective Public Health Practice Project quality assessment tool in the following domains: selection bias, study design, confounders, blinding, data collection methods, withdrawal and dropout, intervention integrity, and analyses. Three reviewers assessed the listed domains independently (JK, LMDA, and SP), and the assessments were compared between the two assessors. Each domain was rated as 1=strong, 2=moderate, and 3=weak, and a global rating was assigned based on the section ratings as follows: 1=strong (no weak ratings), 2=moderate (1 weak rating), and 3=weak (2 or more weak ratings). In case of conflict, assessors discussed the discrepancies and reconciled the ratings. See Multimedia Appendix 2 [26-121] for our Effective Public Health Practice Project assessment. The quality of evidence for the outcome across all studies was assessed using the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) criteria—risk of bias, inconsistency of effect, imprecision, indirectness, and publication bias (JK and LMDA). See Multimedia Appendix 3 for our GRADE assessment.

Stata 17 (StataCorp) was used to compute pooled effect sizes with 95% CIs. As the studies reported outcomes with different measurement tools, we calculated the Hedges *g* values with 95% CI as a standardized mean difference index to estimate the effect size. We chose Hedges *g* over Cohen *d*, because it is less prone to bias for the small sample studies, some of which were contained in our review [122]. Postintervention means and SDs were used for pooled effect size calculation. Weights were assigned to each study by calculating the inverse variance of the outcome scores. If more than one mental health measurement was reported, we included the one that was reported in the largest number of other studies. In the multiple arm studies, digital intervention was compared with each arm, and the effect size for each comparison was estimated. In this case, the frequency (N) of the control was divided by the number of comparisons to avoid overweighting the studies with multiple arms. We selected a random effect model to calculate pooled effect sizes, because we expected the studies to be dissimilar, whereas we wanted to generalize the results to other populations. Higgins and Thompson I^2 values were calculated to evaluate the heterogeneity across studies. We conducted subgroup analyses to investigate the variations in effect sizes by the following study characteristics: intervention content, type of technology use, mode, intervention duration, outcome measurement, depression or anxiety severity, control type, participant age, and study region. Furthermore, we performed a univariate, random-effect meta-regression using prespecified study characteristics (eg, study quality and blinding, in addition to the characteristics examined for the subgroup analysis) to look for notable sources of heterogeneity.

We conducted preplanned sensitivity analyses to test the robustness of the results for the studies where (1) the quality was poor, (2) more than one depression or anxiety score was reported, (3) the baseline psychometric scores were considerably different between the intervention and control groups, and (4) the digital tool was adjunct to the main nondigital intervention. We performed a post hoc sensitivity analysis, excluding outliers defined as the estimates' 95% CI values that did not overlap with the 95% CI values of the pooled effect size [123]. To assess small-study effects and publication bias, we visually examined funnel plots and performed Egger weighted regression test to quantitatively evaluate the degree of asymmetry. To calculate the bias-corrected overall effect sizes after accounting for the publication bias and small-study effects, the Duval and Tweedie trim and fill method was used. The study's review protocol was registered in PROSPERO (CRD42021289709), and the review process was compliant with the PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) guidelines, which can be found in Multimedia Appendix 4.

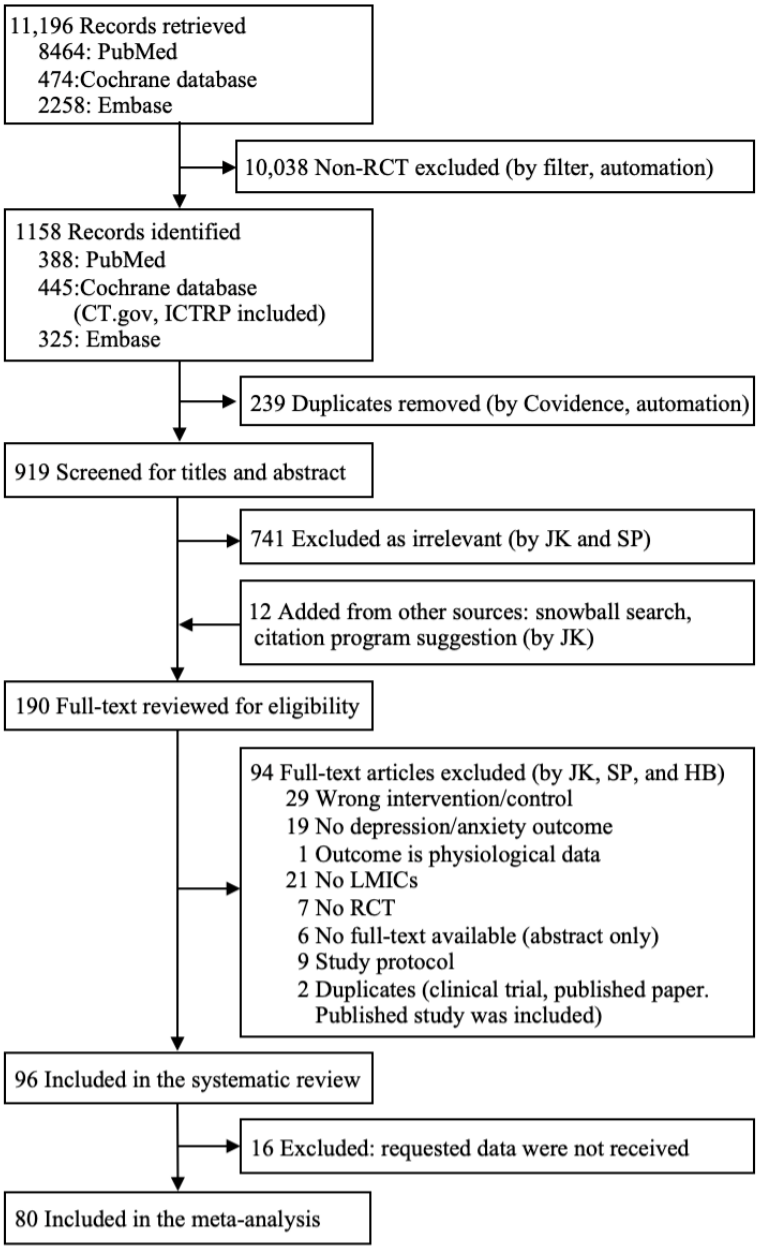
Results

Characteristics of the Included Studies

Initially, 11,196 records were retrieved and 1158 records were left after the RCT filter. When the records were imported to Covidence, 239 duplicates were removed automatically. We screened 919 studies for titles and abstracts and excluded 741 records as irrelevant. In addition, 12 studies were added from other sources; hence, a total of 190 studies were full-text reviewed and 94 studies were further excluded. We included 10 studies out of 22 from the recent meta-analysis, whereas the studies that were conducted in HICs and targeted posttraumatic

stress disorder or substance abuse without depression or anxiety measurements were excluded [22]. A total of 96 studies were selected for the systematic review, and 80 studies were included for the meta-analysis, because data for 16 studies were not available to calculate the postintervention effect size (Figure 1).

Figure 1. Study selection. CT.gov: ClinicalTrials.gov; ICTRP: International Clinical Trials Registry Platform; LMIC: low- and middle-income country; RCT: randomized controlled trial.



Multimedia Appendix 5 [26-121] shows the characteristics of collected data from each study included in this systematic review. More information about the included studies is available in Multimedia Appendix 6 [26-121]. The meta-analysis contained a total of 12,070 participants (n=6052, 50.14% in the technology intervention group and n=6018, 49.85% in the control group). The participants' age range was broad (4-75 years). The average number of participants per study was 144, ranging from 19 to 954. There were 59 and 54 studies reporting depression and anxiety outcomes, respectively. The measurement instruments were various, with 11 types for depression and 17 for anxiety. Symptoms were self-reported by the study participants. All 96 studies included in the systematic review were conducted between 2011 and 2021 in LMICs, including Asia (n=68), Africa (n=6), Europe (n=15), and Latin America (n=7). Studies published in Chinese and Portuguese were translated into English (n=3). In most studies (80/96, 83%), technology itself was a primary intervention, whereas technology was a tool to deliver intervention contents in 16 studies. Mobile apps and internet were the most common technology formats (80/96, 83%). Multiple contents (eg, psychotherapy plus peer support) were most frequently provided (33/96, 34%), followed by psychotherapy, including CBT (24/96, 25%). Playful distraction was primarily observed in relieving preprocedure anxiety (eg, dental treatment, venipuncture, and bone marrow aspiration), which was usually based on a single session and performed for children. The median intervention duration was 6 weeks, ranging from 1 day

to 72 weeks. Some studies (13/96, 14%) had more than 2 study groups (eg, internet-based CBT vs in-person CBT vs no intervention). Usual in-person care or active control (65/96, 68%) was more common than no intervention as a control. Only 15% (14/96) of the studies provided adverse event information.

The study quality assessment results revealed that 80% (77/96) of the studies were considered moderate-to-high quality. The randomization method was described in most reports (72/96, 75%), and 91% (87/96) were considered likely to have representative sample populations. Hence, the possibility of selection bias seemed low. Approximately 77% (74/96) showed less than 20% attrition rates, and 65% (62/96) performed intention-to-treat analysis to account for the missing data. We rated the overall quality of evidence as high because we only included RCTs and did not downgrade it in the key 5 criteria—risk of bias, inconsistency, indirectness, imprecision, and publication bias. First, risk of bias was not considered high regarding randomization, blinding, attrition, and selective reporting. Second, outcomes were considered consistent because the study conclusions were consistent, and pooled effect sizes have narrow CIs. Although heterogeneity was considerable, the likely sources of heterogeneity were comprehensively suggested. Third, indirectness was low, as all the outcomes can directly answer our research question. Fourth, imprecision was considered low because the effect estimate was calculated from a large number of total participants (n=12,070) and was precise according to the GRADE guidelines. Finally, the possibility of publication bias was low to moderate.

Outcomes of the Included Studies

We assessed the effects of DMH intervention compared with control groups with a mean (SD) from 80 studies (67 records for depression and 65 records for anxiety; Figure 2 [26, 28, 29, 32, 33, 35-39, 44-47, 54, 55, 57, 59-61, 63, 65, 68-78, 80, 82-84, 86-88, 90, 91, 93, 95, 97-101, 106, 107, 109, 111, 113, 115, 117, 119-121] and Figure 3 [27, 29, 30, 32, 35-38, 40, 42, 43, 48-50, 52, 55-59, 61, 62, 68, 69, 72-74, 78, 80, 82-84, 86, 88, 89, 93, 99, 101-103, 105-107, 109, 110, 112-118, 120, 121]). The pooled effect sizes were −0.61 (95% CI −0.78 to −0.44) for depression and −0.73 (95% CI −0.93 to −0.53) for anxiety, indicating that groups that used DMH tools had lower postintervention depression and anxiety symptoms compared with the controls. Considerable heterogeneity was observed (I^2=0.94 for depression and 0.95 for anxiety). However, we were able to find the likely sources of heterogeneity, because the effect sizes were considerably different between prespecified subgroups (eg, by intervention content, technology type, mode, depression or anxiety level, outcome, and age) in both depression and anxiety (Table 1). For example, by intervention content, the largest effects on depression and anxiety were found among studies that provided CBT and other types of psychotherapy. By mode of delivery, internet- or mobile app–based interventions showed larger effect sizes compared with telephone- or text message–based approaches. By age, effect sizes on depression were larger among adults than among children, whereas effects on anxiety were larger among children than among adults. Although effect sizes were notably different among subgroups of studies, DMH interventions showed remarkably consistent pooled effects in reducing depression and anxiety symptoms among all subgroups of studies examined.

Notably, the effect size was larger (−0.70, 95% CI −0.90 to −0.51 for depression and −0.85, 95% CI −1.08 to −0.62 for anxiety) when we removed studies having considerable baseline score differences between groups in the sensitivity analysis. However, excluding outliers resulted in decreased effect sizes both in depression (−0.57, 95% CI −0.66 to −0.44) and anxiety (−0.66, 95% CI −0.77 to −0.56). Other sensitivity analyses did not change the effect sizes (Multimedia Appendix 7). Funnel plots appeared to be symmetrical for both anxiety and depression (Multimedia Appendix 8). Egger test results were not significant for depression (P=.19) but significant for anxiety (P=.03), suggesting potential publication bias for anxiety. However, when we ran the Duval and Tweedie trim and fill analysis to statistically assess the publication bias, no imputation was necessary to adjust for the publication bias in both depression and anxiety, and the effect sizes stayed the same.

Figure 2. Effect of digital mental health interventions on depression. *Studies with more than 2 arms. The frequency (N) of the control group was divided by the number of comparisons to avoid being overweight. REML: restricted maximum likelihood.

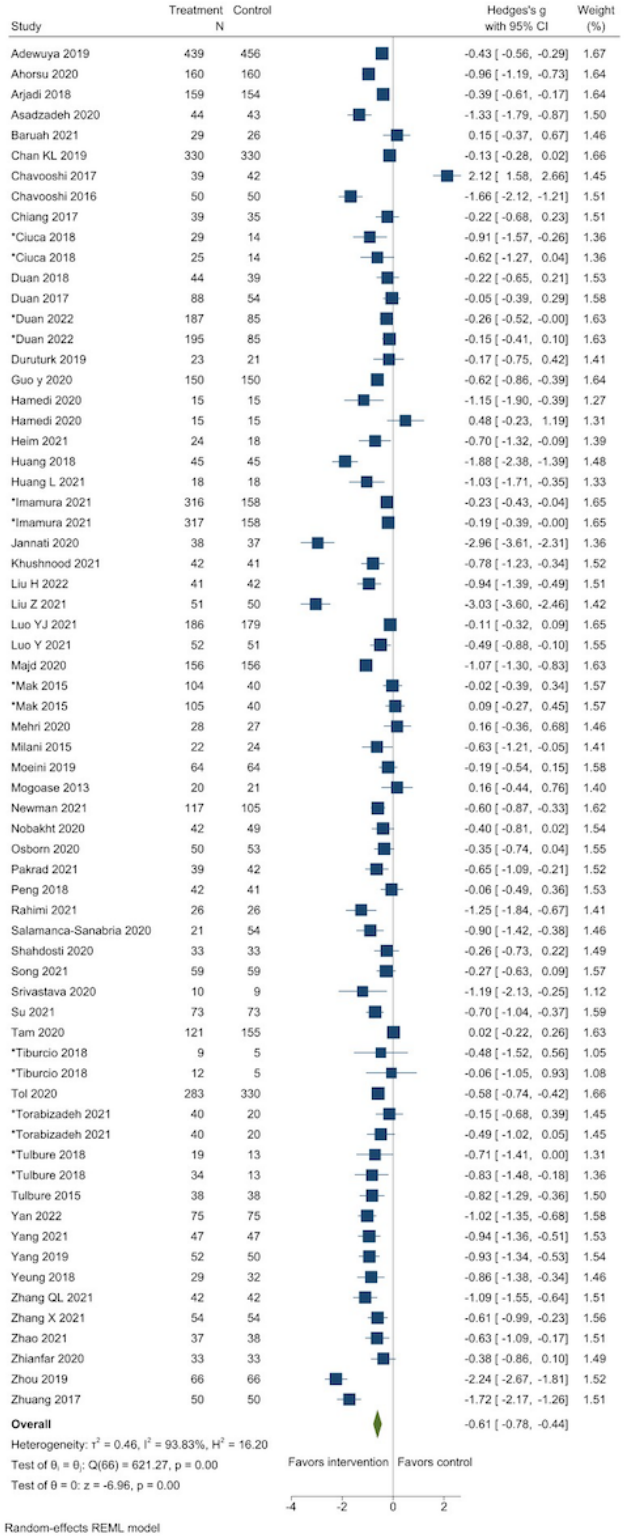


Figure 3. Effect of digital mental health interventions on anxiety. *Studies with more than 2 arms. The frequency (N) of the control group was divided by the number of comparisons to avoid being overweight. REML: restricted maximum likelihood.

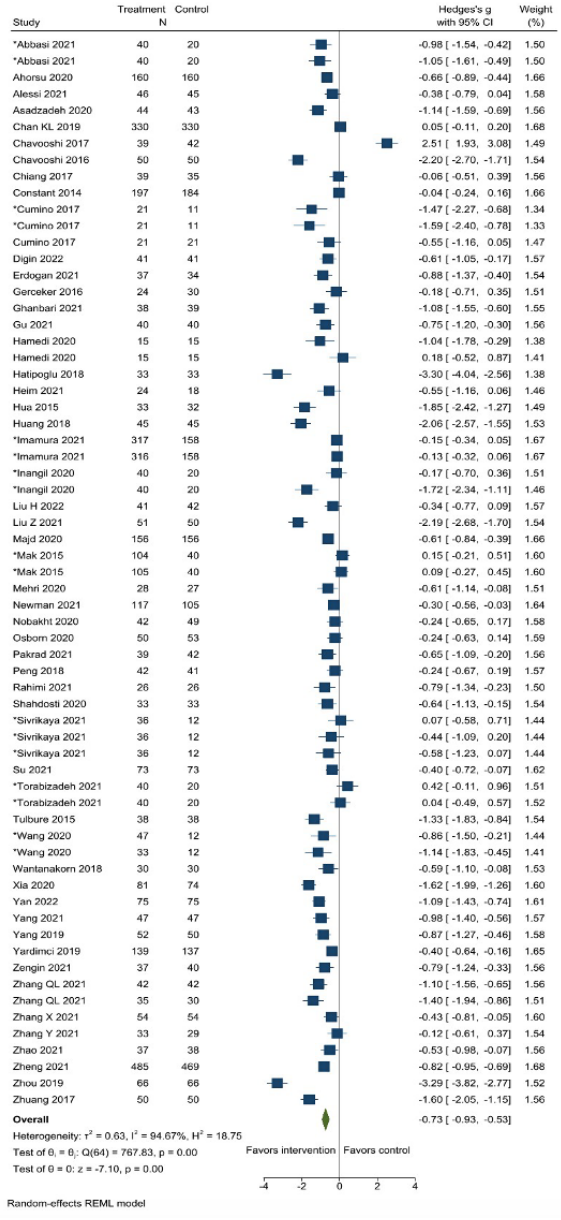


Table 1. Subgroup analyses of digital mental health interventions.

Subgroup[a]	Depression				Anxiety			
	Frequency, n	Hedges *g* (95% CI)	I^2 (%)	*P* value	Frequency, n	Hedges *g* (95% CI)	I^2 (%)	*P* value
By content				<.001				<.001
CBT[b]	18	−0.93 (−1.33 to −0.53)	94		10	−0.78 (−1.20 to −0.35)	94	
Other psychotherapy	7	−0.58 (−1.64 to 0.48)	96		7	−0.82 (−2.15 to 0.52)	98	
Psychoeducation	7	−0.30 (−0.45 to −0.16)	42		11	−0.27 (−0.61 to 0.07)	89	
Multiple content	21	−0.62 (−0.85 to −0.39)	90		19	−0.75 (−0.98 to −0.53)	87	
Physical activity	7	−0.29 (−0.47 to −0.11)	40		1	−0.53 (−0.98 to −0.07)	N/A[c]	
Mindfulness	4	−0.36 (−0.84 to 0.11)	84		4	−0.26 (−0.73 to 0.21)	84	
Social support	3	−0.59 (−1.02 to −0.15)	42		1	−0.64 (−1.13 to −0.16)	N/A	
Playful distraction	N/A	N/A	N/A		12	−1.22 (−1.67 to −0.76)	86	
By type of technology use				<.001				.01
Tech intervention	56	−0.58 (−0.73 to −0.44)	90		54	−0.74 (−0.94 to −0.55)	93	
As a delivery tool	11	−0.74 (−1.52 to 0.05)	96		11	−0.65 (−1.41 to 0.11)	96	
By mode				.03				<.001
Internet-based	37	−0.61 (−0.84 to −0.38)	92		19	−0.65 (−0.96 to −0.35)	91	
Mobile apps	21	−0.75 (−0.50 to −1.00)	92		36	−0.92 (−1.18 to −0.66)	95	
Telephone or Texting	8	−0.24 (−1.00 to −0.50)	96		9	−0.10 (−0.79 to 0.59)	96	
Audio or video files	1	−0.58 (−0.74 to −0.42)	N/A		1	−1.05 (−1.61 to −0.49)	N/A	
By duration				.49				.005
Multiple session	66	−0.62 (−0.79 to −0.44)	94		50	−0.71 (−0.96 to −0.46)	96	
Single session	1	−0.35 (−0.74 to 0.04)	N/A		15	−0.80 (−1.06 to −0.53)	71	
By depression or anxiety severity				<.001				<.001
Disorder	19	−0.91 (−1.25 to −0.56)	95		10	−0.96 (−1.30 to −0.62)	82	
Symptomatic	48	−0.50 (−0.69 to −0.31)	93		55	−0.70 (−0.93 to −0.47)	95	
By outcome				<.001				.007
Primary	60	−0.57 (−0.76 to −0.39)	94		58	−0.73 (−0.95 to −0.51)	94	
Secondary	7	−0.94 (−1.24 to −0.65)	85		7	−0.73 (−1.27 to −0.19)	96	
By control				<.001				.91
No intervention	22	−0.48 (−0.74 to −0.22)	89		22	−0.65 (−0.85 to −0.44)	73	
Usual care or active	45	−0.67 (−0.90 to −0.45)	95		43	−0.77 (−1.06 to −0.48)	97	
By age				<.001				<.001
Children	6	−0.32 (−0.76 to 0.12)	82		15	−1.05 (−1.46 to −0.63)	91	
Adult	60	−0.63 (−0.82 to −0.45)	94		50	−0.64 (−0.87 to −0.41)	95	
Older adult	1	−0.78 (−1.23 to −0.34)	N/A		N/A	N/A	N/A	
By region				.67				<.001
Asia	53	−0.63 (−0.85 to −0.42)	95		46	−0.72 (−0.97 to −0.47)	96	
Africa	3	−0.48 (−0.61 to −0.35)	29		2	−0.08 (−0.26 to 0.10)	0	
Europe	8	−0.57 (−0.84 to −0.38)	35		13	−0.82 (−1.27 to −0.37)	91	
Latin America	3	−0.62 (−1.14 to −0.10)	23		4	−0.93 (−1.54 to −0.32)	75	

[a]Intervention content (CBT, psychotherapy, psychoeducation, multiple contents, physical activity, mindfulness, social support, and playful distraction). Although CBT and mindfulness-based cognitive therapy are types of psychotherapy, we separated these from other types of psychotherapy. As CBT was the single most frequent, and mindfulness was emerging content for digital mental health tools, we intended to evaluate the effects separately: type

of technology use (technology itself as an intervention or technology as a delivery tool); mode (internet-based, mobile apps, telephone or text messages, or audio or video files); intervention duration (multiple session or single session); depression or anxiety severity (depression or anxiety disorder or depression or anxiety at risk or simply symptomatic); outcome measurement (depression or anxiety is primary outcome or depression or anxiety is secondary outcome); control type (usual care or active control, no intervention, or waitlist control); participant age (children, adults, or older adults); and study region (Asia, Africa, Europe, or Latin America).

[b]CBT: cognitive behavioral therapy.

[c]N/A: not applicable.

Discussion

Principal Findings and Comparisons With Previous Work

With our 96 studies for systematic review and 80 studies for meta-analysis, we found that DMH interventions showed moderate to high effectiveness in reducing depression and anxiety symptoms in LMICs. Our findings contribute to knowledge-building in the effectiveness of DMH tools in LMICs, especially on depression and anxiety, which are the 2 most common mental disorders. The aggregated results from 80 RCTs with moderate to high effect sizes (Hedges g=−0.61 for depression and −0.73 for anxiety) provide a comprehensive up-to-date review (up to February 2022), with some promising evidence. Moreover, our findings are well aligned with previous systematic reviews in that DMH tools improved mental health outcomes in low-resource settings [20] with a moderate overall effect size (Hedges g=0.60) [22]. Our results are also comparable with the outcomes of digital psychological interventions in HICs that were found to be moderately effective in reducing depression (Hedges g=0.51-0.58) [124,125] and highly effective for anxiety (Hedges g=0.80) [126].

This study has some clinical implications for patient care. First, when psychotherapy, including CBT, was delivered via digital formats, it effectively relieved depression and anxiety symptoms compared with the usual care. Moreover, perinatal and postpartum depression was the single most frequent target for DMH intervention, followed by depression or anxiety care for the caregivers of children with chronic or congenital health conditions. Patients with chronic diseases were also often targeted. Thus, mobile app– or internet-based psychotherapy could be one way to care for the mental health of perinatal or postpartum women, patients with chronic disease, or their caregivers, who may be marginalized in mental health care in low-resource settings. Finally, our findings showed that digital tools (eg, social media and audiovisual materials) effectively alleviated preprocedure-related anxiety in the clinical setting.

Limitations

Substantial heterogeneity should be acknowledged as a limitation of this study. It was anticipated because the main components of the study varied, including intervention formats, control types, and participants' ages, which could lead to the large variance in effect sizes between studies. As I^2 is an indicator of inconsistency across the study outcomes, high I^2 signals that the observed variability could be real. We are also aware of the possibility of bias, including the small-study effect, because half of the studies had small sample sizes (eg, the total participants were <100). Small and pilot studies tend to include participants whose symptoms are likely more prominent or who are likely proactive and compliant with the study instructions. This could allow the studies to be well controlled and managed, potentially leading to a large overall effect size compared with what may occur in a real-world setting. Considering that DMH interventions tested through RCTs are still at an early stage in LMICs, it is not surprising that the sample size tends to be small, and the contents and study protocols vary substantially between studies. Moreover, the Egger test detected the possibility of publication bias (P=.03) for anxiety, although the trim and fill test revealed that no imputation was necessary for both outcomes to account for the publication bias. There is a possibility of duplication bias as we observed some studies conducted by the same research team, using similar study designs and interventions but for different target populations. We also need to recognize citation bias, as the notable findings are likely to be cited and included in our study. However, we believe that the possibility of language bias and availability or cost bias is minimal; we included non-English studies published in their local journals, and no papers were excluded owing to journal unavailability for full-text review. Finally, we were unable to adjust for the baseline scores, which could lead to the biased effect size or conservative results. For example, we witnessed some preintervention score differences between groups exceeding the postintervention score differences, meaning that the postintervention score differences could not correctly reflect the genuine effect of the intervention. In this study, this made the pooled effect size conservative for both depression and anxiety, resulting in increased effect sizes when these studies were excluded for sensitivity analysis.

Study Strengths and Future Directions

This study has several strengths that can be acknowledged. First, our findings are aggregated results of 96 RCTs conducted in LMICs from 2011 to 2021. Hence, the moderate and high pooled effect sizes for depression and anxiety are supported by a large number of participants included. Second, our study focused on the most common mental disorders—depression and anxiety. The participants included those with both light depression or anxiety symptoms and moderate depression or anxiety disorders. The participants' age range was broad, from preschool children to older adults, and study settings varied, including hospitals, schools, or fully web based. Thus, the diversity of study populations and treatment delivery settings may contribute to the generalizability of our results and qualitative and quantitative comparisons between various subgroups. Third, we were able to detect potential sources of variability, which could possibly be the intervention content, control type, baseline depression or anxiety symptom severity, and participant age. Notable effect size differences by subgroup signaled that the effectiveness of DMH tools could be more effective in reducing depressive or anxiety symptoms when a certain content was applied (eg, CBT) or it was administered for a certain level of baseline depression or anxiety (eg, depression disorder). Notably, our findings

showed that using mobile apps could be more effective than non–mobile app–based DMH tools in relieving depressive and anxiety symptoms, although all these interpretations need caution. The evidence of the source of variability for both depression and anxiety is novel, as this information was not available in previous reviews. Fourth, noticeably, the effect size was higher when the DMH tools were compared with usual care or active control than no intervention for depression. This could highlight that DMH could be as effective as usual in-person care, not just better than nothing. Finally, studies conducted during the COVID-19 pandemic were included, showing that DMH tools could be useful to deal with depression and anxiety even when in-person social connections were reduced during an unprecedented worldwide infectious disease outbreak. Our findings suggest that DMH tools could be viable options when infectious diseases increase the risk of in-person–based care. This could be encouraging, as we are still in the extended COVID-19 pandemic era, and reportedly, the overall mental distress had been elevated nearly universally among all populations, especially in LMICs. For future studies, we would suggest that more effort would be needed for economic evaluation of DMH tools to gauge the feasibility of implementation and economic consequences and implications in LMICs.

Conclusions

Our findings from a systematic review and meta-analysis of 96 RCTs showed that DMH tools could be an effective method to care for depression and anxiety in low-resource settings where usual care is minimally available or feasible despite the care necessity and urgency. This study provides ample evidence-based insight regarding future directions of DMH use for depression and anxiety in LMICs, given the anticipated increasing demand, development, and implementation of DMH tools.

Data Availability

All data extracted from the reported papers for this systematic review and meta-analysis will be made available with publication on reasonable request to the corresponding author.

Authors' Contributions

JK conceptualized the study, and JK and ELP designed the methods and protocol. JK, SP, and HB contributed to the literature review and study selection for inclusion. JK, SP, LMDA, and HB contributed to data extraction (JK did data double extraction), data check, and study quality assessment. JK and HB were responsible for statistical analysis and data interpretation. JK prepared the first draft of the manuscript, and HB, ELP, JSH, and YKC substantially contributed to manuscript revision. All authors had full access to all the data, read the final manuscript, and approved its submission for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 101 KB - mental_v10i1e43066_app1.pdf]

Multimedia Appendix 2
Study quality assessment (Effective Public Health Practice Project).
[PDF File (Adobe PDF File), 230 KB - mental_v10i1e43066_app2.pdf]

Multimedia Appendix 3
Overall quality of evidence (Grading of Recommendations, Assessment, and Evaluation).
[PDF File (Adobe PDF File), 103 KB - mental_v10i1e43066_app3.pdf]

Multimedia Appendix 4
PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) 2020 Checklist.
[PDF File (Adobe PDF File), 145 KB - mental_v10i1e43066_app4.pdf]

Multimedia Appendix 5
Characteristics of studies included in the systematic review.
[PDF File (Adobe PDF File), 301 KB - mental_v10i1e43066_app5.pdf]

Multimedia Appendix 6
Supplemental information of the studies for the systematic review.

[PDF File (Adobe PDF File), 250 KB - mental_v10i1e43066_app6.pdf]

Multimedia Appendix 7
Supplemental information on the sensitivity analysis.
[PDF File (Adobe PDF File), 71 KB - mental_v10i1e43066_app7.pdf]

Multimedia Appendix 8
Funnel plots for depression and anxiety.
[PDF File (Adobe PDF File), 237 KB - mental_v10i1e43066_app8.pdf]

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150 [FREE Full text] [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]
2. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/W?sequence=1 [accessed 2023-01-19]
3. Mental health atlas 2014. World Health Organization. 2015. URL: https://apps.who.int/iris/handle/10665/178879 [accessed 2022-07-08]
4. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016 May;3(5):415-424 [FREE Full text] [doi: 10.1016/S2215-0366(16)30024-4] [Medline: 27083119]
5. Pathare S, Brazinova A, Levav I. Care gap: a comprehensive measure to quantify unmet needs in mental health. Epidemiol Psychiatr Sci 2018 Oct;27(5):463-467 [FREE Full text] [doi: 10.1017/S2045796018000100] [Medline: 29521609]
6. Fit Mind, Fit Job: From Evidence to Practice in Mental Health and Work. Organisation for Economic Co-operation and Development. Paris, France: OECD Publishing; 2016. URL: https://read.oecd-ilibrary.org/employment/fit-mind-fit-job_9789264228283-en#page1 [accessed 2022-07-08]
7. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
8. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
9. Cross SP, Karin E, Staples LG, Bisby MA, Ryan K, Duke G, et al. Factors associated with treatment uptake, completion, and subsequent symptom improvement in a national digital mental health service. Internet Interv 2022 Mar;27:100506 [FREE Full text] [doi: 10.1016/j.invent.2022.100506] [Medline: 35242587]
10. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022 Jan 31;9(1):e30766 [FREE Full text] [doi: 10.2196/30766] [Medline: 35099398]
11. Povey J, Sweet M, Nagel T, Lowell A, Shand F, Vigona J, et al. Determining priorities in the aboriginal and islander mental health initiative for youth app second phase participatory design project: qualitative study and narrative literature review. JMIR Form Res 2022 Feb 18;6(2):e28342 [FREE Full text] [doi: 10.2196/28342] [Medline: 35179498]
12. Gerhards SA, de Graaf LE, Jacobs LE, Severens JL, Huibers MJ, Arntz A, et al. Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2010 Apr;196(4):310-318. [doi: 10.1192/bjp.bp.109.065748] [Medline: 20357309]
13. Hedman E, Andersson E, Ljótsson B, Andersson G, Rück C, Lindefors N. Cost-effectiveness of internet-based cognitive behavior therapy vs. cognitive behavioral group therapy for social anxiety disorder: results from a randomized controlled trial. Behav Res Ther 2011 Nov;49(11):729-736. [doi: 10.1016/j.brat.2011.07.009] [Medline: 21851929]
14. Carswell K, Harper-Shehadeh M, Watts S, Van't Hof E, Abi Ramia J, Heim E, et al. Step-by-step: a new WHO digital mental health intervention for depression. Mhealth 2018 Aug 13;4:34 [FREE Full text] [doi: 10.21037/mhealth.2018.08.01] [Medline: 30225240]
15. Bergström A, Fottrell E, Hopkins H, Lloyd D, Stevenson O, Willats P, et al. mHealth: can mobile technology improve health in low-and middle-income countries? University College London. 2015 Jul. URL: https://www.ucl.ac.uk/public-policy/sites/public-policy/files/migrated-files/mhealth_briefing.pdf [accessed 2022-07-15]
16. Mobile for Development: The State of Mobile Internet Connectivity Report 2019. GSM Association. 2019 Jul 16. URL: https://www.gsma.com/mobilefordevelopment/resources/the-state-of-mobile-internet-connectivity-report-2019/#targetText=The [accessed 2022-07-15]

17. Burki T. Developing countries in the digital revolution. Lancet 2018 Feb 03;391(10119):417. [doi: 10.1016/S0140-6736(18)30191-0] [Medline: 29407022]
18. Stephani V, Opoku D, Quentin W. A systematic review of randomized controlled trials of mHealth interventions against non-communicable diseases in developing countries. BMC Public Health 2016 Jul 15;16:572 [FREE Full text] [doi: 10.1186/s12889-016-3226-3] [Medline: 27417513]
19. Kaonga NN, Morgan J. Common themes and emerging trends for the use of technology to support mental health and psychosocial well-being in limited resource settings: a review of the literature. Psychiatry Res 2019 Nov;281:112594. [doi: 10.1016/j.psychres.2019.112594] [Medline: 31605874]
20. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017 Jun;4(6):486-500 [FREE Full text] [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]
21. Jiménez-Molina Á, Franco P, Martínez V, Martínez P, Rojas G, Araya R. Internet-based interventions for the prevention and treatment of mental disorders in Latin America: a scoping review. Front Psychiatry 2019 Sep 13;10:664 [FREE Full text] [doi: 10.3389/fpsyt.2019.00664] [Medline: 31572242]
22. Fu Z, Burger H, Arjadi R, Bockting CL. Effectiveness of digital psychological interventions for mental health problems in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Psychiatry 2020 Oct;7(10):851-864 [FREE Full text] [doi: 10.1016/S2215-0366(20)30256-X] [Medline: 32866459]
23. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
24. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]
25. Dragioti E, Li H, Tsitsas G, Lee KH, Choi J, Kim J, et al. A large-scale meta-analytic atlas of mental health problems prevalence during the COVID-19 early pandemic. J Med Virol 2022 May;94(5):1935-1949 [FREE Full text] [doi: 10.1002/jmv.27549] [Medline: 34958144]
26. Arjadi R, Nauta MH, Scholte WF, Hollon SD, Chowdhary N, Suryani AO, et al. Internet-based behavioural activation with lay counsellor support versus online minimal psychoeducation without support for treatment of depression: a randomised controlled trial in Indonesia. Lancet Psychiatry 2018 Sep;5(9):707-716. [doi: 10.1016/S2215-0366(18)30223-2] [Medline: 30006262]
27. Abbasi H, Saqib M, Jouhar R, Lal A, Ahmed N, Ahmed MA, et al. The efficacy of little lovely dentist, dental song, and tell-show-do techniques in alleviating dental anxiety in paediatric patients: a clinical trial. Biomed Res Int 2021 May 23;2021:1119710 [FREE Full text] [doi: 10.1155/2021/1119710] [Medline: 34124238]
28. Adewuya AO, Momodu O, Olibamoyo O, Adegbaju A, Adesoji O, Adegbokun A. The effectiveness and acceptability of mobile telephone adherence support for management of depression in the Mental Health in Primary Care (MeHPriC) project, Lagos, Nigeria: a pilot cluster randomised controlled trial. J Affect Disord 2019 Jun 15;253:118-125. [doi: 10.1016/j.jad.2019.04.025] [Medline: 31035212]
29. Ahorsu DK, Lin CY, Imani V, Carlbring P, Nygårdh A, Broström A, et al. Testing an app-based intervention to improve insomnia in patients with epilepsy: a randomized controlled trial. Epilepsy Behav 2020 Nov;112:107371. [doi: 10.1016/j.yebeh.2020.107371] [Medline: 32861897]
30. Alessi J, de Oliveira GB, Franco DW, Becker AS, Knijnik CP, Kobe GL, et al. Telehealth strategy to mitigate the negative psychological impact of the COVID-19 pandemic on type 2 diabetes: a randomized controlled trial. Acta Diabetol 2021 Jul;58(7):899-909 [FREE Full text] [doi: 10.1007/s00592-021-01690-1] [Medline: 33723649]
31. Araya R, Menezes PR, Claro HG, Brandt LR, Daley KL, Quayle J, et al. Effect of a digital intervention on depressive symptoms in patients with comorbid hypertension or diabetes in Brazil and Peru: two randomized clinical trials. JAMA 2021 May 11;325(18):1852-1862 [FREE Full text] [doi: 10.1001/jama.2021.4348] [Medline: 33974019]
32. Asadzadeh L, Jafari E, Kharaghani R, Taremian F. Effectiveness of midwife-led brief counseling intervention on post-traumatic stress disorder, depression, and anxiety symptoms of women experiencing a traumatic childbirth: a randomized controlled trial. BMC Pregnancy Childbirth 2020 Mar 06;20(1):142 [FREE Full text] [doi: 10.1186/s12884-020-2826-1] [Medline: 32138707]
33. Baruah U, Varghese M, Loganathan S, Mehta KM, Gallagher-Thompson D, Zandi D, et al. Feasibility and preliminary effectiveness of an online training and support program for caregivers of people with dementia in India: a randomized controlled trial. Int J Geriatr Psychiatry 2021 Apr;36(4):606-617. [doi: 10.1002/gps.5502] [Medline: 33491811]
34. Byonanebye DM, Nabaggala MS, Naggirinya AB, Lamorde M, Oseku E, King R, et al. An interactive voice response software to improve the quality of life of people living with HIV in Uganda: randomized controlled trial. JMIR Mhealth Uhealth 2021 Feb 11;9(2):e22229 [FREE Full text] [doi: 10.2196/22229] [Medline: 33570497]
35. Chan KL, Leung WC, Tiwari A, Or KL, Ip P. Using smartphone-based psychoeducation to reduce postnatal depression among first-time mothers: randomized controlled trial. JMIR Mhealth Uhealth 2019 May 14;7(5):e12794 [FREE Full text] [doi: 10.2196/12794] [Medline: 31094354]

36. Chavooshi B, Mohammadkhani P, Dolatshahee B. Telemedicine vs. in-person delivery of intensive short-term dynamic psychotherapy for patients with medically unexplained pain: a 12-month randomized, controlled trial. J Telemed Telecare 2017 Jan;23(1):133-141. [doi: 10.1177/1357633X15627382] [Medline: 27036877]
37. Chavooshi B, Mohammadkhani P, Dolatshahi B. A randomized double-blind controlled trial comparing Davanloo intensive short-term dynamic psychotherapy as internet-delivered vs treatment as usual for medically unexplained pain: a 6-month pilot study. Psychosomatics 2016;57(3):292-300. [doi: 10.1016/j.psym.2016.01.001] [Medline: 27017450]
38. Chiang VC, Lee RL, Ho MF, Leung CK, Tang PY, Wong SW, et al. Fulfilling the psychological and information need of the family members of critically ill patients using interactive mobile technology: a randomised controlled trial. Intensive Crit Care Nurs 2017 Aug;41:77-83. [doi: 10.1016/j.iccn.2017.03.006] [Medline: 28438476]
39. Ciuca AM, Berger T, Crişan LG, Miclea M. Internet-based treatment for panic disorder: a three-arm randomized controlled trial comparing guided (via real-time video sessions) with unguided self-help treatment and a waitlist control. PAXPD study results. J Anxiety Disord 2018 May;56:43-55. [doi: 10.1016/j.janxdis.2018.03.009] [Medline: 29625834]
40. Constant D, de Tolly K, Harries J, Myer L. Mobile phone messages to provide support to women during the home phase of medical abortion in South Africa: a randomised controlled trial. Contraception 2014 Sep;90(3):226-233. [doi: 10.1016/j.contraception.2014.04.009] [Medline: 24850188]
41. Craveiro MA, Caldeira CL. Influence of an audiovisual resource on the preoperative anxiety of adult endodontic patients: a randomized controlled clinical trial. J Endod 2020 Jul;46(7):909-914. [doi: 10.1016/j.joen.2020.03.024] [Medline: 32389383]
42. Cumino DO, Vieira JE, Lima LC, Stievano LP, Silva RA, Mathias LA. Smartphone-based behavioural intervention alleviates children's anxiety during anaesthesia induction: a randomised controlled trial. Eur J Anaesthesiol 2017 Mar;34(3):169-175. [doi: 10.1097/EJA.0000000000000589] [Medline: 28146459]
43. Dığın F, Özkan ZK, Şahin A. Effect of sending SMS, which reminds about the intake of medication, on reducing postoperative anxiety in patients undergoing cataract surgery: a randomized controlled study. J Perianesth Nurs 2022 Feb;37(1):75-79. [doi: 10.1016/j.jopan.2021.07.002] [Medline: 34815165]
44. Duan YP, Liang W, Guo L, Wienert J, Si GY, Lippke S. Evaluation of a web-based intervention for multiple health behavior changes in patients with coronary heart disease in home-based rehabilitation: pilot randomized controlled trial. J Med Internet Res 2018 Nov 19;20(11):e12052 [FREE Full text] [doi: 10.2196/12052] [Medline: 30455167]
45. Duan YP, Wienert J, Hu C, Si GY, Lippke S. Web-based intervention for physical activity and fruit and vegetable intake among Chinese university students: a randomized controlled trial. J Med Internet Res 2017 Apr 10;19(4):e106 [FREE Full text] [doi: 10.2196/jmir.7152] [Medline: 28396306]
46. Duan Y, Liang W, Wang Y, Lippke S, Lin Z, Shang B, et al. The effectiveness of sequentially delivered web-based interventions on promoting physical activity and fruit-vegetable consumption among Chinese college students: mixed methods study. J Med Internet Res 2022 Jan 26;24(1):e30566 [FREE Full text] [doi: 10.2196/30566] [Medline: 35080497]
47. Duruturk N, Özköslü MA. Effect of tele-rehabilitation on glucose control, exercise capacity, physical fitness, muscle strength and psychosocial status in patients with type 2 diabetes: a double blind randomized controlled trial. Prim Care Diabetes 2019 Dec;13(6):542-548. [doi: 10.1016/j.pcd.2019.03.007] [Medline: 31014938]
48. Erdogan B, Aytekin Ozdemir A. The effect of three different methods on venipuncture pain and anxiety in children: distraction cards, virtual reality, and Buzzy® (randomized controlled trial). J Pediatr Nurs 2021;58:e54-e62. [doi: 10.1016/j.pedn.2021.01.001] [Medline: 33485746]
49. Özalp Gerçeker G, Karayağız Muslu G, Yardimci F. Children's postoperative symptoms at home through nurse-led telephone counseling and its effects on parents' anxiety: a randomized controlled trial. J Spec Pediatr Nurs 2016 Oct;21(4):189-199. [doi: 10.1111/jspn.12155] [Medline: 27596004]
50. Ghanbari E, Yektatalab S, Mehrabi M. Effects of psychoeducational interventions using mobile apps and mobile-based online group discussions on anxiety and self-esteem in women with breast cancer: randomized controlled trial. JMIR Mhealth Uhealth 2021 May 18;9(5):e19262 [FREE Full text] [doi: 10.2196/19262] [Medline: 34003138]
51. Ghawadra SF, Lim Abdullah K, Choo WY, Danaee M, Phang CK. The effect of mindfulness-based training on stress, anxiety, depression and job satisfaction among ward nurses: a randomized control trial. J Nurs Manag 2020 Jul;28(5):1088-1097. [doi: 10.1111/jonm.13049] [Medline: 32432795]
52. Gu S, Ping J, Xu M, Zhou Y. TikTok browsing for anxiety relief in the preoperative period: a randomized clinical trial. Complement Ther Med 2021 Aug;60:102749 [FREE Full text] [doi: 10.1016/j.ctim.2021.102749] [Medline: 34118388]
53. Guo L, Zhang J, Mu L, Ye Z. Preventing postpartum depression with mindful self-compassion intervention: a randomized control study. J Nerv Ment Dis 2020 Feb;208(2):101-107. [doi: 10.1097/NMD.0000000000001096] [Medline: 31868776]
54. Guo Y, Hong YA, Cai W, Li L, Hao Y, Qiao J, et al. Effect of a WeChat-based intervention (Run4Love) on depressive symptoms among people living with HIV in China: a randomized controlled trial. J Med Internet Res 2020 Feb 09;22(2):e16715 [FREE Full text] [doi: 10.2196/16715] [Medline: 32044751]
55. Hamedi V, Hamid N, Beshlideh K, Marashi SA, Hashemi Sheikh Shabani SE. Effectiveness of conventional cognitive-behavioral therapy and its computerized version on reduction in pain intensity, depression, anger, and anxiety in children with cancer: a randomized, controlled trial. Iran J Psychiatry Behav Sci 2020 Dec 5;14(4):e83110. [doi: 10.5812/ijpbs.83110]

56. Hatipoglu Z, Gulec E, Lafli D, Ozcengiz D. Effects of auditory and audiovisual presentations on anxiety and behavioral changes in children undergoing elective surgery. Niger J Clin Pract 2018 Jun;21(6):788-794 [FREE Full text] [doi: 10.4103/njcp.njcp_227_17] [Medline: 29888729]
57. Heim E, Ramia JA, Hana RA, Burchert S, Carswell K, Cornelisz I, et al. Step-by-step: feasibility randomised controlled trial of a mobile-based intervention for depression among populations affected by adversity in Lebanon. Internet Interv 2021 Apr;24:100380 [FREE Full text] [doi: 10.1016/j.invent.2021.100380] [Medline: 33747798]
58. Hua Y, Qiu R, Yao W, Zhang Q, Chen X. The effect of virtual reality distraction on pain relief during dressing changes in children with chronic wounds on lower limbs. Pain Manage Nursing 2015 Oct 18;16(5):685-691 [FREE Full text] [doi: 10.1016/J.PMN.2015.03.001]
59. Huang LX. Effect of WeChat platform-based nursing care on negative emotion and quality of life in women with chronic hepatitis B during late pregnancy. World Chinese J Dig 2018;26:126-130. [doi: 10.11569/WCJD.V26.I2.126]
60. Huang L, Shen Q, Fang Q, Zheng X. Effects of internet-based support program on parenting outcomes for primiparous women: a pilot study. Int J Environ Res Public Health 2021 Apr 21;18(9):4402 [FREE Full text] [doi: 10.3390/ijerph18094402] [Medline: 33919112]
61. Imamura K, Tran TT, Nguyen HT, Sasaki N, Kuribayashi K, Sakuraya A, et al. Effect of smartphone-based stress management programs on depression and anxiety of hospital nurses in Vietnam: a three-arm randomized controlled trial. Sci Rep 2021 May 31;11(1):11353 [FREE Full text] [doi: 10.1038/s41598-021-90320-5] [Medline: 34059737]
62. İnangil D, Şendir M, Büyükyılmaz F. Efficacy of cartoon viewing devices during phlebotomy in children: a randomized controlled trial. J Perianesth Nurs 2020 Aug;35(4):407-412. [doi: 10.1016/j.jopan.2020.01.008] [Medline: 32360128]
63. Jannati N, Mazhari S, Ahmadian L, Mirzaee M. Effectiveness of an app-based cognitive behavioral therapy program for postpartum depression in primary care: a randomized controlled trial. Int J Med Inform 2020 Sep;141:104145. [doi: 10.1016/j.ijmedinf.2020.104145] [Medline: 32480319]
64. Jareethum R, Titapant V, Chantra T, Sommai V, Chuenwattana P, Jirawan C. Satisfaction of healthy pregnant women receiving short message service via mobile phone for prenatal support: a randomized controlled trial. J Med Assoc Thai 2008 Apr;91(4):458-463. [Medline: 18556852]
65. Khushnood K, Altaf S, Sultan N, Ali Awan MM, Mehmood R, Qureshi S. Role Wii Fit exer-games in improving balance confidence and quality of life in elderly population. J Pak Med Assoc 2021 Sep;71(9):2130-2134 [FREE Full text] [doi: 10.47391/JPMA.319] [Medline: 34580500]
66. Korkmaz S, Iyigun E, Tastan S. An evaluation of the influence of web-based patient education on the anxiety and life quality of patients who have undergone mammaplasty: a randomized controlled study. J Cancer Educ 2020 Oct;35(5):912-922. [doi: 10.1007/s13187-019-01542-1] [Medline: 31119709]
67. Li J, Mo PK, Kahler CW, Lau JT. A three-arm randomised controlled trial to evaluate the efficacy of a positive psychology and social networking intervention in promoting mental health among HIV-infected men who have sex with men in China. Epidemiol Psychiatr Sci 2021 Mar 19;30:e24 [FREE Full text] [doi: 10.1017/S2045796021000081] [Medline: 33736740]
68. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. Internet Interv 2022 Mar;27:100495 [FREE Full text] [doi: 10.1016/j.invent.2022.100495] [Medline: 35059305]
69. Liu Z, Qiao D, Xu Y, Zhao W, Yang Y, Wen D, et al. The efficacy of computerized cognitive behavioral therapy for depressive and anxiety symptoms in patients with COVID-19: randomized controlled trial. J Med Internet Res 2021 May 14;23(5):e26883 [FREE Full text] [doi: 10.2196/26883] [Medline: 33900931]
70. Luo Y, Xia W, Cheung AT, Ho LL, Zhang J, Xie J, et al. Effectiveness of a mobile device-based resilience training program in reducing depressive symptoms and enhancing resilience and quality of life in parents of children with cancer: randomized controlled trial. J Med Internet Res 2021 Nov 29;23(11):e27639 [FREE Full text] [doi: 10.2196/27639] [Medline: 34847060]
71. Luo YJ, Jackson T, Stice E, Chen H. Effectiveness of an internet dissonance-based eating disorder prevention intervention among body-dissatisfied young Chinese women. Behav Ther 2021 Jan;52(1):221-233. [doi: 10.1016/j.beth.2020.04.007] [Medline: 33483119]
72. Rajabi Majd N, Broström A, Ulander M, Lin CY, Griffiths MD, Imani V, et al. Efficacy of a theory-based cognitive behavioral technique app-based intervention for patients with insomnia: randomized controlled trial. J Med Internet Res 2020 Apr 01;22(4):e15841 [FREE Full text] [doi: 10.2196/15841] [Medline: 32234700]
73. Mak WW, Chan AT, Cheung EY, Lin CL, Ngai KC. Enhancing web-based mindfulness training for mental health promotion with the health action process approach: randomized controlled trial. J Med Internet Res 2015 Jan 19;17(1):e8 [FREE Full text] [doi: 10.2196/jmir.3746] [Medline: 25599904]
74. Mehri M, Chehrzad MM, Maleki M, Kousha M, Akhlaghi E, Mardani A. The effect of behavioral parent training of children with attention deficit hyperactivity disorder on parents' mental health. Neurol Psychiatry Brain Res 2020 Sep;37:53-59. [doi: 10.1016/j.npbr.2020.06.003]
75. Shamshiri Milani H, Azargashb E, Beyraghi N, Defaie S, Asbaghi T. Effect of telephone-based support on postpartum depression: a randomized controlled trial. Int J Fertil Steril 2015;9(2):247-253 [FREE Full text] [doi: 10.22074/ijfs.2015.4246] [Medline: 26246884]

76. Moeini B, Bashirian S, Soltanian AR, Ghaleiha A, Taheri M. Examining the effectiveness of a web-based intervention for depressive symptoms in female adolescents: applying social cognitive theory. J Res Health Sci 2019 Aug 19;19(3):e00454 [FREE Full text] [Medline: 31586376]
77. Mogoaşe C, Brăilean A, David D. Can concreteness training alone reduce depressive symptoms? A randomized pilot study using an internet-delivered protocol. Cogn Ther Res 2013 Jan 3;37(4):704-712. [doi: 10.1007/s10608-012-9514-z]
78. Newman MG, Kanuri N, Rackoff GN, Jacobson NC, Bell MJ, Taylor CB. A randomized controlled feasibility trial of internet-delivered guided self-help for generalized anxiety disorder (GAD) among university students in India. Psychotherapy (Chic) 2021 Dec;58(4):591-601 [FREE Full text] [doi: 10.1037/pst0000383] [Medline: 34881930]
79. Ngai FW, Wong PW, Leung KY, Chau PH, Chung KF. The effect of telephone-based cognitive-behavioral therapy on postnatal depression: a randomized controlled trial. Psychother Psychosom 2015;84(5):294-303. [doi: 10.1159/000430449] [Medline: 26278623]
80. Nobakht Z, Rassafiani M, Hosseini SA, Hosseinzadeh S. A web-based daily care training to improve the quality of life of mothers of children with cerebral palsy: a randomized controlled trial. Res Dev Disabil 2020 Oct;105:103731 [FREE Full text] [doi: 10.1016/j.ridd.2020.103731] [Medline: 32659699]
81. Ofoegbu TO, Asogwa U, Otu MS, Ibenegbu C, Muhammed A, Eze B. Efficacy of guided internet-assisted intervention on depression reduction among educational technology students of Nigerian universities. Medicine (Baltimore) 2020 Feb;99(6):e18774 [FREE Full text] [doi: 10.1097/MD.0000000000018774] [Medline: 32028389]
82. Osborn TL, Rodriguez M, Wasil AR, Venturo-Conerly KE, Gan J, Alemu RG, et al. Single-session digital intervention for adolescent depression, anxiety, and well-being: outcomes of a randomized controlled trial with Kenyan adolescents. J Consult Clin Psychol 2020 Jul;88(7):657-668. [doi: 10.1037/ccp0000505] [Medline: 32391709]
83. Pakrad F, Ahmadi F, Grace SL, Oshvandi K, Kazemnejad A. Traditional vs extended hybrid cardiac rehabilitation based on the continuous care model for patients who have undergone coronary artery bypass surgery in a middle-income country: a randomized controlled trial. Arch Phys Med Rehabil 2021 Nov;102(11):2091-101.e3. [doi: 10.1016/j.apmr.2021.04.026] [Medline: 34175270]
84. Peng X, Su Y, Hu Z, Sun X, Li X, Dolansky M, et al. Home-based telehealth exercise training program in Chinese patients with heart failure: a randomized controlled trial. Medicine (Baltimore) 2018 Aug;97(35):e12069 [FREE Full text] [doi: 10.1097/MD.0000000000012069] [Medline: 30170422]
85. Esmaeili Rad M, Ahmadi F. A new method to measure and decrease the online social networking addiction. Asia Pac Psychiatry 2018 Dec;10(4):e12330. [doi: 10.1111/appy.12330] [Medline: 30175904]
86. Rahimi R, Hasanpour S, Mirghafourvand M, Esmaeilpour K. Effect of Hope-oriented group counseling on mental health of infertile women with failed IVF cycles: a randomized controlled trial. BMC Psychiatry 2021 Jun 02;21(1):286 [FREE Full text] [doi: 10.1186/s12888-021-03280-5] [Medline: 34078307]
87. Salamanca-Sanabria A, Richards D, Timulak L, Connell S, Mojica Perilla M, Parra-Villa Y, et al. A culturally adapted cognitive behavioral internet-delivered intervention for depressive symptoms: randomized controlled trial. JMIR Ment Health 2020 Jan 31;7(1):e13392 [FREE Full text] [doi: 10.2196/13392] [Medline: 32003749]
88. Shahdosti H, Mazlom SR, Vaghee S, Amini S. Evaluating the effect of planned online video visitations on anxiety and depression of patients at open heart intensive care unit: a randomized controlled trial. Iran Red Crescent Med J 2020 Aug 04;22(7):e102578 [FREE Full text] [doi: 10.5812/ircmj.102578]
89. Sivrikaya EC, Yilmaz O, Sivrikaya P. Dentist-patient communication on dental anxiety using the social media: a randomized controlled trial. Scand J Psychol 2021 Dec;62(6):780-786. [doi: 10.1111/sjop.12769] [Medline: 34333783]
90. Song Y, Xie X, Chen Y, Wang Y, Yang H, Nie A, et al. The effects of WeChat-based educational intervention in patients with ankylosing spondylitis: a randomized controlled trail. Arthritis Res Ther 2021 Mar 04;23(1):72 [FREE Full text] [doi: 10.1186/s13075-021-02453-7] [Medline: 33663558]
91. Srivastava P, Mehta M, Sagar R, Ambekar A. Smartteen- a computer assisted cognitive behavior therapy for Indian adolescents with depression- a pilot study. Asian J Psychiatr 2020 Apr;50:101970. [doi: 10.1016/j.ajp.2020.101970] [Medline: 32114331]
92. Stamm B, Girardon-Perlini NM, Pasqualoto AS, Beuter M, Magnago TS. Telephone intervention for anxiety management in oncology patients: a randomized clinical trial. Acta Paul Enferm 2018 Mar;31(2):137-143 [FREE Full text] [doi: 10.1590/1982-0194201800021]
93. Su JJ, Yu DS. Effects of a nurse-led eHealth cardiac rehabilitation programme on health outcomes of patients with coronary heart disease: a randomised controlled trial. Int J Nurs Stud 2021 Oct;122:104040. [doi: 10.1016/j.ijnurstu.2021.104040] [Medline: 34333211]
94. Taleban R, Zamani A, Moafi M, Jiryaee N, Khadivi R. Applications of text messaging, and bibliotherapy for treatment of patients affected by depressive symptoms. Int J Prev Med 2016 Mar 1;7:46 [FREE Full text] [doi: 10.4103/2008-7802.177889] [Medline: 27076884]
95. Tam CC, Li X, Benotsch EG, Lin D. A resilience-based intervention programme to enhance psychological well-being and protective factors for rural-to-urban migrant children in China. Appl Psychol Health Well Being 2020 Mar;12(1):53-76. [doi: 10.1111/aphw.12173] [Medline: 31290227]

96. Thitipitchayanant K, Somrongthong R, Kumar R, Kanchanakharn N. Effectiveness of self-empowerment-affirmation-relaxation (Self-EAR) program for postpartum blues mothers: a randomize controlled trial. Pak J Med Sci 2018;34(6):1488-1493 [FREE Full text] [doi: 10.12669/pjms.346.15986] [Medline: 30559809]
97. Tiburcio M, Lara MA, Martínez N, Fernández M, Aguilar A. Web-based intervention to reduce substance abuse and depression: a three arm randomized trial in Mexico. Subst Use Misuse 2018 Nov 10;53(13):2220-2231. [doi: 10.1080/10826084.2018.1467452] [Medline: 29768070]
98. Tol WA, Leku MR, Lakin DP, Carswell K, Augustinavicius J, Adaku A, et al. Guided self-help to reduce psychological distress in South Sudanese female refugees in Uganda: a cluster randomised trial. Lancet Glob Health 2020 Feb;8(2):e254-e263 [FREE Full text] [doi: 10.1016/S2214-109X(19)30504-2] [Medline: 31981556]
99. Torabizadeh C, Rousta S, Gholamzadeh S, Kojouri J, Jamali K, Parvizi MM. Efficacy of education delivery through multimedia and text messaging on the psychological parameters of patients scheduled for coronary angiography: a single-blind randomized controlled clinical trial. BMC Cardiovasc Disord 2021 Jan 04;21(1):3 [FREE Full text] [doi: 10.1186/s12872-020-01820-7] [Medline: 33397300]
100. Tulbure BT, Andersson G, Sălăgean N, Pearce M, Koenig HG. Religious versus conventional internet-based cognitive behavioral therapy for depression. J Relig Health 2018 Oct;57(5):1634-1648. [doi: 10.1007/s10943-017-0503-0] [Medline: 29067598]
101. Tulbure BT, Szentagotai A, David O, tefan S, Månsson KN, David D, et al. Internet-delivered cognitive-behavioral therapy for social anxiety disorder in Romania: a randomized controlled trial. PLoS One 2015 May 4;10(5):e0123997 [FREE Full text] [doi: 10.1371/journal.pone.0123997] [Medline: 25938241]
102. Wang H, Zhao Q, Mu W, Rodriguez M, Qian M, Berger T. The effect of shame on patients with social anxiety disorder in internet-based cognitive behavioral therapy: randomized controlled trial. JMIR Ment Health 2020 Jul 20;7(7):e15797 [FREE Full text] [doi: 10.2196/15797] [Medline: 32347799]
103. Wantanakorn P, Harintajinda S, Chuthapisith J, Anurathapan U, Rattanatamrong P. A new mobile application to reduce anxiety in pediatric patients before bone marrow aspiration procedures. Hosp Pediatr 2018 Oct;8(10):643-650. [doi: 10.1542/hpeds.2018-0073] [Medline: 30213798]
104. Wei N, Huang B, Lu SJ, Hu JB, Zhou XY, Hu CC, et al. Efficacy of internet-based integrated intervention on depression and anxiety symptoms in patients with COVID-19. J Zhejiang Univ Sci B 2020 May;21(5):400-404 [FREE Full text] [doi: 10.1631/jzus.B2010013] [Medline: 32425006]
105. Xia L. The effects of continuous care model of information-based hospital-family integration on colostomy patients: a randomized controlled trial. J Cancer Educ 2020 Apr;35(2):301-311. [doi: 10.1007/s13187-018-1465-y] [Medline: 30685831]
106. Yan Z, Liu F, Lin XP, Chen J, Huang QL, Chen SL, et al. WeChat-based remote follow-up management alleviates the home care burden and anxiety of parents of premature infants: randomized controlled study. Child Care Health Dev 2022 Jul;48(4):651-657. [doi: 10.1111/cch.12973] [Medline: 35083757]
107. Yang B, Liu JF, Xie WP, Cao H, Chen Q. The effects of WeChat follow-up management to improve the parents' mental status and the quality of life of premature newborns with patent ductus arteriosus. J Cardiothorac Surg 2021 Aug 21;16(1):235 [FREE Full text] [doi: 10.1186/s13019-021-01617-5] [Medline: 34419135]
108. Yang L, Wang X, Cui X. Patients' intensive telephone-based care program reduces depression in coronary artery disease patients and may contribute to favorable overall survival by decreasing depression. J Cardiovasc Nurs 2019;34(3):236-243. [doi: 10.1097/JCN.0000000000000571] [Medline: 30973493]
109. Yang M, Jia G, Sun S, Ye C, Zhang R, Yu X. Effects of an online mindfulness intervention focusing on attention monitoring and acceptance in pregnant women: a randomized controlled trial. J Midwifery Womens Health 2019 Jan;64(1):68-77. [doi: 10.1111/jmwh.12944] [Medline: 30695166]
110. Yardımcı T, Mert H. Web-based intervention to improve implantable cardioverter defibrillator patients' shock-related anxiety and quality of life: a randomized controlled trial. Clin Nurs Res 2019 Feb;28(2):150-164. [doi: 10.1177/1054773817741427] [Medline: 29134819]
111. Yeung A, Wang F, Feng F, Zhang J, Cooper A, Hong L, et al. Outcomes of an online computerized cognitive behavioral treatment program for treating Chinese patients with depression: a pilot study. Asian J Psychiatr 2018 Dec;38:102-107. [doi: 10.1016/j.ajp.2017.11.007] [Medline: 29146042]
112. Zengin M, Başoğul C, Yayan EH. The effect of online solution-focused support program on parents with high level of anxiety in the COVID-19 pandemic: a randomised controlled study. Int J Clin Pract 2021 Dec;75(12):e14839 [FREE Full text] [doi: 10.1111/ijcp.14839] [Medline: 34514706]
113. Zhang QL, Lei YQ, Liu JF, Cao H, Chen Q. Using telemedicine to improve the quality of life of parents of infants with CHD surgery after discharge. Int J Qual Health Care 2021 Sep 25;33(3):mzab133. [doi: 10.1093/intqhc/mzab133] [Medline: 34490460]
114. Zhang QL, Xu N, Huang ST, Cao H, Chen Q. WeChat-assisted pre-operative health education improves the quality of life of parents of children with ventricular septal defects: a prospective randomised controlled study. J Paediatr Child Health 2021 May;57(5):664-669. [doi: 10.1111/jpc.15307] [Medline: 33656232]

115. Zhang X, Lin P, Sun J, Sun Y, Shao D, Cao D, et al. Prenatal stress self-help mindfulness intervention via social media: a randomized controlled trial. J Ment Health (forthcoming) 2021 Jul 15:1-10. [doi: 10.1080/09638237.2021.1952947] [Medline: 34264775]
116. Zhang Y, Zhang B, Gan L, Ke L, Fu Y, Di Q, et al. Effects of online bodyweight high-intensity interval training intervention and health education on the mental health and cognition of sedentary young females. Int J Environ Res Public Health 2021 Jan 03;18(1):302 [FREE Full text] [doi: 10.3390/ijerph18010302] [Medline: 33401605]
117. Zhao M, You Y, Chen S, Li L, Du X, Wang Y. Effects of a web-based parent-child physical activity program on mental health in parents of children with ASD. Int J Environ Res Public Health 2021 Dec 07;18(24):12913 [FREE Full text] [doi: 10.3390/ijerph182412913] [Medline: 34948523]
118. Zheng Y, Wang W, Zhong Y, Wu F, Zhu Z, Tham YC, et al. A peer-to-peer live-streaming intervention for children during COVID-19 homeschooling to promote physical activity and reduce anxiety and eye strain: cluster randomized controlled trial. J Med Internet Res 2021 Apr 30;23(4):e24316 [FREE Full text] [doi: 10.2196/24316] [Medline: 33882021]
119. Zhianfar L, Nadrian H, Asghari Jafarabadi M, Espahbodi F, Shaghaghi A. Effectiveness of a multifaceted educational intervention to enhance therapeutic regimen adherence and quality of life amongst Iranian hemodialysis patients: a randomized controlled trial (MEITRA study). J Multidiscip Healthc 2020 Apr 14;13:361-372 [FREE Full text] [doi: 10.2147/JMDH.S247128] [Medline: 32341649]
120. Zhou K, Li J, Li X. Effects of cyclic adjustment training delivered via a mobile device on psychological resilience, depression, and anxiety in Chinese post-surgical breast cancer patients. Breast Cancer Res Treat 2019 Nov;178(1):95-103. [doi: 10.1007/s10549-019-05368-9] [Medline: 31342310]
121. Zhuang JX. Psychosomatic care plus WeChat communication relaxation therapy in patients with functional dyspepsia. World Chin J Dig 2017 May;25(13):1191 [FREE Full text] [doi: 10.11569/wcjd.v25.i13.1191]
122. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. J Educ Stat 1981;6(2):107-128. [doi: 10.2307/1164588]
123. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis with R: A Hands-On Guide. Boca Raton, FL, USA: Chapman and Hall; 2021.
124. Serrano-Ripoll MJ, Zamanillo-Campos R, Fiol-DeRoque MA, Castro A, Ricci-Cabello I. Impact of smartphone app-based psychological interventions for reducing depressive symptoms in people with depression: systematic literature review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 Jan 27;10(1):e29621 [FREE Full text] [doi: 10.2196/29621] [Medline: 35084346]
125. Wang Y, Lin Y, Chen J, Wang C, Hu R, Wu Y. Effects of internet-based psycho-educational interventions on mental health and quality of life among cancer patients: a systematic review and meta-analysis. Support Care Cancer 2020 Jun;28(6):2541-2552. [doi: 10.1007/s00520-020-05383-3] [Medline: 32179998]
126. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. In: Database of Abstracts of Reviews of Effects (DARE): Quality-Assessed Reviews. York, UK: Centre for Reviews and Dissemination; 1995.

Abbreviations

CBT: cognitive behavioral therapy
DMH: digital mental health
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
HIC: high-income country
LMIC: low- and middle-income country
PRISMA: Preferred Reporting Item for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by J Torous; submitted 29.09.22; peer-reviewed by P Watson, A Daros; comments to author 14.01.23; revised version received 27.01.23; accepted 30.01.23; published 20.03.23.

Please cite as:
Kim J, Aryee LMD, Bang H, Prajogo S, Choi YK, Hoch JS, Prado EL
Effectiveness of Digital Mental Health Tools to Reduce Depressive and Anxiety Symptoms in Low- and Middle-Income Countries: Systematic Review and Meta-analysis
JMIR Ment Health 2023;10:e43066
URL: https://mental.jmir.org/2023/1/e43066
doi:10.2196/43066
PMID:36939820

Review

The Role of Compassionate Technology in Blended and Digital Mental Health Interventions: Systematic Scoping Review

Charlotte van Lotringen[1], MSc; Benedetta Lusi[2], MA; Gerben J Westerhof[1], PhD; Geke D S Ludden[2], IR, PhD; Hanneke Kip[1], PhD; Saskia M Kelders[1], PhD; Matthijs L Noordzij[1], PhD

[1]Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

[2]Department of Design, Production and Management, University of Twente, Enschede, Netherlands

Corresponding Author:
Charlotte van Lotringen, MSc
Department of Psychology, Health and Technology
University of Twente
Drienerlolaan 5
Enschede, 7500 AE
Netherlands
Phone: 31 534892100
Email: c.m.vanlotringen@utwente.nl

Abstract

Background: An essential value in mental health care is compassion: awareness of suffering, tolerating difficult feelings in the face of suffering, and acting or being motivated to alleviate suffering. Currently, technologies for mental health care are on the rise and could offer several advantages, such as more options for self-management by clients and more accessible and economically viable care. However, digital mental health interventions (DMHIs) have not been widely implemented in daily practice. Developing and evaluating DMHIs around important mental health care values, such as compassion, could be key for a better integration of technology in the mental health care context.

Objective: This systematic scoping review explored the literature for previous instances where technology for mental health care has been linked to compassion or empathy to investigate how DMHIs can support compassion in mental health care.

Methods: Searches were conducted in the PsycINFO, PubMed, Scopus, and Web of Science databases, and screening by 2 reviewers resulted in 33 included articles. From these articles, we extracted the following data: technology types, goals, target groups, and roles of the technologies in the intervention; study designs; outcome measures; and the extent to which the technologies met a 5-step proposed definition of compassion.

Results: We found 3 main ways in which technology can contribute to compassion in mental health care: by showing compassion to people, by enhancing self-compassion in people, or by facilitating compassion between people. However, none of the included technologies met all 5 elements of compassion nor were they evaluated in terms of compassion.

Conclusions: We discuss the potential of compassionate technology, its challenges, and the need to evaluate technology for mental health care on compassion. Our findings could contribute to the development of compassionate technology, in which elements of compassion are explicitly embedded in its design, use, and evaluation.

(JMIR Ment Health 2023;10:e42403) doi:10.2196/42403

KEYWORDS

compassionate technology; digital mental health interventions; eHealth; e–mental health; compassion; mental health care

Introduction

Background

Currently, digital technologies for mental health care are on the rise [1,2]. Examples include not only internet-delivered cognitive behavioral therapy (CBT) but also virtual reality (VR) and artificial intelligence–enabled programs [3], mobile apps [4], socially assistive robotics [5], and serious games [6,7]. Given the increasing costs of care and growing waiting lists in this field, technology could be an important element of sustainable mental health care [3]. Technology offers a wealth of possibilities to complement health care professionals by taking over certain tasks, lightening their workload, and providing

innovative ways to deliver health care [8]. In addition, technology can make mental health care more accessible to clients, give them more insight into their own mental health, and offer flexibility through personalization [9,10]. Digital mental health interventions (DMHIs) can be used as stand-alone interventions or in blended treatments: combining *traditional* face-to-face psychotherapy techniques with digital interventions [11]. Blended treatment has the potential to offer the *best of both worlds*, as it could be used to better match the intensity of the treatment to the severity of a client's complaints and to enhance a client's self-management [12].

Many forms of DMHIs exist and are shown to be effective in clinical trials [13], and if used well, they can often be as effective as face-to-face treatments [14-17]. However, the acceptance and use of these technologies remain low among both clients and professionals [18,19]; therefore, the actual implementation of DMHIs in daily practice has been limited. Among clients, there are low levels of adherence to DMHIs [20,21]. A potential reason for this could be that current DMHIs are often very direct translations of nondigital interventions, such as web-based CBT interventions, which closely follow their evidence-based nondigital versions [22]. Owing to the larger focus on content (eg, established therapeutic techniques) than on the user (eg, how the user prefers to engage with them), DMHIs are often perceived as impersonal [22]. This seems to be a missed opportunity, given that allowing for personalization is one of the advantages that technology could offer. Thus, the possibilities technology offers are not being optimally used. A further challenge is that professionals show hesitation and a lack of digital skills to implement DMHIs [18]. Similar to what we see from the client's perspective, this hesitation in part stems from the doubt whether personal, empathic connections can be made or supported if DMHIs are used [18]. Moreover, professionals indicate that an explicit conceptual foundation for the use of DMHIs is currently lacking, meaning that it is not clear, and at best implicit, why and how DMHIs can be embedded in the mental health care system [19].

Compassionate Technology

Given these current barriers, a new conceptual foundation is needed to build a bridge between technology and the context of mental health care. This could be done by expanding our view of the design, use, and evaluation of DMHIs to include and center on the values that are foundational to mental health care. Values are personal or societal judgments of what is valuable and important in life [23]. Compassion is widely recognized as central and essential in mental health care or health care in general [24-26]. In short, compassion refers to the awareness of suffering and motivation to act to alleviate suffering. Although conceptions and practices surrounding compassion have existed in Buddhism for >2500 years, in Western psychology, the construct of compassion has become a topic of study only more recently [27].

After comparing and synthesizing earlier definitions and measures of compassion from science, religion, and health care, Strauss et al [28] proposed that compassion is a cognitive, affective, and behavioral process containing five elements: (1) recognizing suffering, (2) understanding the universality of suffering in the human experience, (3) feeling empathy for the person suffering and connecting with the distress (emotional resonance), (4) tolerating one's own uncomfortable feelings that arise in the face of suffering (eg, distress, anger, and fear), to remain open and accepting of the person suffering, and (5) acting or feeling motivated to alleviate suffering [28].

Compassion can be directed not only toward ourselves (often referred to as self-compassion [29]) and loved ones but also toward strangers and ultimately toward all humankind [30].

Although empathy is part of the compassionate elements, compassion is conceptually different from empathy. Empathy has been defined as the vicarious experience of another's emotions [31], where one understands, is affected by, and shares another's emotions and perspectives [32]. It does not involve a motivation to act to alleviate another person's suffering. More specifically, although empathy can also lead to behavioral outcomes, it is not part of the concept of empathy itself [33]. In contrast, the process of compassion is specifically a response to suffering and not to emotions in general. It not only entails connecting with another's distress but also understanding the universality of suffering in the human experience and tolerating uncomfortable feelings that can be aroused in response to the suffering person, so that one can remain open and accept this person and then act to alleviate their suffering [28]. The ability to tolerate uncomfortable feelings is essential for preventing empathic distress. Empathic distress occurs when one is more upset *by* another's suffering than one is concerned *for* the other [34]. It can lead people to close themselves off from suffering and tends to inhibit compassion [35,36]. Thus, compassion contains elements that are specifically relevant to mental health care, more so than empathy alone.

Embedding compassion in health care improves clinical outcomes, perceived quality of care, and patient satisfaction; strengthens the therapeutic alliance; and supports patient-centered care [28,37,38]. Although research is slowly beginning to include compassion as a factor in technology in general [39], research on the link between compassion and technology for mental health care is still scarce. An exception is a recent scoping review by Kemp et al [37], who investigated how digital technologies were being used by patients and professionals in the delivery of compassionate mental health care. In addition, it investigated the facilitators and barriers for the use of digital technology in the delivery of compassionate mental health care. The authors found that when used appropriately, digital technologies can facilitate and strengthen compassion and meaningful human connections in mental health care. Moreover, technology can create new means for relationships between mental health professionals and patients. Kemp et al [37] focused their review on compassionate care and examined DMHIs with a model of digital intersections with compassionate care [40]. In this review, we take a different but complementary approach, examining DMHIs with the elements of compassion in itself [28]. This way, we aimed to explore how DMHIs could support the different components of compassion as a process. We expect this process-view of compassion could be a helpful and practical guiding force to shape compassionate blended treatment.

As compassion is a pivotal value in the mental health care context, emphasizing the process of compassion as a central value in the design, evaluation, and use of mental health care technology could be key to make it more suitable to the needs of clients and professionals. In turn, this could increase the uptake and integration of technologies in current treatments, ultimately ensuring compassionate blended mental health care that realizes its potential. On the basis of the proposed definition of compassion by Strauss et al [28], technology that is designed, developed, and evaluated around the value of compassion would enable and facilitate elements of compassion, or in short, the recognition and alleviation of suffering. To the best of our knowledge, this concept of *compassionate technology* for mental health care has not been systematically studied. This highlights the need for a clear conceptualization of compassionate technology based on research in which DMHIs have been explicitly linked to compassion.

Research Objectives

This systematic scoping review provides an overview of how and to what degree elements of compassion have been linked to digital technologies for mental health or mental health care in previous studies. We used this information to describe the current status and scope of research on technology that is connected to compassion and to inform the future development of compassionate technology.

To reach these aims, we formulated the following research questions:

1. What types of technology for mental health care have been connected to compassion in previous studies, for what goals and which target groups were they developed, and what was the role of the technology in the intervention (eg, stand-alone or blended treatment)?
2. What study designs have previously been used to study these technologies, and what outcome measures are used?
3. To what extent and how do these technologies meet the 5 elements of compassion as distinguished by Strauss et al [28]?

Methods

Research Design

A systematic scoping review was conducted in accordance with existing guidelines [41]. This approach is particularly useful for bringing together the literature in disciplines with emerging evidence when a body of the literature has not yet been reviewed or exhibits a large, complex, or heterogeneous nature that is not suitable for a more precise systematic review [41]. Because of the novelty of the field and the presumed limited previous research on technologies that foster compassion, a systematic scoping review was deemed the appropriate method, as it focuses on appraising a body of literature on a specified topic in terms of extent, range, and nature [42].

Search Strategy

The search strategy was developed iteratively in consultation with an information specialist, as suggested by Horsley [43]. A systematic search was conducted between October and November 2020 using the PsycINFO, PubMed, Scopus, and Web of Science databases. The 4 databases were chosen because PsycINFO focuses on research on behavioral science and mental health, PubMed focuses on biomedicine (including psychiatry), and the other 2 databases include research from all disciplines. In this way, psychology, psychiatry, and technology research fell within the scope of our search. The query used a combination of terms related to *compassion*, *technology*, and *mental health* occurring in the title, abstract, or keywords of published articles (see Multimedia Appendix 1 for the search strings). The search terms used were chosen to focus specifically on the mental health field as well as on explicit mentions of either compassion or the related term empathy.

Eligibility Criteria

Because of the novelty of the subject, this review aims to identify articles covering any form of digital technology for the support of mental health that fosters the presence of compassion. Digital technologies linked to empathy instead of compassion were also eligible to avoid excluding articles that used the term "empathy" instead of compassion while pertaining to a similar construct. Our review did not focus on one specific type of participant in the included studies, so that articles including different types of people in and around the mental health field were eligible (eg, therapists, clients or patients, [informal] caregivers and also the public). All types of original research studies were eligible from any year of publication, country of origin, or original language (if a copy was available in English, German, Italian, or Dutch).

Exclusion criteria were as follows:

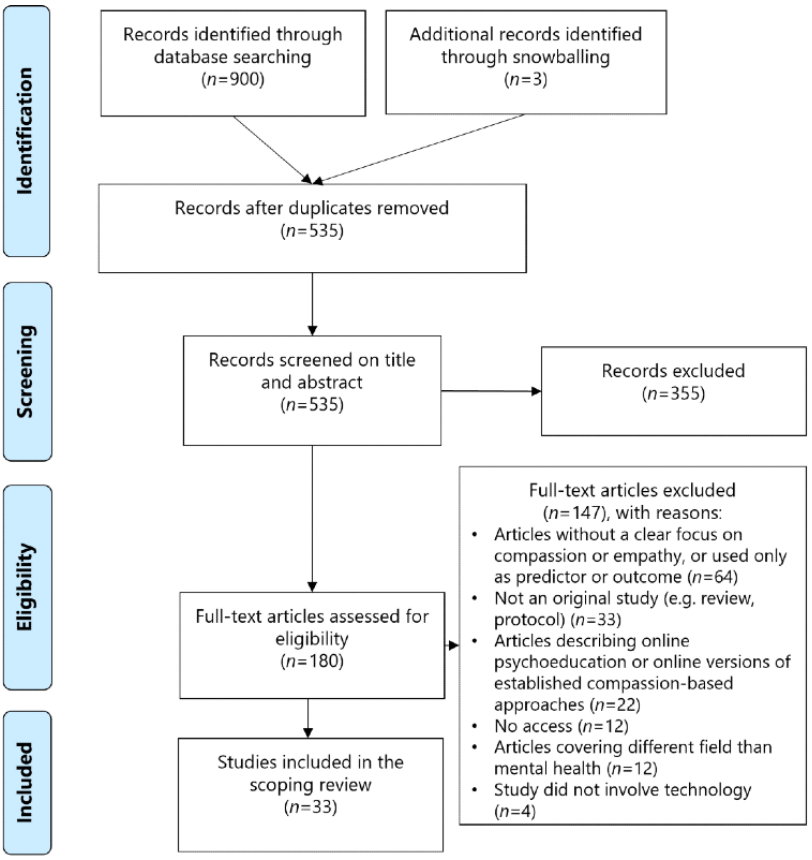
1. Articles without a clear focus on compassion or empathy, for example, merely mentioning compassion or empathy without further elaboration or using these factors solely as predictors or outcomes in a study.
2. Articles in which the link with compassion or empathy was *only* found as part of web-based versions of established compassion-based approaches (eg, Compassion-Focused Therapy [44]) or in the delivery of written psychoeducation (eg, a website with information on self-compassion), and compassion or empathy did not refer to (interactions with or through) the technology itself. These articles were excluded because a lot of research has already been conducted on these therapeutic approaches [45], and the role technology plays is relatively small, so that it is not directly relevant for our current purposes.
3. Articles related to compassion or empathy in a different field than mental health, for example, education or health care in general without a focus on mental health.

Study Selection

Covidence (a literature review screening software recommended by Cochrane [46]) was used to filter duplicate articles and facilitate study selection in 3 steps (Figure 1). First, the titles and abstracts of all retrieved articles were screened for eligibility by 2 authors (CvL and BL). Disagreements on the inclusion or exclusion of papers were discussed until an agreement was reached. Second, the full text of all remaining articles was checked for inclusion by one author (CvL) and doubts were

discussed with a second author (BL). If an agreement could not be reached between the 2 authors, a third researcher was consulted (MLN). Third, to check whether seminal works were overlooked during the initial search process, forward and backward snowballing by one author (CvL) was used, based on the reference lists and citations of the included papers.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection process for the systematic scoping review.



Data Extraction

Overview

Data that were extracted from each included article were author or authors, country of origin, year of publication, type of technology, target group or groups, goal or goals of the technology, the role of technology in the intervention, study design, outcome measures, elements of compassion, and compassionate role of technology. The role of technology in the intervention was coded depending on whether a technology was used by the target group on their own without guidance ("stand-alone"); *only* to deliver treatment sessions with a therapist or coach ("telecommunication"); with guidance from a coach or therapist, which was not face-to-face ("guided"); or was integrated into a treatment with face-to-face therapy sessions ("blended treatment"). Study designs were classified as qualitative, quantitative cross-sectional, quasi-experimental, experimental, or a combination (adapted from Centre for Evidence-Based Medicine [47]), with quasi-experimental studies referring to nonrandom allocation to groups and experimental studies referring to random allocation. For consistency, the term "outcome measures" was used for both quantitative and qualitative study designs, in the latter cases referring to the investigated variables. Elements of compassion were first described by closely following the author's wording and then coded in the matching compassionate element (recognition, universality, empathy, tolerance, acting) by comparing them to their descriptions in Strauss et al [28]. Thus, the final results table includes the coded compassionate element and how this was described in the included study.

As we noticed that mentions of empathy or compassion in the included articles referred to different processes that technology could support, we divided articles into 3 "compassionate roles," used to structure our results. These compassionate roles of technology have been categorized into different codes through inductive coding [48] using the method of constant comparison [49]. The roles were coded depending on whether the mentions of compassionate elements were mainly used to describe the features of the technology itself (Role A), to describe interventions for self-compassion (Role B), or to describe technology to facilitate compassionate elements between people (Role C). The characteristics of all included studies were extracted by 1 author (CvL), and data extraction of 15% of the included studies was validated by a second author (BL) with 82% agreement. This percentage of agreement was deemed sufficient to continue data extraction by one author. Data extraction occurred iteratively in consultation with several of the coauthors, where doubts were discussed until consensus was reached.

Main Perspective for Qualitative Analysis

In qualitative research, it is valuable to consider and describe the perspective held by researchers and how this could have influenced the research [50,51]. Therefore, we provide a short description of the background and perspective of the first author (CvL), focusing on elements that might have influenced the research process and vision presented in this paper.

CvL is a Dutch woman born in the Netherlands and raised in a nonreligious environment. Compassionate technology for mental health care is the topic of her PhD research project. The aim of this project was to investigate how technology in mental health care can be integrated into daily mental health care practice, where compassion is a fundamental value. This project took place in collaboration with researchers from different disciplines, a mental health care organization, and an eHealth company. CvL's prior understanding of the topic comprises literature research on compassion, technology, and mental health care, as well as interviews with mental health care professionals, clients, and developers of DMHIs and observations of ways of working of mental health care professionals. Compassion is also an important personal value to her, in the sense of feeling responsibility for the world around you and wanting to contribute positively to society.

Results

Overview

On the basis of the inclusion and exclusion criteria, we included 33 studies (Tables 1-3), covering 31 unique DMHIs. These studies were published between 2008 and 2020, with the majority published between 2016 and 2020 (25/33, 76%) and conducted in Western countries (31/33, 94%). The technologies we found could be divided into 3 roles based on the main way the technology contributed to the presence of compassion (Figure 2). The technology could show elements of compassion *to* people (Role A, n=8), technology could enhance elements of self-compassion *in* people (Role B, n=12), and technology could facilitate elements of compassion *between* people (Role C, n=13).

Figure 2. Schematic display of the 3 identified roles in which technology can contribute to compassion.

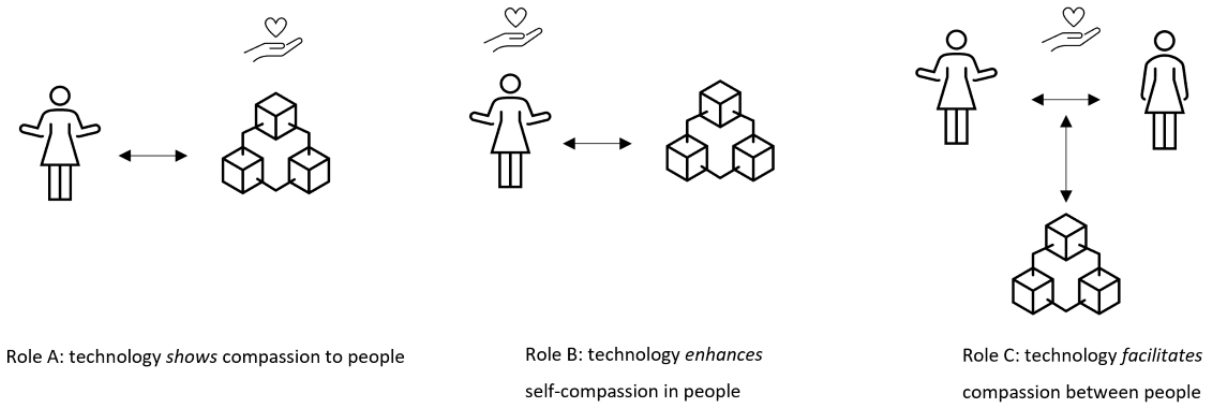


Table 1. Extracted data from the included studies classified as Role A: technology shows elements of compassion to people (n=8).

Role and study	Technology (technology, target group, goal, and role in intervention)	Study (study design and outcome measures)	Compassionate elements (on the basis of the 5 elements proposed by Strauss et al [28])
Hauser-Ulrich et al [52], 2020, Switzerland	• Technology: mobile health intervention using a fully automated text-based health care chatbot (SELMA) • Target group: patients with ongoing or cyclic pain • Goal: to deliver personalized intervention modules for mental health and psychoeducation for pain management, and to build a working alliance between a participant and the chatbot • Role in intervention: stand-alone	• Design: experimental (pilot RCT[a]) • Outcome measures: symptom and well-being outcomes, working alliance, adherence, and acceptability	• Recognition[b]: chatbot enquires about the participants' mood • Empathy[c]: chatbot replies in an empathic way, and is also experienced as such by participants • Acting[d]: chatbot addresses participants' accountability by supporting the completion of activities and tasks and motivating participants to repeat them
Morris et al [53], 2018, United States	• Technology: text-based conversational agent in a peer-to-peer platform, that draws and repurposes responses from a pool of peer-support data (Koko) • Target group: users of Koko • Goal: to express empathic support • Role in intervention: stand-alone	• Design: qualitative and experimental • Outcome measures: ratings of the conversational agent's responses ("good, okay, bad")	• Empathy: conversational agent gives a nuanced and empathic reply addressing specific elements of the user's message
Javed and Park [54], 2019, United States	• Technology: socially assistive robot • Target group: children with autism spectrum disorder • Goal: to regulate the user's emotions and prime higher social engagement in a user • Role in intervention: stand-alone	• Design: qualitative • Outcome measures: engagement in intervention	• Recognition: the user selects their current emotion state from an input panel • Empathy: after the user indicated their current emotion state, the robot's initial display of emotions closely follow the user's • Acting: the robot's subsequent emotion selections lead the user toward a desired (goal) emotional state, to support emotion regulation
Pontier and Siddiqui [55], 2008, Netherlands	• Technology: web-based agent • Target group: people filling in the BDI[e] • Goal: to guide the user through the BDI and respond empathically to the answers given by the user • Role in intervention: stand-alone	• Design: quasi-experimental • Outcome measures: ratings of how friendly, interested, trustworthy, and kind the web-based agent was	• Empathy: the web-based agent responds to the answers given on the BDI by the user, either by showing happiness or a neutral expression when the user seems fine, or sadness when the user appears to be more depressed
Kohori et al [56], 2018, Japan	• Technology: interactive therapy robot • Target group: people who have difficulty living with pets • Goal: to induce a "healing mood" • Role in intervention: stand-alone	• Design: quantitative cross-sectional and qualitative • Outcome measures: design elements for inducing a healing mood, acceptability	• Empathy: the robot can recognize and track human faces and nod while someone is talking
Fitzpatrick et al [57], 2017, United States	• Technology: fully automated conversational agent, desktop or mobile accessible (Woebot) • Target group: college students self-reporting symptoms of anxiety and depression • Goal: to deliver a self-help program in a convenient and engaging way • Role in intervention: stand-alone	• Design: experimental (RCT) • Outcome measures: symptom and well-being outcomes, acceptability, and usability	• Recognition: participants could indicate their mood, and received a weekly mood description • Empathy: the bot replied in an empathic way, appropriate to participants' inputted mood • Acting: the bot presented CBT[f]-based content based on users' mood state, asked for and checked up on users' personal goals and sent personalized messages for motivation

Role and study	Technology (technology, target group, goal, and role in intervention)	Study (study design and outcome measures)	Compassionate elements (on the basis of the 5 elements proposed by Strauss et al [28])
Bickmore et al [58], 2010, United States	• Technology: animated conversational agent • Target group: patients with schizophrenia • Goal: to promote antipsychotic medication adherence, physical activity, system use and foster the therapeutic alliance • Role ins intervention: stand-alone	• Design: quasi-experimental (pilot evaluation study) • Outcome measures: system use and interaction times, medication adherence, and physical activity	• Empathy: the agent responds empathically, by expressing verbal and nonverbal concern for a negative state of the user • Acting: agent asks about medication-taking behavior, provides feedback based on self-monitoring charts, and reminds user of goal behavior and when to get medication refills
Rouaix et al [59], 2017, France	• Technology: humanoid robot (NAO), with Wizard of Oz approach • Target group: people with dementia • Goal: to act as therapist's assistant in psychomotor therapy, as a mediating tool • Role in intervention: blended treatment	• Design: experimental • Outcome measures: well-being outcomes, engagement, satisfaction with intervention, appreciation of robot, and empathy-related behaviors in human-robot interaction	• Recognition: robot is able to acknowledge the participants emotions and feelings • Empathy: robot exhibits empathic gestures such as giving confirmation signs by head movements • Acting: robot gives positive feedback and acknowledges participant's performance to boost confidence and motivation

[a]RCT: randomized controlled trial.

[b]Recognizing suffering.

[c]Feeling empathy for the person suffering and connecting with the distress (emotional resonance).

[d]Motivation to act or acting to alleviate suffering.

[e]BDI: Beck Depression Inventory.

[f]CBT: cognitive behavioral therapy.

Table 2. Extracted data from the included studies classified as Role B: technology enhances elements of self-compassion in people (n=12).

Role and study	Technology (technology, target group, goal, and role in intervention)	Study (study design and outcome measures)	Compassionate elements (on the basis of the 5 elements proposed by Strauss et al [28])
Ascone et al [60], 2020, Germany	• Technology: immersive compassion-focused imagery VR[a] intervention • Target group: student sample with slightly elevated symptoms of paranoia • Goal: to enable a sense of presence and induce specific emotional responses to support the development of self-compassionate feelings • Role in intervention: stand-alone	• Design: experimental • Outcome measures: symptom and well-being outcomes, self-compassion, and self-rated intervention benefit	• Universality[b]: the VR intervention was set in space, to evoke an overview effect and elicit feelings of connectedness with humanity and self-transcendence • Acting[c]: participants were guided into opening for feelings of kindness, warmth, wisdom, courage, and strength while also interacting with a space nebula representing compassion, which reacted to touch by intensifying its glow
Brown et al [61], 2020, United Kingdom	• Technology: VR with compassionate coach imagery • Target group: individuals scoring highly for paranoia • Goal: to reduce negative beliefs about the self, and hence paranoia • Role in intervention: stand-alone	• Design: experimental (RCT[d]) • Outcome measures: symptom and well-being outcomes and self-compassion	• Empathy[e]: participants created a personal compassionate coach who provided strength, kindness, and warmth • Acting: the compassionate coach helped participants to feel better able to cope with everyday challenges and to be more self-compassionate
Pennesi and Wade [62], 2018, Australia	• Technology: web-based imagery re-scripting intervention • Target group: body-dissatisfied young women at risk of developing an eating disorder • Goal: to reduce disordered eating by strengthening protective factors (ie, self-compassion) • Role in intervention: stand-alone	• Design: experimental (RCT) • Outcome measures: symptom and well-being outcomes and self-compassion	• Empathy: patients were asked to imagine and write about their earliest memory of an unpleasant body experience from both an observers' perspective and in the present with their wiser and more compassionate self in the room
Falconer et al [63], 2016, United Kingdom	• Technology: VR • Target group: patients with major depressive disorder • Goal: to facilitate self-compassion through embodiment, also for people who find it difficult to be self-compassionate • Role in intervention: stand-alone	• Design: quasi-experimental • Outcome measures: symptom and well-being outcomes, self-compassion, and "VR experience" (eg, feeling of being comforted)	• Acting: participants were provided with sentences to reduce distress based on compassion-focused therapy and asked to deliver them compassionately to a virtual child and then experienced them from the perspective of the child
Falconer et al [64], 2014, United Kingdom	• Technology: VR • Target group: healthy females high in self-criticism • Goal: to foster self-compassion through embodiment • Role in intervention: stand-alone	• Design: experimental • Outcome measures: symptom and well-being outcomes and "VR experience"	• Acting: participants were provided with sentences to reduce distress based on compassion-focused therapy and asked to deliver them compassionately to a virtual crying child, and then experienced them from the perspective of the child
Donovan et al [65], 2016, United States	• Technology: mobile phone app (BodiMojo) • Target group: adolescents • Goal: to foster coping and well-being during adolescence • Role in intervention: stand-alone	• Design: quantitative and qualitative • Outcome measures: use data and satisfaction	• Recognition[f]: the app allows the user to track their feelings in a mood cloud, giving a visual representation of their mood and get personalized feedback • Universality: daily wellness tips can be linked to emphasizing common humanity • Acting: daily wellness tips prompt users to engage in mindfulness and self-compassion activities
Rodgers et al [66], 2018, United States	• Technology: mobile phone app (BodiMojo) • Target group: adolescents • Goal: to promote positive body image and self-compassion • Role in intervention: stand-alone	• Design: experimental (RCT) • Outcome measures: symptom and well-being outcomes and self-compassion	• Recognition: the app allows the user to track their feelings daily in a mood cloud, giving a visual representation of their mood, and clicking the feelings gave them access to supportive emotional regulation statements • Universality: the intervention messages can be linked to emphasizing common humanity • Acting: users get intervention messages twice daily, and these prompt users to engage in mindfulness and self-compassion activities

Role and study	Technology (technology, target group, goal, and role in intervention)	Study (study design and outcome measures)	Compassionate elements (on the basis of the 5 elements proposed by Strauss et al [28])
Fonseca et al [67], 2019, Portugal	• Technology: self-guided, web-based intervention (Be a Mom) • Target group: at risk postpartum women and women presenting early-onset postpartum depressive symptoms • Goal: to prevent persistent postpartum depression symptoms • Role in intervention: stand-alone	• Design: experimental (pilot RCT) • Outcome measures: symptom and well-being outcomes and self-compassion	• Recognition: intervention helps women to be aware of and understand their emotions and thoughts • Universality: intervention offered exercises to help women accept that they are vulnerable and human like all mothers • Tolerance[g]: intervention helps women to nonjudgmentally accept difficult emotions • Acting: intervention helps women to use more self-compassionate ways to deal with their experiences, and addresses perinatal-specific concerns
Schnepper et al [68], 2020, Austria	• Technology: mobile self-compassion intervention • Target group: people that want to lose weight or develop healthier eating behavior • Goal: to improve eating behavior, self-compassion, and stress levels • Role in intervention: stand-alone	• Design: experimental (RCT) • Outcome measures: symptom and well-being outcomes and self-compassion	• Universality: participants learned to see negative emotions as part of being human • Tolerance: participants learned to be mindful and less critical about negative emotions • Acting: journaling exercises explored how participants could find less critical ways to motivate themselves to improve eating behavior
Raymond [69], 2019, United States	• Technology: SMS text messaging • Target group: undergraduate psychology students • Goal: to enhance psychological interventions • Role in intervention: blended treatment	• Design: experimental (RCT) • Outcome measures: well-being outcomes	• Acting: in addition to 3 daily texts with self-compassion content, participants could request additional texts to receive a recommendation to help manage strong emotions or a self-compassion quote
Lee et al [70], 2019, Netherlands	• Technology: chatbot that was either caregiving or care-receiving (Vincent) • Target group: nonclinical population • Goal: to increase self-compassion • Role in intervention: stand-alone	• Design: experimental • Outcome measures: self-compassion, opinions about the agent, inclusion of other in the self (identification with chatbot), and engagement	• Universality: the care-receiving chatbot likely increased self-compassion through a substantial change in participants' sense of common humanity
Köhle et al [71], 2017, Netherlands	• Technology: web-based self-help intervention based on Acceptance and Commitment Therapy and self-compassion (Hold on, for each other) • Target group: partners of patients with cancer • Goal: to help partners to positively persevere during the difficult times they find themselves facing • Role in intervention: guided	• Design: qualitative • Outcome measures: user experiences (appreciation of intervention and their lessons learned)	• Recognition: partners learn to recognize and be aware of their own emotions • Universality: partners have the option to connect with peers, eg, to share their answers on exercises and read those given by others • Tolerance: partners learn how to be accepting of their difficult emotions

[a]VR: virtual reality.

[b]Understanding the universality of suffering in the human experience.

[c]Motivation to act or acting to alleviate suffering.

[d]RCT: randomized controlled trial.

[e]Feeling empathy for the person suffering and connecting with the distress (emotional resonance).

[f]Recognizing suffering.

[g]Tolerating uncomfortable feelings aroused in response to the suffering person (eg, distress, anger, fear) so remaining open to and accepting of the person suffering.

Table 3. Extracted data from the included studies classified as Role C: technology supports elements of compassion between people (n=13).

Role and study	Technology (technology, target group, goal, and role in intervention)	Study (study design and outcome measures)	Compassionate elements (on the basis of the 5 elements proposed by Strauss et al [28])
Okita [72], 2013, United States	• Technology: therapeutic robot companion (Paro) • Target group: pediatric patients and their parents • Goal: to reduce pain and emotional anxiety in patients and their parents • Role in intervention: stand-alone	• Design: experimental • Outcome measures: child's pain and parent's empathetic pain and perception of child's pain	• Recognition[a]: parents could acknowledge the patient's pain more accurately through robot-assisted therapy • Empathy[b]: robot was used as a social agent to generate perspective taking through a shared common experience and seemed to enhance parent's ability to empathize directly with the child • Acting[c]: engaging with the robot together reduced pain and emotional anxiety in the patients and reduced empathetic pain in the parents
Choo et al [73], 2016, United States	• Technology: web-based intervention with a "booster" phone call • Target group: women with coexisting intimate partner violence and substance use disorders • Goal: to address violence and drug use among women patients in the emergency department • Role in intervention: guided	• Design: qualitative and quantitative cross-sectional • Outcome measures: satisfaction, usability, and consistency with motivational interviewing	• Universality[d]: technology is a link to social support, to prevent isolation • Empathy: the experience of the intervention was personal and empathetic
Bar-Lev [74], 2008, Israel	• Technology: online support group • Target group: people with HIV or AIDS • Goal: to provide a web-based community center, with medical information, job postings, links to community services, and a public discussion group • Role in intervention: stand-alone	• Design: qualitative • Outcome measures: emotional dynamics in online support groups	• Universality: participants in online support groups create emotionally vibrant, empathic communities by describing and sharing their experiences
Wijma et al [75], 2018, Netherlands	• Technology: VR[e] intervention with a VR simulation movie and e-course (Through the D'mentia Lens) • Target group: informal caregivers of people with dementia • Goal: to enhance understanding and empathy in caregivers of people with dementia • Role in intervention: stand-alone	• Design: quasi-experimental (pilot study) • Outcome measures: feasibility, acceptance, caregiver's person-centeredness, empathy, perceived pressure from care and perceived competence, and quality of relationship	• Empathy: the intervention strengthened the ability of the participant to empathize with the person with dementia they take care of • Tolerance[f]: and gave informal caregivers more confidence in their care task and a more positive attitude toward it by increasing their resilience
Han et al [76], 2011, United States	• Technology: online support groups • Target group: low-income women with breast cancer • Goal: to help cope with illness • Role in intervention: stand-alone	• Design: qualitative • Outcome measures: empathy and emotional support expression and reception and breast cancer–related concerns	• Universality: online support groups provide a community to connect with people going through similar experiences • Empathy: online support groups provide patients with a space to share illness experiences, feelings, and concerns, and these self-disclosing activities stimulate empathic responses from others • Acting: participants can read and write empathic messages to respond compassionately to each other's distress, and this seems to reduce breast cancer concerns

Role and study	Technology (technology, target group, goal, and role in intervention)	Study (study design and outcome measures)	Compassionate elements (on the basis of the 5 elements proposed by Strauss et al [28])
Högberg et al [77], 2018, Sweden	• Technology: web-based communication service • Target group: patients with hematological diseases • Goal: to offer patients the possibility to request support from a nurse • Role in intervention: telecommunication	• Design: qualitative • Outcome measures: nurse's abilities of compassion, competence, and upholding trust	• Recognition: patients can share personal everyday experiences and their worries, and the nurse can explicitly express recognition • Empathy: the nurse can express caring and interest to respond compassionately and reflect the tone of the message • Tolerance: the nurse can respond in a supportive way by expressing acceptance and validation • Acting: patients can explicitly request for direct actions, which are achieved because of the communication rather than via the communication per se
Steinwachs et al [78], 2011, United States	• Technology: interactive web-based intervention (YourSchizophreniaCare) • Target group: patients with schizophrenia • Goal: to help patients to discuss their mental health treatment with their therapist • Role in intervention: stand-alone	• Design: experimental • Outcome measures: client's and clinician's respective contributions to dialogue and clinician's empathy	• Recognition: patients answer questions about their current status and treatment, eg, how often their medication makes them restless • Acting: on the basis of their answers, they can get recommendations to discuss a topic with the therapist, and feedback on how to do it effectively
Kysely et al [79], 2020, Australia	• Technology: videoconferencing • Target group: couples • Goal: to deliver psychotherapy, specifically relationship interventions • Role in intervention: telecommunication	• Design: qualitative • Outcome measures: expectations and experiences with videoconferencing, ie, empathy	• Empathy: ambiguous; the distance between therapist and client in videoconferencing can stimulate the client to open up more because of feeling safe and empowered, but it can also be experienced as impersonal
Blair Irvine et al [80], 2012, United States	• Technology: web-based training program, multimedia (Caring Skills: working with Mental Illness) • Target group: licensed care staff working in long-term care facilities • Goal: to increase empathy and decrease stigmatization toward residents with mental illness • Role in intervention: telecommunication	• Design: experimental (RCT[s]) • Outcome measures: care staff member's self-efficacy, attitudes toward people with mental illnesses, behavior intentions, empathy, acceptance, and usability	• Empathy: intervention stimulates engaging the resident in conversation to attempt to understand their perspective, listening to them and acknowledging their emotion • Tolerance: intervention conveys a person-centered care philosophy, where resident's potentially problematic behavior is seen as an expression of an unmet need, rather than just behavior to be managed, and promotes self-care for the caregiver after upsetting interactions with residents • Acting: training includes behavioral skills to work with mental illness behavior, such as acknowledging the resident's emotions and finding a suitable redirection
van Rijn et al [81], 2017, United Kingdom	• Technology: VR-on a laptop, not "immersive" (ProReal) • Target group: prisoners in a therapeutic community prison • Goal: to improve mental health outcomes and mental well-being, as an addition to the existing therapeutic intervention • Role in intervention: blended treatment	• Design: qualitative and quasi-experimental • Outcome measures: symptom and well-being outcomes, engagement, and quality of relationships between participants and with counselor	• Recognition: the program enables clients to make their thoughts, feelings, and experiences visible in a web-based representation with avatars, supporting their self-expression and understanding • Empathy: the program helps to reflect on situations and experiences and to see them from different perspectives that clients could empathize with
Fordham and Ball [82], 2019, United States	• Technology: digital game (Hellblade: Senua's Sacrifice) • Target group: general public • Goal: to create embodied experiences of mental health and promote empathic understanding (in this case focused on representations of psychosis) • Role in intervention: stand-alone	• Design: qualitative • Outcome measures: the design of embodied experiences of mental illness	• Empathy: the game allows the player to experience symptoms similar to a psychosis (eg, auditory and visual hallucinations)

Role and study	Technology (technology, target group, goal, and role in intervention)	Study (study design and outcome measures)	Compassionate elements (on the basis of the 5 elements proposed by Strauss et al [28])
Tippin and Maranzan [83], 2019, Canada	• Technology: web-based antistigma video intervention (Photovoice) • Target group: general public • Goal: to reduce mental illness stigma • Role in intervention: stand-alone	• Design: experimental (RCT) • Outcome measures: stigma toward mental illness and empathic concern	• Empathy: intervention conveys lived experiences with mental illness and evokes empathic concern in watcher • Tolerance: intervention led to decreased anger and fear toward people with a mental illness
Milbury et al [84], 2020, United States	• Technology: web-based meditation intervention for couples via FaceTime (led by counselor) • Target group: patients with primary and metastatic brain tumors and their partners • Goal: to target symptom and well-being outcomes • Role in intervention: telecommunication	• Design: experimental (pilot RCT) • Outcome measures: symptoms and well-being outcomes, compassion in couples' relationship, and feasibility	• Recognition: couples were introduced to meditation techniques by guiding their awareness to their current experience • Empathy: couples participated in meditations that focused their attention on their interconnectedness and their feelings of compassion for the self and the partner • Tolerance: participants were asked to share experiences with each other in a state of nonjudgmental and accepting awareness

[a]Recognizing suffering.

[b]Feeling empathy for the person suffering and connecting with the distress (emotional resonance).

[c]Motivation to act or acting to alleviate suffering.

[d]Understanding the universality of suffering in the human experience.

[e]VR: virtual reality.

[f]Tolerating uncomfortable feelings aroused in response to the suffering person (eg, distress, anger, fear) so remaining open to and accepting of the person suffering.

[g]RCT: randomized controlled trial.

Technologies' Types, Goals, Target Groups, and Roles in the Intervention

To answer the first research question, we will describe the different types, goals, target groups, and roles in technological interventions for each compassionate role. Table 4 shows a numeric summary of the results, including the frequencies and references of the variables that had a limited number of clear subtypes (technology types, roles in intervention, and compassionate elements). In Role A, we found 8 technologies showing elements of compassion for a person. An example is a chatbot that replies empathically to the person's input and motivates them to complete certain activities, such as modules based on CBT [52]. For this role, the types of technology were mainly chatbots or conversational agents [52,53,57,58] and social robots [54,56,59]. The most identified goal was for technology to express empathic support and to foster the therapeutic alliance between the person and technology [52,53,55,58]. The target groups included different groups of people, such as people with dementia [59], schizophrenia [58], and ongoing or cyclic pain [52]. Some target groups were more general, such as "people who have difficulty living with pets" [56], or "people filling in the Beck's Depression Inventory" [55]. For Role A, we found almost exclusively stand-alone interventions [52-58], meaning that they were used by the target group without any guidance from a coach or therapist. One exception was a social robot used in blended treatment [59], where the robot functioned as an assistant to the therapist in psychomotor therapy.

Table 4. Numerical summary of the extracted variables that had a limited number of clear subtypes, including their frequencies and references, per compassionate role.

Compassionate role, variable, and subtypes	Frequency, n (%)	References
Role A (n=8)		
Technologies		
Chatbot or conversational agent	4 (50)	[52,53,57,58]
Social robot	3 (38)	[54,56,59]
Web-based agent	1 (12)	[55]
Roles in interventions		
Stand-alone	7 (88)	[52-58]
Blended treatment	1 (12)	[59]
Compassionate elements		
Recognition	4 (50)	[52,54,57,59]
Universality	N/A[a]	N/A
Empathy	8 (100)	[52-59]
Tolerance	N/A	N/A
Acting	5 (63)	[52,54,57-59]
Role B (n=12)		
Technologies		
Virtual reality	4 (33)	[60,61,63,64]
Mobile phone app or SMS text messaging	4 (33)	[65,66,68,69]
Web-based intervention	3 (25)	[62,67,71]
Chatbot	1 (8)	[70]
Roles in interventions		
Stand-alone	10 (83)	[60-64,66-68,70]
Guided	1 (8)	[71]
Blended treatment	1 (8)	[69]
Compassionate elements		
Recognition	4 (33)	[65-67,71]
Universality	7 (58)	[60,65-68,70,71]
Empathy	2 (17)	[61,62]
Tolerance	3 (25)	[67,68,71]
Acting	9 (75)	[60,61,63-69]
Role C (n=13)		
Technologies		
Web-based intervention	5 (38)	[73,77,78,80,83]
Online support group	2 (15)	[74,76]
Videoconferencing	2 (15)	[79,84]
Virtual reality	2 (15)	[75,81]
Social robot	1 (8)	[72]
Digital game	1 (8)	[82]
Roles in interventions		
Stand-alone	7 (54)	[72,74-76,78,82,83]
Telecommunication	4 (31)	[77,79,80,84]

Compassionate role, variable, and subtypes	Frequency, n (%)	References
Guided	1 (8)	[73]
Blended treatment	1 (8)	[81]
Compassionate elements		
Recognition	5 (38)	[72,77,78,81,84]
Universality	3 (23)	[73,74,76]
Empathy	11 (85)	[72,73,75-77,79-84]
Tolerance	5 (38)	[75,77,80,83,84]
Acting	5 (38)	[76-78,80]

[a]N/A: not applicable.

In Role B, we found 12 technologies that were used to enhance the elements of self-compassion in the person using them. An example is a VR intervention set in space to support the development of self-compassion [60]. The most frequently used technology types are VR [60,61,63,64], mobile phone apps, and SMS text messaging [65,66,68,69]. The goals mentioned most often for technologies with this role were to enhance self-compassion [60,62-64,66,68,70] and to decrease psychological symptoms [61,62,67,68]. Target groups were sometimes specific, for example, people with paranoia symptoms [60,61], and sometimes broader, for example, adolescents [65,66]. For Role B, most technologies served as a stand-alone intervention [60-68,70]. We found 1 study in which the intervention was guided in the form of written feedback from a personal web-based counselor [71]. Finally, in 1 study, the technology was part of a blended treatment [69]. Here, SMS text messaging was used to enhance a physical psychological intervention on self-compassion [69].

In Role C, we found 13 technologies that were used to facilitate elements of compassion between people, for instance, a therapeutic robot that acts as a social agent between pediatric patients and their parents in robot-assisted therapy [72]. The types of technologies were mainly web-based interventions [73,74,76-78], online support groups [74,76], videoconferencing [79,84], and VR [75,81]. Various goals were found, these included decreasing psychological symptoms [72,73,81,84] and enhancing empathy [75,80,82]. Target groups were often people with physical illnesses [72,74,76,77,84] but also included informal caregivers [75,84] and licensed care staff [80]. For Role C, again, most technologies formed a stand-alone intervention [72,74-76,78,82,83]. We also found several technologies that were used for telecommunication [77,79,80,84]. We found 1 study where the technological intervention was guided (albeit minimally), pertaining to the use of a “booster phone call” to review the process and challenges [73]. Finally, we also found 1 study in which technology was part of blended treatment: a VR intervention that was used in physical sessions with a counselor [81].

Study Designs and Outcome Measures

A range of different study designs and outcome measures was found in the included studies. Regarding outcome measures, none of the studies evaluated whether the technology as such was (perceived to be) compassionate. For the 8 articles categorized as Role A (technology showing compassion), we found the following study designs: experimental [52,57,59], quasi-experimental [55,58], qualitative [54], a combination of qualitative and experimental [53], and a combination of quantitative cross-sectional and qualitative [56]. Outcome measures included the acceptability of the intervention [52,56,57], evaluations of the technology [53,55,59], and symptoms and well-being [52,57,59]. One study investigated empathy in human-robot interaction by observing people with dementia interacting with a therapy robot during psychomotor therapy [59].

For the 12 articles categorized as Role B (enhancing self-compassion), the study designs were almost exclusively experimental [60-62,64,66-70], followed by qualitative [71] and quasi-experimental [63]. All but 1 study used specific outcome measures, most often symptoms and well-being [60-64,66-69] and self-compassion [60-63,66-68,70].

Finally, for the 13 articles categorized as Role C (facilitating compassion), the study designs were mostly qualitative [74,76,77,79,82] or experimental [72,78,80,83,84], followed by quasi-experimental [75], a combination of qualitative and quantitative cross-sectional [73], and a combination of qualitative and quasi-experimental [81]. Found outcome measures were mostly empathy between humans [72,75,76,78-80,83] and symptoms and well-being [72,76,81,84]. One study had compassion as an outcome measure, referring to nurses’ abilities to be compassionate through web-based communication [77].

Elements of Compassion

Overview

In general, empathy was used slightly more often as the main term to describe certain elements of technologies in the included studies (18/33, 55%) than compassion (15/33, 45%). Compassion was most often found in the form of increasing self-compassion (Role B). We did not find any technologies that embedded all the 5 elements of compassion. In the following sections, we will further specify how the 5 elements were embedded in the included DMHIs.

Recognition

The first element of compassion proposed by Strauss et al [28] is the recognition of suffering. This was observed in 13 of the

included studies. In Role A (technology showing compassion), recognition referred to users being able to indicate their thoughts and feelings toward the technology. For example, people could indicate their mood to a fully automated conversational agent and receive a weekly mood description [57]. Similarly, in Role B (enhancing self-compassion), technology could allow users to track their feelings daily in a mood cloud, prompting personalized feedback or supportive emotional regulation statements to enhance self-compassion [66]. In Role C (facilitating compassion), technology could be used to help caregivers recognize suffering. An example is a robot seal that helped parents to acknowledge their children's pain more accurately in robot-assisted therapy [72]. In another study, people with schizophrenia could use a web-based application to answer questions about their status and treatment, so the app could then act by giving recommendations on what they could discuss with their therapist and how [78]. Thus, unsuitable medications or other issues that could arise could be recognized faster.

Universality

This element refers to the understanding of the universality of suffering in human experiences [28]. It was only found in 10 of the included studies, of which none belonged to Role A, 7 studies were categorized as Role B, and 3 as Role C. In Role B, the technology could give shape to universality by giving user messages or exercises that emphasize common humanity, and that negative emotions are part of being human. For instance, in 1 study, women at risk for postpartum depressive symptoms received exercises in a web application to help them accept that they are susceptible and human, similar to all mothers [67]. Another study involved the chatbot Vincent, who "talked" about his mistakes, supporting common humanity [70]. In Role C, technology supported universality by providing a link to social support [73] so that people could exchange experiences and prevent isolation.

Empathy

This element refers to feeling empathy for those who are suffering and connecting with distress (emotional resonance [28]). Overall, the element "empathy" was the most frequently found (n=29). For Role A, empathy was found in all the studies (n=8). Most studies simply mentioned that the technology used an empathic tone in the messages it sent [52]; in some cases, it was tailored to the user's input (current mood or text). In Role B, the element of empathy was found only twice. It refers to technology fostering the user's empathy with the goal of targeting symptoms and well-being outcomes by letting them imagine an unpleasant experience they had from the perspective of an observer [62]. The other time it was found in a study where participants created a personal compassionate coach in VR that provided kindness and warmth, to help them be more self-compassionate [61]. In Role C, technology supports empathy among people in diverse ways and was found 11 times. Usually, technology facilitates empathy between a client and an informal or formal caregiver, for example, through a VR intervention that allows informal caregivers to experience dementia [75]. Technology can also facilitate empathy between a client and other people with similar experiences, for example, through an online support group [76]. Finally, technology could foster empathy in the general public for those with mental illnesses [83].

Tolerance

The element that was found least in the included studies was "tolerance" (n=8), referring to tolerating uncomfortable feelings aroused in response to the suffering person (such as distress, anger, or fear) and remaining open to and accepting of the person suffering [28]. Tolerance was not found in Role A. For Role B, tolerance was found 3 times. In all cases, it referred to supporting tolerance in humans and was described as helping people to be nonjudgmental and accepting toward themselves [68]. Finally, for Role C, tolerance was found the most frequently (n=5). In one instance, it was expressed in a VR intervention, giving informal caregivers of people with dementia more confidence in their care tasks and a more positive attitude toward them by increasing their resilience and proactive competence [75]. Similarly, a web-based training program for licensed care staff conveyed the philosophy that a resident's potentially problematic behavior is an expression of an unmet need, instead of just behavior that needs to be managed [80]. In another study, tolerance was mentioned more indirectly, stating that a web-based communication service allowed a nurse to respond by expressing acceptance and validation [77].

Acting

Finally, the element "acting" refers to the motivation to act or actually acting to alleviate suffering [28]. This element was identified in 20 of the included studies. In Role A, acting was found 6 times. It was often found in very practical ways, stimulating the user to complete activities [52]. In some cases, the prompted content or activities were offered by the technology based on the user's input, such as their mood state [57] or medication-taking behavior [58]. In one case, a social robot was used to provide positive feedback and motivate a person with dementia to engage in psychomotor therapy with their therapist [59]. For Role B, acting was found 9 times, where the technology acted to enhance the user's self-compassion. For instance, VR was used to allow people to interact with a space nebula representing compassion, guiding them to open up compassionate feelings to increase self-compassion and decrease paranoia [60]. In another case with VR, people delivered compassionate sentences to a virtual child and then experienced these again from the perspective of the child [64]. To facilitate compassion in Role C, acting was found 5 times and could take different shapes. Technology could act as a social agent to stimulate empathic social interactions with other people [72]. Other times, the element of acting referred to human action facilitated by technology, such as the possibility of requesting support from a nurse [77].

Discussion

Principal Findings

This systematic scoping review investigated how and to what extent technology for mental health care has been connected to compassion in previous research. We could identify 3 roles that technology can play to support compassion in mental health

care: showing compassion *to* the client, enhancing self-compassion *in* people, and facilitating compassion *between* people, such as between a client and a caretaker or therapist. The main types of technologies and their goals, as well as the main study designs and outcome measures, differed by role, whereas the main target groups varied widely across all roles. We found a large majority of technologies that were described as stand-alone interventions, to be used by the target group without guidance from a coach or therapist. Only a few included studies described technologies as part of blended treatment: integrated in face-to-face sessions with a coach or therapist. This is in line with earlier research showing a lack of clarity on how to embed technologies in blended treatment [19]. Furthermore, none of the technologies included all elements of a proposed definition of compassion [28], nor was any technology evaluated on whether it was experienced as compassionate. However, we recognized certain elements of compassion in the technologies, showing that all elements could potentially be embedded in technology.

Of the 5 proposed elements of compassion, we found descriptions matching the elements "empathy" and "acting" most often in the included studies, followed by "recognition." In the scoping review care by Kemp et al [37], DMHIs were examined using a model of digital intersections with compassionate care. The authors found DMHIs that could support 4 of the 6 categories in this model. However, for 2 categories, no DMHIs were found in their review. These were "awareness of suffering" (the use of a DMHI to become more aware of one's suffering) and "mediated response" (the use of a DMHI to mediate the response to suffering) [37]. "Awareness of suffering" seems to be related to "recognition of suffering" in our current review, whereas "mediated response" could have overlap with "acting to alleviate suffering." We found several examples of how DMHIs could support both compassion components.

The elements of "universality" and "tolerance" were rarely found "Universality" was mostly found in the studies in Role B, describing DMHIs with the aim to enhance self-compassion. These studies often followed the definition of self-compassion by Neff [29], which overlaps with the definition of compassion proposed by Strauss et al [28]. Furthermore, as mentioned before, none of the included studies described technologies that fostered all elements of compassion. As Sinclair et al [38] discussed in their review on compassion in health care, the separate elements of compassion are not inherently compassionate; rather, their combination constitutes compassion. Following this logic, combined with the lack of evaluation of technology on (perceived) compassion, we cannot be sure if truly compassionate technology exists within our scope.

The Added Value of Compassion

Empathy was the element of compassion that we identified most frequently and that was used more often as a main term to describe technologies than compassion itself. As discussed in the Introduction section, there are some important differences between the constructs [28,33]. In short, compassion is specifically a response to suffering, whereas empathy can be felt for any emotion. Further, compassion includes the motivation to act to alleviate suffering, whereas empathy can also be followed by behavior in general but does not necessarily include this. These differences, as well as the additional elements of understanding the universality of human suffering and distress tolerance, make compassion especially valuable in the field of mental health or mental health care [37]. Therefore, we believe that explicitly embedding elements of compassion in the design and use of technology for mental health care, not just empathy, is a promising approach that is currently lacking.

In the current review, the lesser-known elements of compassion that are just as important in mental health care, such as understanding the universality of suffering in human experience and tolerating uncomfortable feelings, were rarely found. This indicates that a too narrow (or no) definition of compassion is used, at least when referring to features of technology. The overarching construct of compassion in "traditional" mental health care is a fundamental value and has multiple beneficial effects [28,37]. Moreover, considering the intentions, motivations, and values of stakeholders in mental health care or health care is argued to be essential in successfully blending technology into this field [12,37,85]. Thus, considering the elements of compassion in the design and use of future digital interventions could be a promising approach to improving the acceptance of DMHIs in the mental health context.

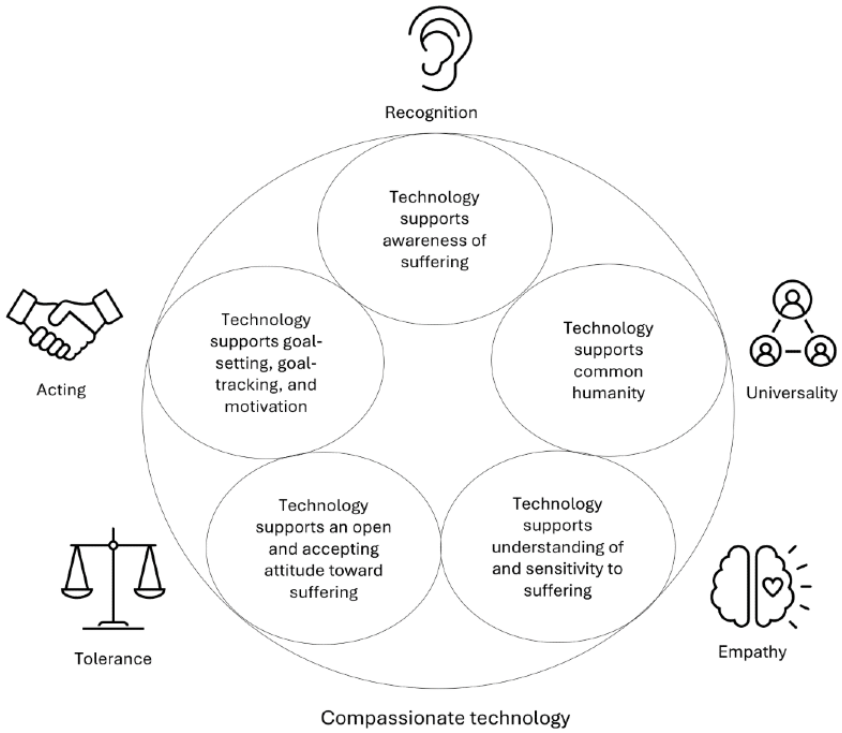
The Potential of Compassionate Technology

Our findings show that compassion offers a versatile and potentially transdiagnostic lens with which to examine technology for mental health care. First, as discussed in the Introduction, mental health professionals expressed the need for a clear conceptual basis for embedding technologies into mental health care [19]. Compassion could be a suitable basis and guiding force for the integration of technologies in mental health care: blending in technology in such a way that the therapeutic process as a whole is optimally in line with all elements of compassion. Thus, recognizing and alleviating suffering (ie, compassion) would be the central goal of both the design and evaluation of DMHIs as well as protocols for working with technology. Conventional values such as efficiency or effectiveness would be considered as means to achieve this goal. The design methodology of "Values that Matter" would lend itself very well to the design of DMHIs around compassion, as it aims to embed ethical values in technology [86].

Second, not only can compassion offer the basis for the design of DMHIs and protocols for the integration of compassionate technology in treatment, but the other way around technology can also bring new and additional ways to foster compassion that are not possible in traditional treatment (see Figure 3 for a conceptual overview). Examples include being present anywhere and anytime without getting tired or frustrated or being easier and more accessible to share suffering. On a microlevel, every interaction with a DMHI could convey all elements of compassion, and the elements of compassion could also be conveyed in the treatment as a whole by the therapist and technology together or at different times during the therapeutic process. Here, we illustrate what the latter could look like with an extended treatment scenario based on the 5 proposed elements

of compassion [28] and the elements of compassion found in the different technologies included in our review. We also referred to relevant related initiatives that did not fully meet our inclusion criteria and were therefore excluded but still provided interesting additional possibilities.

Figure 3. Conceptual overview of compassionate technology, which supports all elements of compassion.



First, the DMHI could help the client and therapist to (1) *recognize* moments of suffering anytime and anywhere. This could be based on input from the client (eg, indicating their mood and feelings [57]) but could also be combined with physiological measurements from a wearable (an example can be found in Fletcher et al [87]). The DMHI could convey the (2) *universality* of suffering in the human experience, for example, by connecting the client to similar experiences of others [73]. Furthermore, the DMHI could support the presence of (3) *empathy* for clients' suffering by sending empathic messages based on the user's input [52] or by helping the therapist to offer emotional support specifically during moments of potential distress [77].

Furthermore, the tone in communication with the DMHI and the therapist would remain open, accepting, and (4) *tolerant* of distress by acknowledging and validating difficult feelings (eg, [67,88]), where the DMHI has the advantage that it is not susceptible to empathic distress and can always be there for the client when it is needed. DMHIs could also be used to help increase the resilience and distress tolerance of mental health professionals [80]. Finally, together with the therapist, the client could look back on the recorded and experienced difficult moments and successes and set therapeutic goals. The DMHI could help to keep track of and work on these goals [57], until the next appointment with the therapist, for instance, by activating tailored exercises at the right moment and giving reminders or rewards to support (5) *acting* to alleviate suffering. Of course, this example raises new questions that need to be explored, such as how the tasks of the therapist and technology can be optimally blended to offer all the elements of recognizing and relieving suffering in an iterative manner.

Critical Notes on Digital Environments in Relation to Compassion and Empathy

This review has shown the possibilities for DMHIs to contribute to compassion in the field of mental health in several ways. However, it is important to acknowledge that concerns also exist regarding the use of technology in (mental) health care. First, there is concern about technology detracting from empathy and compassion. This could happen because of a lack of emotional signals and cues, and the possibility of hiding behind anonymity, and easily escaping the reactions of others [37,89]. In our review, we saw this in an article describing ambiguous experiences with videoconferencing, which could be considered impersonal [79]. These concerns point to the importance of placing more emphasis on empathy and compassion if DMHIs are used, such as by preparing (mental) health professionals to use DMHIs in a compassionate way for both themselves and their clients [89]. One possible way to enhance empathy among health professionals could be VR interventions [90,91]. In addition, there is debate about whether DMHIs actually contribute to equal access to care or whether they enlarge the existing inequities in society. For example, groups that have limited access to digital health care and limited digital literacy skills could be left behind [92]. These concerns should be considered in the development, evaluation, and use of DMHIs in mental health care. For instance, future evaluation of DMHIs should critically assess for the presence and quality of different compassion elements.

Limitations

This review is a systematic scoping review into an unexplored and diverse field. As we searched for mentions of compassion or empathy to delimit our research area, there could be technologies for mental health care that we did not include here but do show elements of compassion. There are several related initiatives in technology development, such as calm technology and affective computing, which focus on unobtrusive and emotion-sensitive technologies, respectively [93,94]. However, because this scoping review aims to provide an overview of technology explicitly related to compassion or empathy, these related initiatives did not fall within our scope.

Furthermore, it is important to keep in mind that our conceptualization of compassion, and especially the acceptability and value of suffering within Western psychological theories and therapies, may have cross-cultural limitations [95,96]. As the majority of studies included in the current review were conducted in Western countries, our findings may represent a mostly Western view of compassion and suffering. For instance, non-Western cultures generally do not see suffering as a purely negative life experience but instead believe that by appreciating it for what it offers, it can actually contribute to living a good life [95].

In addition, because the term "compassionate technology" was not found in previous studies, we used the definition of compassion proposed by Strauss et al [28] to assess the extent to which the technology in the included studies matched elements of compassion. In doing so, we closely followed the phrasing of the study authors. However, authors often did not explain (in detail) what they understood as "empathic" or "compassionate." Therefore, we may have misinterpreted the meaning of the authors. To make future studies on compassionate technology more transparent and comparable, we recommend that the authors include their definitions of compassion or its elements.

Because the present review was a scoping review, we did not assess the methodological quality of the included articles but instead focused on creating an overview of the scope of this field. Moreover, although some studies have measured compassion or self-compassion as an outcome, no studies have measured how compassionate the technology was, for example, as experienced by its users. We do not know whether the compassionate elements actually contributed to the presence of compassion, and if so, how and which elements did. Thus, we cannot be sure that the technologies we included measurably showed, enhanced, or facilitated compassion in the mental health care process.

Future Research

On the basis of the increasing frequency we saw over the publication years of the included articles linking compassion and empathy to DMHIs, the attention for such values in this field seems to be growing. This makes it a promising area for further research, but also one where much remains to be discovered. Most importantly, compassion is not yet seen as a foundation and goal for embedding technology into mental health care, and research is needed to learn more about how to design technologies and blended ways of working around compassion, focusing on the optimal recognition and alleviation of suffering. Furthermore, no scale exists to evaluate DMHIs on compassion. To advance the field of compassionate technology, it is essential to be able to determine which types of technologies with which design features actually support the presence of compassion across the 3 roles and for whom. Thus, a scale needs to be developed to measure compassion as shown, enhanced or facilitated by technology for both clients and professionals in mental health care.

Finally, as a scoping review, this review focused on the scope and degree to which we could find compassion linked to DMHIs but did not consider the effects of technology with compassionate elements on, for example, adherence [97], engagement [98], or effectiveness. Future research is needed to study which validated measures have been used in this field, so that the effects on the aforementioned constructs could be researched. Although the types of research and outcome measures we found varied widely, research on technology enhancing self-compassion *in* people consisted mainly of experimental studies measuring participants' psychological symptoms or self-compassion or both. Therefore, research in this role would lend itself well to a meta-analysis of the effectiveness of these technologies. Moreover, we used search terms related to mental health but found many studies related to mental well-being, with some components of mental health. Thus, it could be interesting for future research to expand the scope to explicitly include DMHIs around mental well-being or lifestyle, because there might be additional relevant work in these fields as well.

Conclusions

Compassion is an essential value in mental health care, pertaining to recognizing suffering, being moved by it, and acting to alleviate it. Given the importance and benefits of compassion in mental health care, shifting the focus of the design, evaluation, and use of DMHIs to center on compassion seems to be a new, fascinating, and perhaps even necessary direction in research and clinical practice. This scoping review explored how and to what degree elements of compassion have been linked to technologies for mental health care in previous studies. Our review shows that compassion is a widely applicable construct across different technologies, target groups, and for different aims in mental health care and is potentially a guiding force in embedding technology in mental health care. Moreover, it provides new input for the design and development of technology around compassion and demonstrates the necessity of evaluating technology on this foundational value in mental health care. Overall, this review serves as a first step toward "compassionate technology" as a guiding principle in the use and design of technology for mental health care. This principle refers to technology that contributes to the recognition and alleviation of suffering and is appropriately suited to the mental health care context for both clients and professionals.

Acknowledgments

This publication is part of the project “designing compassionate technology with high societal readiness levels for mental health care” (project number 403.19.229) of the research program Transitions and Behavior, financed by the Dutch Research Council (NWO), Minddistrict BV, and Dimence Groep. Funding sources were not involved in the writing of this manuscript or in the decision to submit it for publication.

Authors' Contributions

CvL, GJW, GDSL, SMK, and MLN conceptualized the study and wrote the original draft of the manuscript. CvL and BL screened the literature. CvL extracted data from the included articles and BL validated the data extraction. CvL, BL, and MLN contributed to data interpretation. All authors edited multiple drafts and reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies for each of the included databases.
[DOCX File , 14 KB - mental_v10i1e42403_app1.docx]

References

1. Tal A, Torous J. The digital mental health revolution: opportunities and risks. Psychiatr Rehabil J 2017 Sep;40(3):263-265. [doi: 10.1037/prj0000285] [Medline: 28891658]
2. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
3. Aboujaoude E, Gega L, Parish MB, Hilty DM. Editorial: digital interventions in mental health: current status and future directions. Front Psychiatry 2020 Feb 27;11:111 [FREE Full text] [doi: 10.3389/fpsyt.2020.00111] [Medline: 32174858]
4. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]
5. Rabbitt SM, Kazdin AE, Hong JH. Acceptability of robot-assisted therapy for disruptive behavior problems in children. Arch Sci Psychol 2015 Aug 03;3(1):101-110 [FREE Full text] [doi: 10.1037/arc0000017]
6. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2017 Jan 10;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
7. Mandryk RL, Birk MV. Toward game-based digital mental health interventions: player habits and preferences. J Med Internet Res 2017 Apr 20;19(4):e128 [FREE Full text] [doi: 10.2196/jmir.6906] [Medline: 28428161]
8. Bergmo TS. How to measure costs and benefits of eHealth interventions: an overview of methods and frameworks. J Med Internet Res 2015 Nov 09;17(11):e254 [FREE Full text] [doi: 10.2196/jmir.4521] [Medline: 26552360]
9. Doherty G, Coyle D, Matthews M. Design and evaluation guidelines for mental health technologies. Interact Comput 2010 Jul;22(4):243-252. [doi: 10.1016/j.intcom.2010.02.006]
10. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 01;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
11. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
12. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJ. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14:355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
13. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
14. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
15. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
16. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]

17. Andersson G, Topooco N, Havik O, Nordgreen T. Internet-supported versus face-to-face cognitive behavior therapy for depression. Expert Rev Neurother 2016;16(1):55-60. [doi: 10.1586/14737175.2015.1125783] [Medline: 26610160]
18. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of eMental health by psychologists: the construction of the levels of adoption of eMental health model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
19. Titzler I, Saruhanjan K, Berking M, Riper H, Ebert DD. Barriers and facilitators for the implementation of blended psychotherapy for depression: a qualitative pilot study of therapists' perspective. Internet Interv 2018 Jun;12:150-164 [FREE Full text] [doi: 10.1016/j.invent.2018.01.002] [Medline: 30135779]
20. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
21. Sharry J, Davidson R, McLoughlin O, Doherty G. A service-based evaluation of a therapist-supported online cognitive behavioral therapy program for depression. J Med Internet Res 2013 Jun 27;15(6):e121 [FREE Full text] [doi: 10.2196/jmir.2248] [Medline: 23807565]
22. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
23. Simpson JA, Weiner ES. Oxford English Dictionary. 2nd edition. Oxford, UK: Clarendon Press; 1989.
24. Shea S, Lionis C. The call for compassion in health care. In: Seppälä EM, Simon-Thomas E, Brown SL, Worline MC, Cameron CD, Doty JD, editors. The Oxford Handbook of Compassion Science. Oxford, UK: Oxford University Press; Sep 28, 2017:457-474.
25. Gilbert P. Compassion and cruelty: a biopsychosocial approach. In: Gilbert P, editor. Compassion: Conceptualisations, Research and Use in Psychotherapy. London, UK: Routledge; 2005:9-74.
26. Gilbert P. Introducing compassion-focused therapy. Adv Psychiatr Treat 2009;15(3):199-208 [FREE Full text] [doi: 10.1192/apt.bp.107.005264]
27. Feldman C, Kuyken W. Compassion in the landscape of suffering. Contemp Buddhism 2011 Jun 14;12(1):143-155 [FREE Full text] [doi: 10.1080/14639947.2011.564831]
28. Strauss C, Lever Taylor B, Gu J, Kuyken W, Baer R, Jones F, et al. What is compassion and how can we measure it? A review of definitions and measures. Clin Psychol Rev 2016 Jul;47:15-27 [FREE Full text] [doi: 10.1016/j.cpr.2016.05.004] [Medline: 27267346]
29. Neff KD. The science of self-compassion. In: Germer C, Siegel R, editors. Compassion and Wisdom in Psychotherapy. New York, NY, USA: Guilford Press; 2012:79-92.
30. Elices M, Carmona C, Pascual JC, Feliu-Soler A, Martin-Blanco A, Soler J. Compassion and self-compassion: construct and measurement. Mindfulness Compassion 2017 Jan;2(1):34-40 [FREE Full text] [doi: 10.1016/j.mincom.2016.11.003]
31. Lazarus RS. Emotion and Adaptation. Oxford, UK: Oxford University Press; 1991.
32. Davis MH. Measuring individual differences in empathy: evidence for a multidimensional approach. J Pers Soc Psychol 1983;44(1):113-126 [FREE Full text] [doi: 10.1037/0022-3514.44.1.113]
33. Cuff BM, Brown SJ, Taylor L, Howat DJ. Empathy: a review of the concept. Emot Rev 2016 Apr;8(2):144-153 [FREE Full text] [doi: 10.1177/1754073914558466]
34. Batson CD. Altruism in Humans. New York, NY, USA: Oxford University Press; Jul 2011.
35. Eisenberg N, Eggum ND. Empathic responding: sympathy and personal distress. In: Decety J, Ickes W, editors. The Social Neuroscience of Empathy. Cambridge, MA, USA: MIT Press; Mar 2009:71-83.
36. Goetz JL, Simon-Thomas E. The landscape of compassion: definitions and scientific approaches. In: Seppälä EM, Brown SL, Worline MC, Cameron CD, Doty JR, editors. The Oxford Handbook of Compassion Science. Oxford, UK: Oxford University Press; 2017:3-15.
37. Kemp J, Zhang T, Inglis F, Wiljer D, Sockalingam S, Crawford A, et al. Delivery of compassionate mental health care in a digital technology-driven age: scoping review. J Med Internet Res 2020 Mar 06;22(3):e16263 [FREE Full text] [doi: 10.2196/16263] [Medline: 32141833]
38. Sinclair S, Norris JM, McConnell SJ, Chochinov HM, Hack TF, Hagen NA, et al. Compassion: a scoping review of the healthcare literature. BMC Palliat Care 2016 Jan 19;15:6 [FREE Full text] [doi: 10.1186/s12904-016-0080-0] [Medline: 26786417]
39. Day J, Finkelstein JC, Field BA, Matthews B, Kirby JN, Doty JR. Compassion-focused technologies: reflections and future directions. Front Psychol 2021 May 13;12:603618 [FREE Full text] [doi: 10.3389/fpsyg.2021.603618] [Medline: 34054636]
40. Wiljer D, Strudwick G, Crawford A. Caring in a digital age: exploring the interface of humans and machines in the provision of compassionate healthcare. In: Hodges BD, Paech G, Bennett J, editors. Without Compassion, There Is No Healthcare: Leading with Care in a Technological Age. Montreal, Quebec: McGill-Queen's University Press; Nov 2020:33-41.
41. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]

42. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
43. Horsley T. Tips for improving the writing and reporting quality of systematic, scoping, and narrative reviews. J Contin Educ Health Prof 2019;39(1):54-57. [doi: 10.1097/CEH.0000000000000241] [Medline: 30789378]
44. Gilbert P. Compassion Focused Therapy: Distinctive Features. Oxfordshire, UK: Routledge; 2010.
45. Leaviss J, Uttley L. Psychotherapeutic benefits of compassion-focused therapy: an early systematic review. Psychol Med 2015 Apr;45(5):927-945 [FREE Full text] [doi: 10.1017/S0033291714002141] [Medline: 25215860]
46. Covidence systematic review software. Covidence. 2020. URL: https://get.covidence.org/literature-review?campaignid=18238395256&adgroupid=138114520982&gclid=EAIaIQobChMI6_7lgZ30_QIVe5lmAh0qpg22EAAYAiAAEgKUcvD_BwE [accessed 2020-09-15]
47. Study designs. Centre for Evidence-Based Medicine. Oxford, UK: University of Oxford URL: https://www.cebm.ox.ac.uk/resources/ebm-tools/study-designs [accessed 2021-09-15]
48. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006 Jun 01;27(2):237-246 [FREE Full text] [doi: 10.1177/1098214005283748]
49. Coghlan A, Filo K. Using constant comparison method and qualitative data to understand participants' experiences at the nexus of tourism, sport and charity events. Tour Manag 2013 Apr;35:122-131 [FREE Full text] [doi: 10.1016/j.tourman.2012.06.007]
50. Levitt HM, Bamberg M, Creswell JW, Frost DM, Josselson R, Suárez-Orozco C. Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: the APA publications and communications board task force report. Am Psychol 2018 Jan;73(1):26-46 [FREE Full text] [doi: 10.1037/amp0000151] [Medline: 29345485]
51. Morrow SL. Quality and trustworthiness in qualitative research in counseling psychology. J Couns Psychol 2005;52(2):250-260 [FREE Full text] [doi: 10.1037/0022-0167.52.2.250]
52. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
53. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
54. Javed H, Park CH. Interactions with an empathetic agent: regulating emotions and improving engagement in autism. IEEE Robot Autom Mag 2019 Jun;26(2):40-48 [FREE Full text] [doi: 10.1109/mra.2019.2904638] [Medline: 34887653]
55. Pontier M, Siddiqui GF. A virtual therapist that responds empathically to your answers. In: Proceedings of the 8th International Conference on Intelligent Virtual Agents. 2008 Sep Presented at: IVA '08; September 1-3, 2008; Tokyo, Japan p. 417-425. [doi: 10.1007/978-3-540-85483-8_42]
56. Kohori T, Hirayama S, Hara T, Muramatsu M, Naganuma H, Yamano M, et al. Development and evaluation of an interactive therapy robot. In: Proceedings of the 14th International Conference on Advances in Computer Entertainment Technology. 2017 Presented at: ACE' 17; December 14-16, 2017; London, UK p. 66-83. [doi: 10.1007/978-3-319-76270-8_6]
57. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
58. Bickmore TW, Puskar K, Schlenk EA, Pfeifer LM, Sereika SM. Maintaining reality: relational agents for antipsychotic medication adherence. Interact Comput 2010 Jul;22(4):276-288 [FREE Full text] [doi: 10.1016/j.intcom.2010.02.001]
59. Rouaix N, Retru-Chavastel L, Rigaud AS, Monnet C, Lenoir H, Pino M. Affective and engagement issues in the conception and assessment of a robot-assisted psychomotor therapy for persons with dementia. Front Psychol 2017 Jun 30;8:950 [FREE Full text] [doi: 10.3389/fpsyg.2017.00950] [Medline: 28713296]
60. Ascone L, Ney K, Mostajeran F, Steinicke F, Moritz S, Gallinat J, et al. Virtual reality for individuals with occasional paranoid thoughts. In: Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Apr Presented at: CHI EA '20; April 25-30, 2020; Honolulu, HI, USA p. 1-8. [doi: 10.1145/3334480.3382918]
61. Brown P, Waite F, Rovira A, Nickless A, Freeman D. Virtual reality clinical-experimental tests of compassion treatment techniques to reduce paranoia. Sci Rep 2020 May 22;10(1):8547 [FREE Full text] [doi: 10.1038/s41598-020-64957-7] [Medline: 32444619]
62. Pennesi JL, Wade TD. Imagery rescripting and cognitive dissonance: a randomized controlled trial of two brief online interventions for women at risk of developing an eating disorder. Int J Eat Disord 2018 May;51(5):439-448. [doi: 10.1002/eat.22849] [Medline: 29500828]
63. Falconer CJ, Rovira A, King JA, Gilbert P, Antley A, Fearon P, et al. Embodying self-compassion within virtual reality and its effects on patients with depression. BJPsych Open 2016 Jan;2(1):74-80 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002147] [Medline: 27703757]

64. Falconer CJ, Slater M, Rovira A, King JA, Gilbert P, Antley A, et al. Embodying compassion: a virtual reality paradigm for overcoming excessive self-criticism. PLoS One 2014 Nov 12;9(11):e111933 [FREE Full text] [doi: 10.1371/journal.pone.0111933] [Medline: 25389766]
65. Donovan E, Rodgers RF, Cousineau TM, McGowan KM, Luk S, Yates K, et al. Brief report: feasibility of a mindfulness and self-compassion based mobile intervention for adolescents. J Adolesc 2016 Dec;53:217-221. [doi: 10.1016/j.adolescence.2016.09.009] [Medline: 27814499]
66. Rodgers RF, Donovan E, Cousineau T, Yates K, McGowan K, Cook E, et al. BodiMojo: efficacy of a mobile-based intervention in improving body image and self-compassion among adolescents. J Youth Adolesc 2018 Jul;47(7):1363-1372. [doi: 10.1007/s10964-017-0804-3] [Medline: 29349593]
67. Fonseca A, Monteiro F, Alves S, Gorayeb R, Canavarro MC. Be a mom, a web-based intervention to prevent postpartum depression: the enhancement of self-regulatory skills and its association with postpartum depressive symptoms. Front Psychol 2019 Feb 18;10:265 [FREE Full text] [doi: 10.3389/fpsyg.2019.00265] [Medline: 30873060]
68. Schnepper R, Reichenberger J, Blechert J. Being my own companion in times of social isolation - a 14-day mobile self-compassion intervention improves stress levels and eating behavior. Front Psychol 2020 Oct 09;11:595806 [FREE Full text] [doi: 10.3389/fpsyg.2020.595806] [Medline: 33162922]
69. Raymond RR. Experience sampling assessment of a brief text-message enhanced self-compassion intervention. The Catholic University of America. 2019. URL: https://cuislandora.wrlc.org/islandora/object/cuislandora%3A213710 [accessed 2020-11-30]
70. Lee M, Ackermans S, Van As N, Chang H, Lucas E, IJsselsteijn WA. Caring for vincent: a chatbot for self-compassion. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 May Presented at: CHI '19; May 4-9, 2019; Glasgow, UK p. 702. [doi: 10.1145/3290605.3300932]
71. Köhle N, Drossaert CH, Jaran J, Schreurs KM, Verdonck-de Leeuw IM, Bohlmeijer ET. User-experiences with a web-based self-help intervention for partners of cancer patients based on acceptance and commitment therapy and self-compassion: a qualitative study. BMC Public Health 2017 Feb 28;17(1):225 [FREE Full text] [doi: 10.1186/s12889-017-4121-2] [Medline: 28245794]
72. Okita SY. Self-other's perspective taking: the use of therapeutic robot companions as social agents for reducing pain and anxiety in pediatric patients. Cyberpsychol Behav Soc Netw 2013 Jun;16(6):436-441. [doi: 10.1089/cyber.2012.0513] [Medline: 23505968]
73. Choo E, Guthrie KM, Mello M, Wetle TF, Ranney M, Tapé C, et al. "I need to hear from women who have 'been there'": developing a woman-focused intervention for drug use and partner violence in the emergency department. Partner Abuse 2016 Apr;7(2):193-220 [FREE Full text] [doi: 10.1891/1946-6560.7.2.193] [Medline: 27695604]
74. Bar-Lev S. "We are here to give you emotional support": performing emotions in an online HIV/AIDS support group. Qual Health Res 2008 Apr;18(4):509-521. [doi: 10.1177/1049732307311680] [Medline: 18192435]
75. Wijma EM, Veerbeek MA, Prins M, Pot AM, Willemse BM. A virtual reality intervention to improve the understanding and empathy for people with dementia in informal caregivers: results of a pilot study. Aging Ment Health 2018 Sep;22(9):1115-1123. [doi: 10.1080/13607863.2017.1348470] [Medline: 28691861]
76. Han JY, Shah DV, Kim E, Namkoong K, Lee SY, Moon TJ, et al. Empathic exchanges in online cancer support groups: distinguishing message expression and reception effects. Health Commun 2011 Mar;26(2):185-197 [FREE Full text] [doi: 10.1080/10410236.2010.544283] [Medline: 21318917]
77. Högberg KM, Sandman L, Nyström M, Stockelberg D, Broström A. Caring through web-based communication: a qualitative evaluation of a nursing intervention to create holistic well-being among patients with hematological disease. J Holist Nurs 2018 Sep;36(3):218-227. [doi: 10.1177/0898010116667343] [Medline: 27659277]
78. Steinwachs DM, Roter DL, Skinner EA, Lehman AF, Fahey M, Cullen B, et al. A web-based program to empower patients who have schizophrenia to discuss quality of care with mental health providers. Psychiatr Serv 2011 Nov;62(11):1296-1302 [FREE Full text] [doi: 10.1176/ps.62.11.pss6211_1296] [Medline: 22211208]
79. Kysely A, Bishop B, Kane R, Cheng M, De Palma M, Rooney R. Expectations and experiences of couples receiving therapy through videoconferencing: a qualitative study. Front Psychol 2019 Jan 21;10:2992 [FREE Full text] [doi: 10.3389/fpsyg.2019.02992] [Medline: 32038380]
80. Blair Irvine A, Billow MB, Eberhage MG, Seeley JR, McMahon E, Bourgeois M. Mental illness training for licensed staff in long-term care. Issues Ment Health Nurs 2012 Mar;33(3):181-194 [FREE Full text] [doi: 10.3109/01612840.2011.639482] [Medline: 22364430]
81. van Rijn B, Cooper M, Jackson A, Wild C. Avatar-based therapy within prison settings: pilot evaluation. Br J Guid Counc 2017;45(3):268-283 [FREE Full text] [doi: 10.1080/03069885.2015.1068273]
82. Fordham J, Ball C. Framing mental health within digital games: an exploratory case study of hellblade. JMIR Ment Health 2019 Apr 18;6(4):e12432 [FREE Full text] [doi: 10.2196/12432] [Medline: 30998224]
83. Tippin GK, Maranzan KA. Efficacy of a photovoice‐based video as an online mental illness anti‐stigma intervention and the role of empathy in audience response: a randomized controlled trial. J Appl Soc Psychol 2019 Jun;49(6):381-394 [FREE Full text] [doi: 10.1111/jasp.12590]

84. Milbury K, Weathers SP, Durrani S, Li Y, Whisenant M, Li J, et al. Online couple-based meditation intervention for patients with primary or metastatic brain tumors and their partners: results of a pilot randomized controlled trial. J Pain Symptom Manage 2020 Jun;59(6):1260-1267. [doi: 10.1016/j.jpainsymman.2020.02.004] [Medline: 32061834]
85. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022 Oct 05;24(10):e37341 [FREE Full text] [doi: 10.2196/37341] [Medline: 36197718]
86. Smits M, Ludden G, Peters R, Bredie SJ, van Goor H, Verbeek PP. Values that matter: a new method to design and assess moral mediation of technology. Des Issues 2022;38(1):39-54 [FREE Full text] [doi: 10.1162/desi_a_00669]
87. Fletcher RR, Tam S, Omojola O, Redemske R, Kwan J. Wearable sensor platform and mobile application for use in cognitive behavioral therapy for drug addiction and PTSD. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:1802-1805. [doi: 10.1109/IEMBS.2011.6090513] [Medline: 22254678]
88. Wray T, Kahler CW, Simpanen EM, Operario D. Game plan: development of a web app designed to help men who have sex with men reduce their HIV risk and alcohol use. JMIR Form Res 2018 Aug 23;2(2):e10125 [FREE Full text] [doi: 10.2196/10125] [Medline: 30684415]
89. Terry C, Cain J. The emerging issue of digital empathy. Am J Pharm Educ 2016 May 25;80(4):58 [FREE Full text] [doi: 10.5688/ajpe80458] [Medline: 27293225]
90. Louie AK, Coverdale JH, Balon R, Beresin EV, Brenner AM, Guerrero AP, et al. Enhancing empathy: a role for virtual reality? Acad Psychiatry 2018 Dec;42(6):747-752. [doi: 10.1007/s40596-018-0995-2] [Medline: 30324398]
91. Elzie CA, Shaia J. Virtually walking in a patient's shoes-the path to empathy? Med Sci Educ 2020 Dec;30(4):1737-1739 [FREE Full text] [doi: 10.1007/s40670-020-01101-0] [Medline: 34457839]
92. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
93. Case A. Calm Technology: Principles and Patterns for Non-Intrusive Design. New York, NY, USA: O'Reilly; Dec 2015.
94. Picard RW. Affective Computing. Cambridge, MA, USA: MIT Press; 1997.
95. Joshanloo M. Eastern conceptualizations of happiness: fundamental differences with western views. J Happiness Stud 2014;15(2):475-493 [FREE Full text] [doi: 10.1007/s10902-013-9431-1]
96. Ji LJ, Khei M, Yap S, Wang X, Zhang Z, Hou Y. Cultural differences in the construal of suffering and the COVID-19 pandemic. Soc Psychol Personal Sci 2021 Aug;12(6):1039-1047 [FREE Full text] [doi: 10.1177/1948550620958807]
97. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
98. Kelders SM, van Zyl LE, Ludden GD. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020 May 27;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]

Abbreviations

CBT: cognitive behavioral therapy
DMHI: digital mental health intervention
VR: virtual reality

Edited by J Torous, G Eysenbach; submitted 08.09.22; peer-reviewed by M Feijt, R Charow; comments to author 18.11.22; revised version received 16.12.22; accepted 25.01.23; published 07.04.23.

Please cite as:
van Lotringen C, Lusi B, Westerhof GJ, Ludden GDS, Kip H, Kelders SM, Noordzij ML
The Role of Compassionate Technology in Blended and Digital Mental Health Interventions: Systematic Scoping Review
JMIR Ment Health 2023;10:e42403
URL: https://mental.jmir.org/2023/1/e42403
doi:10.2196/42403
PMID:37027207

Review

Adverse Effects of Virtual and Augmented Reality Interventions in Psychiatry: Systematic Review

Robert M Lundin[1,2,3], BSc, MBBCh; Yuhern Yeap[4], MBChB; David B Menkes[3], MD, PhD

[1]Change to Improve Mental Health, Mental Health Drugs and Alcohol Services, Barwon Health, Geelong, Australia

[2]Institute for Mental and Physical Health and Clinical Translation, Deakin University, Geelong, Australia

[3]Waikato Clinical Campus, University of Auckland, Hamilton, New Zealand

[4]Mental Health and Addictions, Waikato District Health Board, Hamilton, New Zealand

Corresponding Author:
Robert M Lundin, BSc, MBBCh
Change to Improve Mental Health
Mental Health Drugs and Alcohol Services
Barwon Health
Level 1/126 Little Malop St
Geelong, 3220
Australia
Phone: 61 34215000
Email: robert@lundin.no

Abstract

Background: Virtual reality (VR) and augmented reality (AR) are emerging treatment modalities in psychiatry, which are capable of producing clinical outcomes broadly comparable to those achieved with standard psychotherapies.

Objective: Because the side effect profile associated with the clinical use of VR and AR remains largely unknown, we systematically reviewed available evidence of their adverse effects.

Methods: A systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework across 3 mental health databases (PubMed, PsycINFO, and Embase) to identify VR and AR interventions targeting mental health diagnoses.

Results: Of 73 studies meeting the inclusion criteria, 7 reported worsening clinical symptoms or an increased fall risk. Another 21 studies reported "no adverse effects" but failed to identify obvious adverse effects, mainly cybersickness, documented in their results. More concerningly, 45 of the 73 studies made no mention of adverse effects whatsoever.

Conclusions: An appropriate screening tool would help ensure that VR adverse effects are correctly identified and reported.

(JMIR Ment Health 2023;10:e43240) doi:10.2196/43240

KEYWORDS

virtual reality; augmented reality; mental health; side effects; adverse events; hardware; VR; software; AR; cybersickness; reporting standards

Introduction

Overview

Substantial unmet need for treatment for mental disorders arises from problems of access, efficacy, and tolerability of conventional treatments [1,2]. These limitations have prompted the development and evaluation of novel interventions, including those based on virtual reality (VR) and augmented reality (AR) [3,4]. Recent systematic reviews and meta-analyses have shown that VR and AR can be usefully applied to the treatment of various psychological disorders [5-7], mainly using versions of cognitive behavioral therapy (CBT) and the specific CBT technique of exposure therapy (ET) [8]. Despite the enhanced accessibility and scalability of digital interventions, individual VR studies are often criticized for their limited quality and sample size [9].

Despite its therapeutic promise, the stance of VR-based psychotherapy cannot be established without a realistic appraisal of its benefits and harms. Unfortunately, little evidence is available regarding VR side effects experienced by mental health cohorts, a deficit shared with psychotherapies in general [10-12]; both lack comprehensive identification of adverse effects,

including symptom deterioration [13]. There is an urgent need to improve detection, reporting, and evaluation of adverse effects of psychotherapies, including VR [14].

Adverse Effects in VR and AR Clinical Trials

We define adverse effects as those perceived as unpleasant or harmful by the patient, clinician, or family, including symptomatic deterioration [11], addiction, or delusions manifesting in relation to VR use [10]. Previous studies refer to VR-induced symptoms and effects involving nausea, dizziness, disorientation, postural instability, and fatigue. These unpleasant physical symptoms experienced are often referred to as "cybersickness," which is found to be negatively correlated with the subjective experience of "presence" in the virtual environment [15]. VR interventions can also induce dissociative symptoms [16], including perceived disconnection from the self (depersonalization) and the environment (derealization).

The full mechanism for the above experiences is not known, but they are thought to be triggered by the discontinuity between digital and objective reality. While the frequency and severity of these are being reduced by technological improvements in hardware and software, as well as a reduction in the time spent in VR, the effects are often subjective experiences that cannot be mitigated in some individuals and vary greatly based on the hardware and software used [17]. Relevant hypotheses primarily involve the mismatch between the 2 sensory systems involved in motion detection. If a user is standing still, no motion is detected by the vestibular system; however, when using a VR device, there is a direct contradiction owing to the motion being observed by the visual system [18]. Measuring and addressing such adverse effects is important because they predict poorer treatment outcomes and increased nausea [19-21]. ET, whether in vivo or delivered via VR, can transiently intensify distress [22] while improving outcomes overall [23].

This has led to the development of rating scales for cybersickness, including the Simulator Sickness Questionnaire (SSQ), Visual Analogue Scale, the Fast Motion Sickness Scale, and Virtual Reality Sickness Questionnaire [24-26]. It has also led to the repurposing of previously reported scales such as Subjective Units of Distress Scale [27] and, more recently, the development of the Virtual Reality Neuroscience Questionnaire [28], which additionally measures user experience with the software, allowing direct comparisons between different interventions. There are also tools specifically scoring the degree of "presence" such as the Igroup Presence Questionnaire [22]. This is important because the intervention's ability to influence the emotional state, fundamental to therapeutic effects, is related to the sense of presence and can be disrupted by adverse effects interrupting the simulation [25].

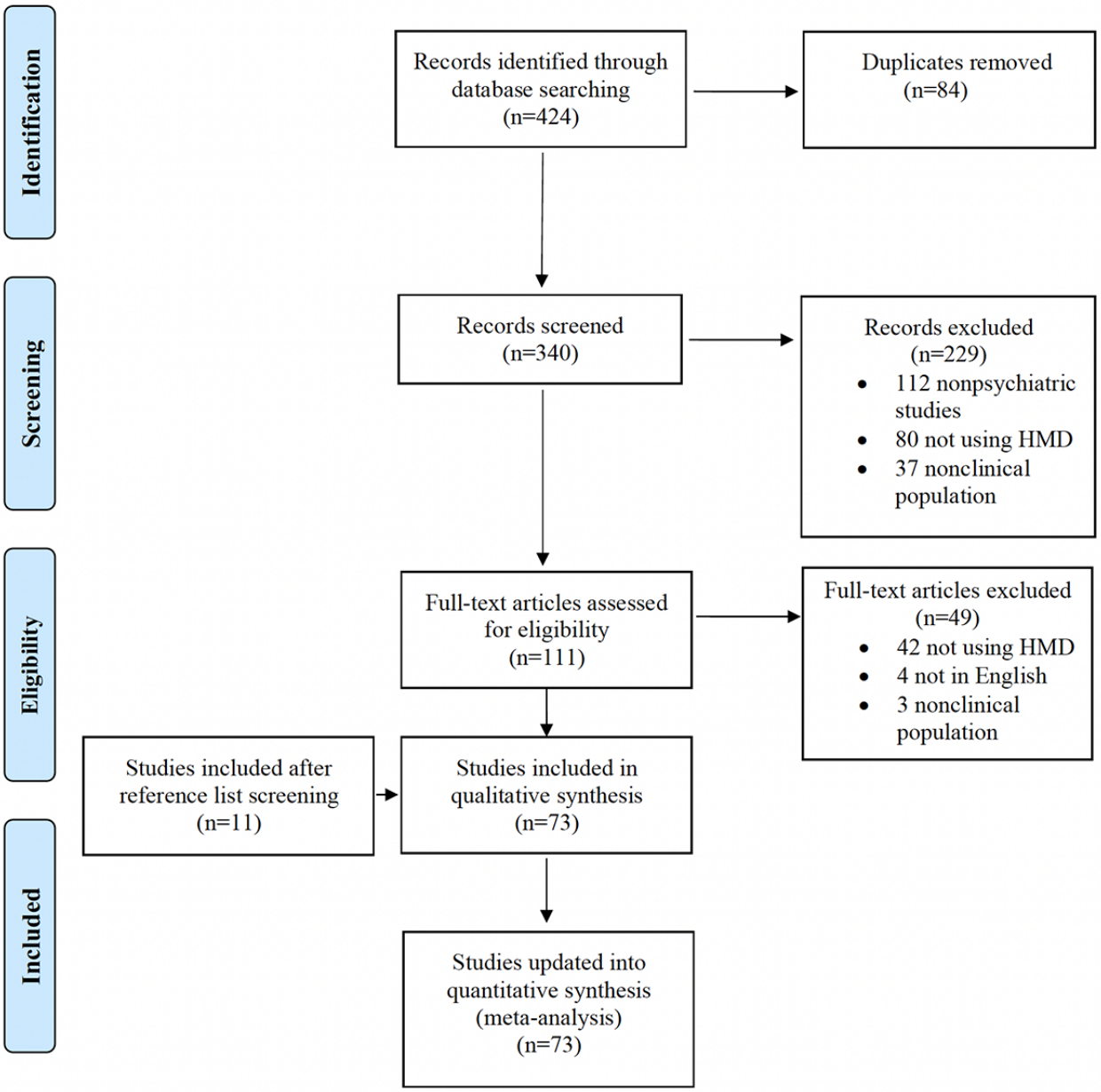
Because VR studies often lack a standardized research design, adverse effect reporting is inconsistent and needs to be improved to reckon VR's appropriate place in mental health [6,29]. Recommendations for the design of VR clinical trials, developed by an international consortium, address the safety and tolerability of equipment, headset, and intervention. Specific recommendations have highlighted the importance of assessing the psychological and emotional experiences of each participant [30]. Considering the prevalence of physical symptoms reported in the wider VR literature, it is concerning that mental health applications rarely report physical or psychological adverse effects [10,31,32], such as reduced cognitive performance [33,34], physical or eye fatigue [35], and cybersickness [36]. This review examines adverse effects reported in trials of VR and AR in mental health.

Methods

Identification and Selection of Studies

A systematic search of 3 databases (PsycINFO, PubMed, and Embase) concluded on September 17, 2022. These databases were selected as the 3 largest repositories for mental health studies. The search included the terms "virtual reality" or "augmented reality" or "computer-assisted" in combination with a range of mental health disorders to capture a breadth of relevant conditions including "mental health" or "psychiatry" and "treatment" or "therapy" or "intervention" or "psychotherapy" or "attention deficit hyperactivity disorder" or "dementia" or "cognitive impairment" or "depression" or "mood disorder" or "schizophrenia" or "psychosis" or "psychotic" or "phobia" or "anxiety" or "bipolar" or "PTSD" or "post-traumatic stress disorder" or "alcohol" or "substance" or "anorexia" or "bulimia" or "eating disorder" or "psychiatric" or "mental illness" or "mental health." The reference lists of studies included for full-text review were used to identify additional articles not captured by the initial search process. The literature search was conducted in accordance with PRISMA (Preferred Reporting Items of Systematic reviews and Meta-Analyses) guidelines as depicted in Figure 1. Duplicate studies were identified and removed; the remaining studies were screened by 2 authors (YY and RL).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) study selection flowchart. HMD: head-mounted display.



Eligibility Criteria

Screened papers were included for analysis if they (1) used a VR- or AR-based intervention administered via a head-mounted display (HMD) and (2) addressed a specific Diagnostic and Statistical Manual of Mental Disorders or International Classification of Diseases mental health disorder. The authors decided to focus on the use of HMDs because the term "virtual reality" is often used to reference nonimmersive alternatives such as computer screens or Cave Automatic Virtual Environments. If both conditions were not met, the paper was excluded. Final decisions on inclusion were reached by consensus, with papers grouped on the basis of disorder and data type. Selected papers were examined for information on the type of mixed reality used, psychological intervention implemented and target disorder, the number of patients in the data set, reported side effects, and any measure of presence or immersion.

Results

Study Selection

The systematic search of databases yielded 424 hits (Figure 1), with 84 duplicates. The examination of abstracts led to the exclusion of 85 articles, leaving 111 for full-text examination. In total, 42 studies were excluded for not using an HMD, 4 had no available translation, and 3 were based on nonclinical populations. Screening of reference lists yielded another 2 admissible reports. In total, 73 studies used VR or AR to treat a psychiatric diagnosis and were selected for analysis.

Study Characteristics

Of the 73 included studies, 67 used VR, 3 used AR, 2 used both VR and AR, and 1 used mixed reality. The main psychological intervention used was ET (n=52, 71%), followed by CBT without ET techniques (n=12, 16%). Non–CBT-based interventions (n=9, 12%) included embodiment illusion (n=2)

and provision of psychoeducation (n=1), relaxation (n=1), cognitive rehabilitation (n=1), physical and cognitive training (n=1), body swapping (n=1), interreality (n=1), and aggression prevention therapy (n=1). A variety of specific problems were targeted by these interventions, which are outlined in Table 1.

Of 73 studies included, only 22 measured levels of immersion or presence. The most popular rating tool was the Igroup Presence Questionnaire (n=6). Other tools included the SSQ (n=3), Presence and Reality Judgment questions (n=3), and an embodiment questionnaire (n=3) where 3 other measures were used once each (n=1). The remaining studies used unknown or subjective descriptions of presence or immersion (n=7).

Table 1. Mental disorders targeted by virtual reality and augmented reality in the included studies.

Mental disorder	Studies, n
Phobias	30
Posttraumatic stress disorder	17
Anxiety	10
Psychosis	7
Eating disorders	6
Stress	3
Cognitive impairment	1
Suicidal ideation	1
Aggression	1
Addiction	1
Depression	1

Reporting of Adverse Effects

Most studies (45/73, 62%) made no reference to adverse or side effects. Of the remaining 28, a total of 21 studies reported no adverse effects, but in 5 of them, adverse effects were apparent but not identified as such (Table 2).

Seven studies reported adverse effects but provided limited information (Table 3). Overall, the use of a tool that measured presence, immersion, or cybersickness was more common in the studies that commented on adverse effects (12/29, 27%) than in those that did not comment on adverse effects (12/44, 31%).

Data on dropouts from treatment groups were reported by 49 of 73 studies, but only 6 reported relevant reasons, including failure to arouse anxiety (n=2), high anxiety (n=2), anxiety along with a lack of engagement (n=1), finding the VR equipment “too distracting,“ distress from previous physical or sexual abuse, and avatars being “too unrealistic” (n=1).

Table 2. Adverse effects detected in studies reporting “no adverse effects.”

Study (year)	Participants in the exposure group, n	Adverse effect described
Pot-Kolder et al (2018) [37]	58	Dropouts due to cybersickness (n=1) and uncomfortable devices (n=2)
Pericot-Valverde et al (2015) [38]	41	Increased cigarette craving (n=4)
Nason et al (2019) [39]	7	Mild motion sickness (n=1)
Gujjar et al (2018) [40]	5	Cybersickness (n=4)
Veling et al (2021) [41]	50	Dropouts due to cybersickness (n=2)

Table 3. Studies reporting adverse effects.

Study (year)	Participants in the exposure group, n	Adverse effect described
Maltby et al (2002) [42]	20	Dropout due to cybersickness (n=1)
Reger et al (2016) [43]	52	Worsened symptoms of posttraumatic stress disorder (n=1)
Botella et al (2016) [44]	32	An unspecified number of participants reported tiredness, dizziness, or back pain, prompting a change in hardware
Krijn et al (2007) [45]	29	Dropout due to simulator sickness (n=1)
Levy et al (2016) [46]	9	"Walked carelessly" after therapy, judged to be at risk of falling (n=1)
Gujjar et al (2019) [47]	15	Mean increase in cybersickness rating post virtual reality
Kim et al (2020) [48]	32	Anxiety scores increased halfway through treatment (session 4) and then decreased

Discussion

This review highlights significant gaps in reporting the adverse effects of VR interventions in mental health. In a majority of included articles, authors made no mention whatsoever of possible adverse effects associated with technology. The identified studies specifically mention cybersickness, worsening of symptoms (posttraumatic stress disorder, anxiety, and cravings), tiredness, dizziness, back pain, and carelessness. Also concerning was the fact that even studies that mentioned possible adverse effects still did not recognize and report them as such. Available evidence indicates that adverse effects associated with VR are likely to be common, but the identified lack of good data makes it difficult to estimate rates with any confidence. Some studies also specifically focus on measures of immersion or presence or only report serious adverse events that would involve significant harm or death. While this is a valid approach for traditional clinical trials with large cohorts focused on medications with serious side effects, it is unlikely to be a sufficient approach for VR interventions in mental health. What is clear from these limited results is that use of VR and AR in mental health studies has adverse effects that include traditional vestibular-related side effects, physical experiences, and psychological impacts, which will all need to be considered.

Although recent attempts have been made to formulate a research framework for these interventions, these have not addressed the detection and reporting of adverse reactions. It is, therefore, important that robust protocols are developed to rectify this shortcoming [30,49]. This review highlights that fewer than half of all VR mental health studies report adverse effects. This points to a significant problem with reporting standards in these studies, which need to be addressed as the initial concern. This is particularly important because adverse effects can significantly influence the subject's emotional state and, therefore, the therapeutic effect of the intervention [16].

It is puzzling that many studies would use simple subjective measures of presence despite the many validated tools available. It has been established that the use of personal digital avatars, and especially seeing your own body in the virtual space, has a strong effect on presence [50]. Such factors are rarely discussed in mental health studies, which is an important omission as it would be particularly key in studies on eating disorders where participants are occasionally asked to assess body sizes.

Future studies will need to consider the range of physical and psychological adverse effects associated with AR and VR interventions and use screening tools that include physical symptoms such as cybersickness and VR-induced symptoms and effects, symptoms of dissociation, negative emotional responses such as increased rumination or thoughts of self-harm, and wider impacts on functioning. Although this would require the use of several tools, an ideal focus for future work would be to use a questionnaire covering all sections to standardize reporting of VR and AR interventions in mental health. This would include a move from older measures such as the SSQ tool to more dimensional assessments using the Virtual Reality Neuroscience Questionnaire tool, which provides grading and a broader approach to assess interventions. This could include the incorporation of established tools such as the Clinician-Administered Dissociative States Scale for dissociation and the Experience of Therapy Questionnaire [51].

As with the evaluation of any health care intervention, a benefit-risk analysis should be standard, where the dimension of risk includes the safety profile identified from the range and frequency of adverse effects. It is essential that all adverse effects, especially those that lead to dropouts or clinical worsening, need to be identified and reported. It is of concern that some studies specifically excluded participants experiencing cybersickness as the adverse events reported will not be representative of the general population. Thorough reporting would allow future systematic reviews to include meta-analyses of outcomes, including adverse effects, and improve our understanding of benefits, harm, and the appropriate place of VR and AR in psychiatric treatment.

This systematic review has multiple limitations. The study design focused specifically on HMDs and does not incorporate the wider (and generally older) literature regarding computer-generated environments projected onto screens. It also focused on diagnosed psychiatric disorders, and the results cannot be generalized to wider health care and nonclinical populations. There was also significant heterogeneity in the definitions of VR and AR in the literature, to the extent that our otherwise detailed search may not have captured some studies. Only papers published in English were included in the review.

Despite promising developments in VR and AR across a range of mental disorders, there is a clear need for standardized detection and reporting of adverse effects associated with these interventions.

Authors' Contributions

RML and DBM conceptualized the project. YY led the literature search and RML wrote the first draft of the manuscript. All authors edited and approved the submitted manuscript.

Conflicts of Interest

None declared.

References

1. Huhn M, Nikolakopoulou A, Schneider-Thoma J, Krause M, Samara M, Peter N, et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. Lancet 2019 Sep 14;394(10202):939-951 [FREE Full text] [doi: 10.1016/S0140-6736(19)31135-3] [Medline: 31303314]
2. Dellazizzo L, Potvin S, Luigi M, Dumais A. Evidence on virtual reality-based therapies for psychiatric disorders: meta-review of meta-analyses. J Med Internet Res 2020 Aug 19;22(8):e20889 [FREE Full text] [doi: 10.2196/20889] [Medline: 32812889]
3. Lundin RM, Menkes DB. Realising the potential of digital psychiatry. Lancet Psychiat 2021 Aug;8(8):655. [doi: 10.1016/s2215-0366(21)00165-6]
4. Lundin R, Menkes D. Commentary: managing virtual hybrid psychiatrist-patient relationships in a digital world. Front Public Health 2021;9:664778 [FREE Full text] [doi: 10.3389/fpubh.2021.664778] [Medline: 33928066]
5. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Mar 22;47(14):2393-2400. [doi: 10.1017/s003329171700040x]
6. Cieślik B, Mazurek J, Rutkowski S, Kiper P, Turolla A, Szczepańska-Gieracha J. Virtual reality in psychiatric disorders: a systematic review of reviews. Complement Ther Med 2020 Aug;52:102480 [FREE Full text] [doi: 10.1016/j.ctim.2020.102480] [Medline: 32951730]
7. Deng W, Hu D, Xu S, Liu X, Zhao J, Chen Q, et al. The efficacy of virtual reality exposure therapy for PTSD symptoms: a systematic review and meta-analysis. J Affect Disord 2019 Oct 01;257:698-709. [doi: 10.1016/j.jad.2019.07.086] [Medline: 31382122]
8. Morina N, Ijntema H, Meyerbröker K, Emmelkamp PM. Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. Behav Res Ther 2015 Nov;74:18-24. [doi: 10.1016/j.brat.2015.08.010] [Medline: 26355646]
9. Jerdan SW, Grindle M, van Woerden HC, Kamel Boulos MN. Head-mounted virtual reality and mental health: critical review of current research. JMIR Serious Games 2018 Jul 06;6(3):e14 [FREE Full text] [doi: 10.2196/games.9226] [Medline: 29980500]
10. Park MJ, Kim DJ, Lee U, Na EJ, Jeon HJ. A literature overview of virtual reality (VR) in treatment of psychiatric disorders: recent advances and limitations. Front Psychiatry 2019 Jul 19;10:505 [FREE Full text] [doi: 10.3389/fpsyt.2019.00505] [Medline: 31379623]
11. Klatte R, Strauss B, Flückiger C, Rosendahl J. Adverse effects of psychotherapy: protocol for a systematic review and meta-analysis. Syst Rev 2018 Sep 08;7(1):135 [FREE Full text] [doi: 10.1186/s13643-018-0802-x] [Medline: 30193585]
12. Berk M, Parker G. The elephant on the couch: side-effects of psychotherapy. Aust N Z J Psychiatry 2009 Sep 01;43(9):787-794. [doi: 10.1080/00048670903107559] [Medline: 19670051]
13. Scott J, Young AH. Psychotherapies should be assessed for both benefit and harm. Br J Psychiatry 2016 Mar 02;208(3):208-209. [doi: 10.1192/bjp.bp.115.169060] [Medline: 26932480]
14. Meister R, von Wolff A, Mohr H, Nestoriuc Y, Härter M, Hölzel L, et al. Adverse event methods were heterogeneous and insufficiently reported in randomized trials on persistent depressive disorder. J Clin Epidemiol 2016 Mar;71:97-108. [doi: 10.1016/j.jclinepi.2015.10.007] [Medline: 26482955]
15. Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: a review. Front Psychol 2019 Feb 4;10:158 [FREE Full text] [doi: 10.3389/fpsyg.2019.00158] [Medline: 30778320]
16. van Heugten-van der Kloet D, Cosgrave J, van Rheede J, Hicks S. Out-of-body experience in virtual reality induces acute dissociation. Psychol Conscious (Wash D C) 2018 Dec;5(4):346-357. [doi: 10.1037/cns0000172]
17. Kourtesis P, Collina S, Doumas LAA, MacPherson SE. Validation of the virtual reality neuroscience questionnaire: maximum duration of immersive virtual reality sessions without the presence of pertinent adverse symptomatology. Front Hum Neurosci 2019 Nov 26;13:417 [FREE Full text] [doi: 10.3389/fnhum.2019.00417] [Medline: 31849627]
18. Virre E. Virtual reality and the vestibular apparatus. IEEE Eng Med Biol Mag 1996;15(2):41-43, 69. [doi: 10.1109/51.486717]

19. Mondellini M, Mottura S, Guida M, Antonietti A. Influences of a virtual reality experience on dissociation, mindfulness, and self-efficacy. Cyberpsychol Behav Soc Netw 2021 Nov 01;24(11):767-771. [doi: 10.1089/cyber.2020.0223] [Medline: 33739862]
20. Aardema F, O'Connor K, Côté S, Taillon A. Virtual reality induces dissociation and lowers sense of presence in objective reality. Cyberpsychol Behav Soc Netw 2010 Aug;13(4):429-435. [doi: 10.1089/cyber.2009.0164] [Medline: 20712501]
21. Martirosov S, Bureš M, Zítka T. Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality. Virtual Real 2022 May 19;26(1):15-32 [FREE Full text] [doi: 10.1007/s10055-021-00507-4] [Medline: 34025203]
22. Lavoie R, Main K, King C, King D. Virtual experience, real consequences: the potential negative emotional consequences of virtual reality gameplay. Virtual Reality 2020 Apr 02;25(1):69-81. [doi: 10.1007/s10055-020-00440-y]
23. Fernández-Álvarez J, Rozental A, Carlbring P, Colombo D, Riva G, Anderson PL, et al. Deterioration rates in virtual reality therapy: an individual patient data level meta-analysis. J Anxiety Disord 2019 Jan;61:3-17. [doi: 10.1016/j.janxdis.2018.06.005] [Medline: 30057346]
24. Sevinc V, Berkman MI. Psychometric evaluation of Simulator Sickness Questionnaire and its variants as a measure of cybersickness in consumer virtual environments. Appl Ergon 2020 Jan;82:102958. [doi: 10.1016/j.apergo.2019.102958] [Medline: 31563798]
25. Pallavicini F, Cipresso P, Raspelli S, Grassi A, Serino S, Vigna C, et al. Is virtual reality always an effective stressors for exposure treatments? Some insights from a controlled trial. BMC Psychiatry 2013 Feb 11;13(1). [doi: 10.1186/1471-244x-13-52]
26. Keshavarz B, Hecht H. Validating an efficient method to quantify motion sickness. Hum Factors 2011 Aug 11;53(4):415-426. [doi: 10.1177/0018720811403736] [Medline: 21901938]
27. Somrak A, Pogačnik M, Guna J. Suitability and comparison of questionnaires assessing virtual reality-induced symptoms and effects and user experience in virtual environments. Sensors (Basel) 2021 Feb 08;21(4):1185 [FREE Full text] [doi: 10.3390/s21041185] [Medline: 33567570]
28. Klein Tuente S, Bogaerts S, van IJzendoorn S, Veling W. Effect of virtual reality aggression prevention training for forensic psychiatric patients (VRAPT): study protocol of a multi-center RCT. BMC Psychiatry 2018 Aug 06;18(1):251 [FREE Full text] [doi: 10.1186/s12888-018-1830-8] [Medline: 30081863]
29. Duggan C, Parry G, McMurran M, Davidson K, Dennis J. The recording of adverse events from psychological treatments in clinical trials: evidence from a review of NIHR-funded trials. Trials 2014 Aug 27;15(1):335 [FREE Full text] [doi: 10.1186/1745-6215-15-335] [Medline: 25158932]
30. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
31. Bouchard S, St-Jacques J, Renaud P, Wiederhold B. Side effects of immersions in virtual reality for people suffering from anxiety disorders. J Cyber Ther Rehabil 2009;2(2):127-137.
32. P. W. Cybersickness and anxiety during simulated motion: Implications for VRET. Front Neuroeng 2009;2. [doi: 10.3389/conf.neuro.14.2009.06.014]
33. Mittelstaedt JM, Wacker J, Stelling D. VR aftereffect and the relation of cybersickness and cognitive performance. Virtual Reality 2018 Nov 10;23(2):143-154. [doi: 10.1007/s10055-018-0370-3]
34. Szpak A, Michalski SC, Saredakis D, Chen CS, Loetscher T. Beyond feeling sick: the visual and cognitive aftereffects of virtual reality. IEEE Access 2019;7:130883-130892. [doi: 10.1109/access.2019.2940073]
35. Smith SP, Burd EL. Response activation and inhibition after exposure to virtual reality. Array 2019 Sep;3-4:100010. [doi: 10.1016/j.array.2019.100010]
36. Szpak A, Michalski SC, Loetscher T. Exergaming with beat saber: an investigation of virtual reality aftereffects. J Med Internet Res 2020 Oct 23;22(10):e19840 [FREE Full text] [doi: 10.2196/19840] [Medline: 33095182]
37. Pot-Kolder RMCA, Geraets CNW, Veling W, van Beilen M, Staring ABP, Gijsman HJ, et al. Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: a single-blind randomised controlled trial. Lancet Psychiatry 2018 Mar;5(3):217-226. [doi: 10.1016/S2215-0366(18)30053-1] [Medline: 29429948]
38. Pericot-Valverde I, García-Rodríguez O, Gutiérrez-Maldonado J, Secades-Villa R. Individual variables related to craving reduction in cue exposure treatment. Addict Behav 2015 Oct;49:59-63. [doi: 10.1016/j.addbeh.2015.05.013] [Medline: 26047836]
39. Nason EE, Trahan M, Smith S, Metsis V, Selber K. Virtual treatment for veteran social anxiety disorder: a comparison of 360° video and 3D virtual reality. J Technol Hum Serv 2019 Nov 28;38(3):288-308. [doi: 10.1080/15228835.2019.1692760]
40. Gujjar KR, van Wijk A, Sharma R, de Jongh A. Virtual reality exposure therapy for the treatment of dental phobia: a controlled feasibility study. Behav Cogn Psychother 2017 Sep 14;46(3):367-373. [doi: 10.1017/s1352465817000534]
41. Veling W, Lestestuiver B, Jongma M, Hoenders HJR, van Driel C. Virtual reality relaxation for patients with a psychiatric disorder: crossover randomized controlled trial. J Med Internet Res 2021 Jan 15;23(1):e17233 [FREE Full text] [doi: 10.2196/17233] [Medline: 33448933]

42. Maltby N, Kirsch I, Mayers M, Allen GJ. Virtual reality exposure therapy for the treatment of fear of flying: a controlled investigation. J Consult Clin Psychol 2002 Oct;70(5):1112-1118. [doi: 10.1037/0022-006x.70.5.1112]
43. Reger GM, Koenen-Woods P, Zetocha K, Smolenski DJ, Holloway KM, Rothbaum BO, et al. Randomized controlled trial of prolonged exposure using imaginal exposure vs. virtual reality exposure in active duty soldiers with deployment-related posttraumatic stress disorder (PTSD). J Consult Clin Psychol 2016 Nov;84(11):946-959. [doi: 10.1037/ccp0000134] [Medline: 27606699]
44. Botella C, Pérez-Ara MÁ, Bretón-López J, Quero S, García-Palacios A, Baños RM. In vivo versus augmented reality exposure in the treatment of small animal phobia: a randomized controlled trial. PLoS One 2016 Feb 17;11(2):e0148237 [FREE Full text] [doi: 10.1371/journal.pone.0148237] [Medline: 26886423]
45. Krijn M, Emmelkamp PMG, Olafsson RP, Bouwman M, van Gerwen LJ, Spinhoven P, et al. Fear of flying treatment methods: virtual reality exposure vs. cognitive behavioral therapy. Aviat Space Environ Med 2007 Feb;78(2):121-128. [Medline: 17310883]
46. Levy F, Rautureau G, Komano O, Millet B, Jouvent R, Leboucher P. Fear of falling: efficacy of virtual reality associated with serious games in elderly people. NDT 2016 Apr:877. [doi: 10.2147/ndt.s97809]
47. Gujjar KR, van Wijk A, Kumar R, de Jongh A. Efficacy of virtual reality exposure therapy for the treatment of dental phobia in adults: a randomized controlled trial. J Anxiety Disord 2019 Mar;62:100-108. [doi: 10.1016/j.janxdis.2018.12.001] [Medline: 30717830]
48. Kim H, Lee S, Jung D, Hur J, Lee H, Lee S, et al. Effectiveness of a participatory and interactive virtual reality intervention in patients with social anxiety disorder: longitudinal questionnaire study. J Med Internet Res 2020 Oct 06;22(10):e23024 [FREE Full text] [doi: 10.2196/23024] [Medline: 33021481]
49. Junqueira D, Phillips R, Zorzela L, Golder S, Loke Y, Moher D, et al. Time to improve the reporting of harms in randomized controlled trials. J Clin Epidemiol 2021 Aug;136:216-220 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.04.020] [Medline: 33984494]
50. Lee J, Kim M, Kim J. A study on immersion and VR sickness in walking interaction for immersive virtual reality applications. Symmetry 2017 May 22;9(5):78. [doi: 10.3390/sym9050078]
51. Parker G, Paterson A, Fletcher K, McClure G, Berk M. Construct validity of the Experiences of Therapy Questionnaire (ETQ). BMC Psychiatry 2014 Dec 31;14:369 [FREE Full text] [doi: 10.1186/s12888-014-0369-6] [Medline: 25551579]

Abbreviations

AR: augmented reality
CBT: cognitive behavioral therapy
ET: exposure therapy
HMD: head-mounted display
SSQ: Simulator Sickness Questionnaire
VR: virtual reality

Edited by J Torous; submitted 05.10.22; peer-reviewed by Y Jung, B Birckhead; comments to author 04.12.22; revised version received 22.12.22; accepted 05.01.23; published 05.05.23.

Please cite as:
Lundin RM, Yeap Y, Menkes DB
Adverse Effects of Virtual and Augmented Reality Interventions in Psychiatry: Systematic Review
JMIR Ment Health 2023;10:e43240
URL: https://mental.jmir.org/2023/1/e43240
doi:10.2196/43240
PMID:37145841

Review

The Use of Photoplethysmography in the Assessment of Mental Health: Scoping Review

Lynnette Nathalie Lyzwinski[1,2*], MPhil, PhD; Mohamed Elgendi[3*], PhD; Carlo Menon[2,3], PhD

[1]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[2]Menrva Research Group, Schools of Mechatronic Systems Engineering and Engineering Science, Simon Fraser University, Vancouver, BC, Canada

[3]Biomedical and Mobile Health Technology Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Carlo Menon, PhD
Biomedical and Mobile Health Technology Lab
Department of Health Sciences and Technology, ETH Zurich
Lengghalde 5
Zurich, 8008
Switzerland
Phone: 41 44 510 72 2
Email: carlo.menon@hest.ethz.ch

Abstract

Background: With the rise in mental health problems globally, mobile health provides opportunities for timely medical care and accessibility. One emerging area of mobile health involves the use of photoplethysmography (PPG) to assess and monitor mental health.

Objective: In recent years, there has been an increase in the use of PPG-based technology for mental health. Therefore, we conducted a review to understand how PPG has been evaluated to assess a range of mental health and psychological problems, including stress, depression, and anxiety.

Methods: A scoping review was performed using PubMed and Google Scholar databases.

Results: A total of 24 papers met the inclusion criteria and were included in this review. We identified studies that assessed mental health via PPG using finger- and face-based methods as well as smartphone-based methods. There was variation in study quality. PPG holds promise as a potential complementary technology for detecting changes in mental health, including depression and anxiety. However, rigorous validation is needed in diverse clinical populations to advance PPG technology in tackling mental health problems.

Conclusions: PPG holds promise for assessing mental health problems; however, more research is required before it can be widely recommended for clinical use.

(JMIR Ment Health 2023;10:e40163) doi:10.2196/40163

KEYWORDS

photoplethysmography; PPG; mental health; depression; anxiety; suicide; mobile phone

Introduction

Background

Photoplethysmography (PPG) is an emerging technology that uses light to measure microvascular blood changes to assess various cardiovascular parameters that are useful in clinical settings [1,2], including blood pressure [3,4] and heart rate (HR) [5]. This technology can quickly and easily assess vasoconstriction and dilation in the skin, autonomic function changes, and vascular functions [6]. PPG assesses these physiological changes through a clip placed on a body part, such as the finger or ear [7,8], and smartphone apps [5,9].

In recent years, PPG technology has been extended to include the assessment of mental health changes in patients [10], which is of great pertinence to patients who experience challenges in accessing mental health care. Many individuals, particularly those living in remote areas, struggle to obtain equitable access to mental health services [11]. Mental health disparities and the need for mental health interventions in remote settings were especially pronounced during the COVID-19 pandemic [12].

In addition, an increase in mental health problems [13] and the risk of suicide [14] has been of concern during the pandemic era, stemming from the lack of social support (interpersonal theory of suicide) [15]. It is well-known that mental health problems are linked to an increased risk of cardiovascular disease through direct effects on the myocardium, including increased HR and blood pressure [16-19]. PPG may help provide a rapid mental health assessment for patients by assessing changes in HR variability (HRV), and a subsequent medical referral to a specialist can be made [3]. Rural and remote patients might possibly also feel greater support if PPG is provided alongside remote telemental health consultations [20].

As PPG is an emerging technology and its application in assessing mental health problems has only recently been used, its overall effectiveness, accuracy, and challenges are not entirely clear. There is a need to determine whether PPG is an effective technological medium for the screening and prediagnosis of mental health problems as a complement to the currently used methods, including clinical evaluations based on self-reported symptoms [21]. In other words, PPG may be used by a wide range of mental health practitioners as a useful adjunct tool. For example, mental health professionals may use it as a prescreener before making referrals. In addition, psychiatrists, psychologists, and other allied health professionals [22] may use it to monitor patients' symptoms in a remote setting.

There is also a need to evaluate how PPG-based technology can be integrated into current standard methods for diagnosing patients with mental health problems. Currently, standard guidelines require practitioners to assess patients using the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [23]. Diagnosis is usually face-to-face but may be virtual through telehealth for remote patients [20]. A crucial research question is whether PPG is a valid and reliable method for assessing mental health and psychological problems. In other words, how much in-depth information can PPG blood flow changes offer compared with traditional face-to-face assessments by mental health professionals for assessing mental health? The other question is whether PPG can be used alone (without requiring questionnaires) for screening. Furthermore, there is also a need to determine what types of studies have been conducted on PPG and mental health, and whether there is any evidence supporting the use of this technology for managing mental health problems.

Objectives

The objectives of this review were as follows:

1. To identify the types of PPG studies conducted on mental health. Have any smartphone apps (mobile health approaches) been developed for mental health using PPG?
2. To determine the effectiveness of PPG for assessing, screening, or monitoring common mental health problems [24] (with an emphasis on depression and anxiety). Symptoms relevant to mental health, such as stress, are also of interest.
3. To assess the validity and reliability of the PPG-based mental-health methods
4. To identify challenges associated with this new technology and make recommendations for future research

Our main outcome of interest was efficacy and accuracy (validity and reliability) in screening for depression, anxiety, or psychological distress in individuals. In other words, how accurately can PPG assess the mental health end point? The secondary outcomes of interest were any identified barriers or challenges with existing PPG-based technologies proposed for mental health problems.

Methods

Overview

PubMed and Google Scholar were searched for articles published in January 2021. The search terms included word variations for PPG and mental health including depression, anxiety, or stress. The search strings were combined into one large search string that was the most relevant to PPG and mental health. Keywords included Medical Subject Headings (MeSH) terms, free text, word variations, and truncation. We discussed the search strategy, including keyword variations, with a medical librarian during the search planning phase. In addition, manual searches of articles meeting the inclusion criteria were performed to identify additional articles. It should be noted that we reran the search, with the term "photoplethysmogram" again in December 2022, to determine if any studies were missed in PubMed. This generated 33 additional results that were irrelevant.

Inclusion and Exclusion Criteria

The literature search was limited to articles published over the past decade. The search strategy is defined as follows: "Photoplethysmography"[Mesh] OR Photoplethysmography[tiab] OR PPG [ti] AND "Depression"[Mesh] OR depression[tiab] OR depress*[tiab] OR "anxiety"[mesh] OR "anxiety"[tiab] OR "Mental Health"[Mesh] OR "Stress, Psychological"[Mesh] OR stress[tiab] OR "Quality of Life"[Mesh] OR quality of life[tiab] OR QoL[tiab] OR well-being[tiab].

All intervention studies were included if they evaluated mental health conditions or symptoms in relation to the PPG. Studies must have evaluated clinical mental health problems, such as depression, anxiety, or schizophrenia, without other comorbidities. Studies that evaluated psychological stress were included only if they were in a population with a history of a clinical mental health problem, especially anxiety. High psychological stress has been linked to anxiety and depression [25,26]; hence, evaluating it in patients with clinical anxiety is relevant for monitoring its progression. Stress is also a risk factor for mental health problems and can be managed through psychological interventions [27,28]. In addition, studies evaluating other symptoms in patients with mental health problems that came up in the search were included if they involved intrusive thoughts or suicidal thoughts and suicidal ideation. There has also been a consideration for including suicidal behavior as a separate mental health disorder in the DSM-5. General stress studies were included only if they evaluated at least one mental health disorder in clinical populations, particularly anxiety. Furthermore, the studies must have used a screening tool for diagnosing anxiety, such as the General Anxiety Scale [29], when comparing it with PPG.

General studies on psychological stress in nonclinical populations and physiological stress were excluded. Studies involving participants diagnosed and treated by clinicians in a face-to-face setting (nontechnological) or over telehealth without the assistance of artificial intelligence for advancing technology such as PPG (this allows a comparison of the accuracy of diagnosis) were excluded. Studies that did not use PPG technology and those that used PPG technology for purposes other than mental health assessment or preassessment, complementary screening or press-screening, and monitoring were also excluded. Finally, the studies had to have been published in the English language.

Screening

The citations of the included papers were imported into EndNote (version X7.1; Clarivate). Relevant titles were screened, followed by abstract screening of titles that met the inclusion criteria. Two independent reviewers, LNL and ME, screened the abstracts to ensure that the articles fulfilled the inclusion criteria. Once there was an agreement, full texts were retrieved from the articles that met the inclusion criteria. Where there was any disagreement between the reviewers, the articles in question were discussed until agreement was reached.

Data Extraction and Synthesis

Data on the general study characteristics and participant demographics were extracted and summarized in a tabular format. These data included the authors, years of publication, study design, country, participants (age and sex [male or female]), mental health history, anatomical site of PPG measurement, and mental health measures (anxiety, stress, depression, relaxation, and happiness). It also included outcomes, such as the effectiveness of PPG for diagnosing depression, anxiety, or stress. Other outcomes included data on the validity and reliability of PPG (eg, Bland-Altman plots comparing PPG with standard methods of diagnosis), as well as qualitative data on barriers associated with PPG technology. This review followed the PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) guidelines [30].

Results

Search Results

The details of the search process are shown in Figure 1. We identified 3393 records. After title screening and review of 114 abstracts against the inclusion and exclusion criteria, followed by full-text retrieval, 24 studies were included in the review [31-54]. It should be noted that the 2 sets of studies were sister papers [32,37,38,51,52]. The sister papers are those authored by the same author who published multiple articles using the same data set. However, each article is treated independently in this study.

The main characteristics of the included articles, including mental health conditions, diagnostic methods if provided, and population (including the control group, if provided), are described in Table 1.

The studies were conducted in Japan [33,40,51,54], China [44,45], South Korea [52], Abu Dhabi [38], Taiwan [43], Spain [42], Turkey [31], Sweden [39], Canada [35], Germany [32,53], and the United States [41,48]. Most of these were proof-of-concept studies with a small number of participants. Three studies had a control group with sample sizes ranging from 61 to 308 [37,42,52]. The age range was broad across the studies, ranging from 16 to 85 years. The studies were relatively balanced in terms of sex (male or female), although not all reported this information. One study was conducted only on female participants [47].

Most studies used finger-based PPG methods to evaluate mental health outcomes and psychological stress [31-33,38,39,43,46,49,51]. Some researchers have used this arm for PPG measurement [10]. A few have used smartphone-based technology, which includes a PPG finger sensor and video camera [35,44] or face-based webcam technology [54]. This implies that mobile methods, including smartphones, for PPG are scarce and are emerging as novel methods for assessing mental health.

Figure 1. Summary of articles included and excluded in this study. PPG: photoplethysmography.

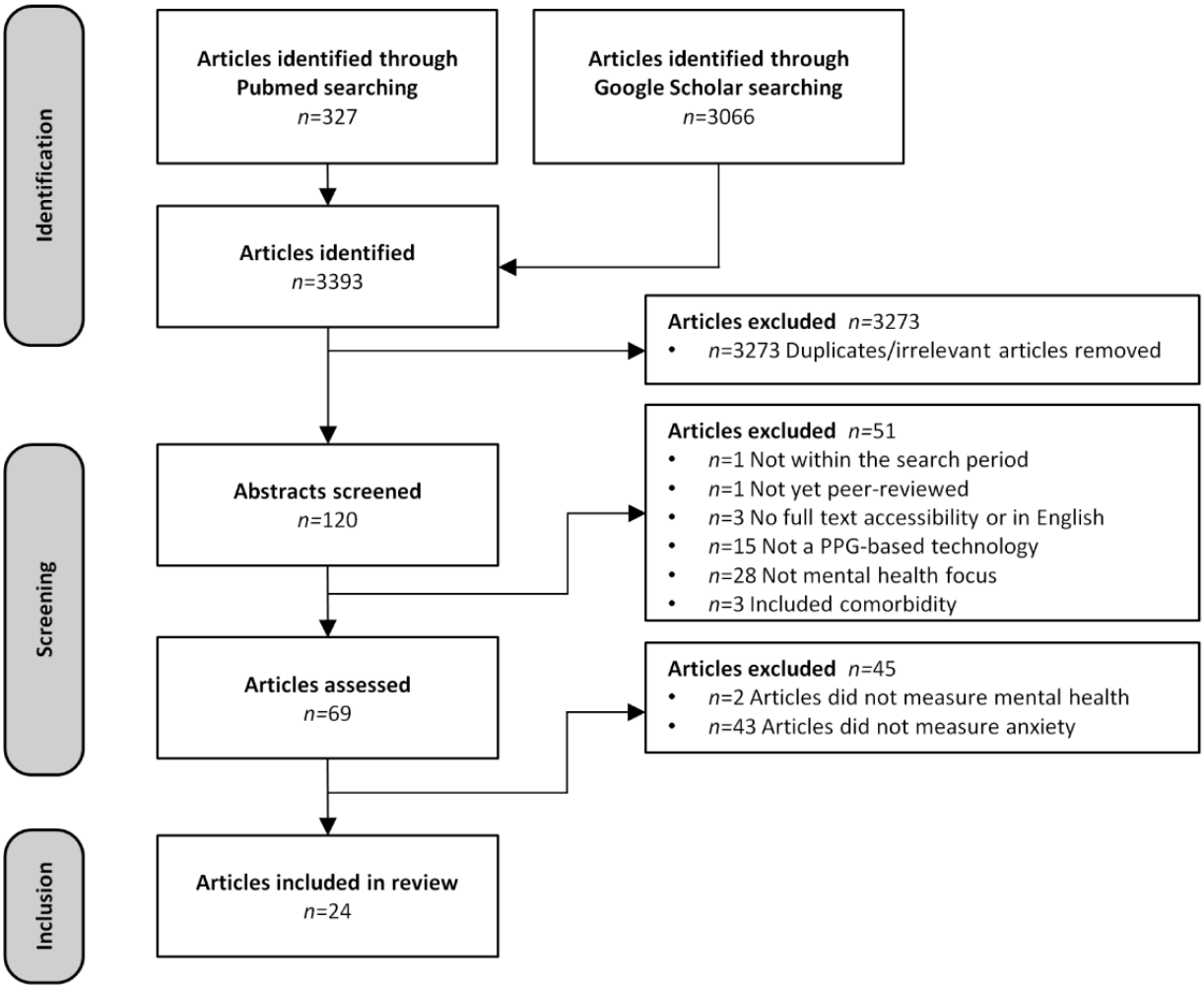


Table 1. Study characteristics.

Year	Study	Sample size, N	Sex ratio (% F[a])	Age range (years)	Mental health measures	Gold standard	Comorbidity or confounding factors	Evaluation metric (mental health vs control)
2022	de Vries et al [34]	9	22.3	26-52	Depression and anxiety	4DSQ[b]	To account for confounders, control variables for TST[c], MVPA[d], and alcohol use were included	Association between HRV[e] fluctuations and stress increase
2021	Sigrist et al [53]	27	100	13-17	Borderline personality disorder	SCID-II[f], severity of psychopathology (CGI-S[g]), and GAF[h]	Controlled for the potential confounders of age, BMI, menstrual cycle, cardioactive medication, oral contraceptives, smoking behavior, alcohol consumption and drug consumption, physical activity level, and illness	HR[i] significantly correlated with gold standard with $P<.05$
2021	Long et al [45]	36	53	22-43	Depression	PHQ-9[j]	N/R[k]	Developed a classifier to differentiate between depressed and health subjects with accuracy=85.3%
2021	Harvie et al [35]	104	35	17-50	Stress	TSST[l]	N/R	Significant acceleration in HR and heightened self-reported stress and anxiety in the TSST condition relative to a placebo version of the TSST with $P<.05$
2021	Unursaikhan et al [54]	53	42	18-60	MDD[m]	SDS[n]	N/R	HRV and SDS correlated significantly ($P<.05$) and can detect MDD with sensitivity=73% and specificity=85%
2021	Kontaxis et al [42]	80	60	32-59	MDD	HDRS[o] and BDI[p]	N/R	HR and amplitude difference between systolic and diastolic peaks are significantly different in healthy and MDD subjects
2021	Kimmel et al [39]	126	100	22-43	Panic disorder, phobia, trait anxiety, and traumatic events	EPDS[q] and the STAI[r]	Comorbidities were reported	LF[s]/HF[t] ratio significantly correlated with anxiety with $P<.001$
2021	Moshe et al [48]	55	54.5	41-54	Depression and anxiety	DASS-21[u]	N/R	HRV and anxiety are significantly correlated with $P<.05$
2020	Perpetuini [50]	102	47	20-70	Anxiety	STAI-Y[v]	N/R	HRV and STAI-Y score are correlated with $r=0.81$
2020	Liu et al [44]	93	54	21-28	Depression	SWLS[w]	N/R	HRV significantly differentiates between high and low SWLS scores with $P<.05$
2020	Ismail et al [36]	34	61.8	20-27	SAD[x]	DSM-5[y] and LSAS[z]	N/R	HR can detect SAD with accuracy=88.9%

Year	Study	Sample size, N	Sex ratio (% F[a])	Age range (years)	Mental health measures	Gold standard	Comorbidity or confounding factors	Evaluation metric (mental health vs control)
2019	Mistuhashi et al [47]	7	28.5	22-24	Stress and trait anxiety	N/R	N/R	HR detects stress with accuracy=80%
2018	Ngamprmuan et al [49]	40	65	20-43	MDD	Wisconsin Card Sorting Test	N/R	HRV significantly differentiates between MDD and control subjects with *P*<.001
2018	Koening et al [41]	90	67.8	14-19	MDD	Clinicians' ratings of depression severity and BDI-II	Comorbidities were reported	HRV significantly differentiates between depression and control subjects with *P*<.05
2018	Dagdanpurev et al [33]	20	50	23-60	MDD	N/R	N/R	Developed a classifier to differentiate between MDD and control subjects with sensitivity=83% and specificity=93%
2017	Khandoker et al [38]	61[aa]	68[aa]	26-48[aa]	MDD with suicidal ideation	Neuropsychiatric Interview (MINI) version 5 and the Hamilton Depression Rating Scale	N/R	A multivariate logistic regression classifier to differentiate between MDD and control with accuracy=93.3%
2017	Kobayashi et al [40]	55	49	21-59	MDD with paced respiration	N/R	N/R	HR significantly differentiates between MDD and control subjects with *P*<.05
2016	Khandoker et al [37]	61	68	22-40	MDD with suicidal ideation	Neuropsychiatric Interview (MINI) version 5 and the Hamilton Depression Rating Scale	N/R	A decision tree classifier to differentiate between MDD and control with accuracy=93.4%
2014	Akar et al [31]	39	43.5	16-29	Schizophrenia	DSM-IV criteria	Subjects with diabetes mellitus, hypertension, respiratory diseases, cardiovascular diseases, and comorbidity were excluded	HRV significantly differentiates between patients with schizophrenia and control subjects with *P*<.05
2014	Clamor et al [32]	115	41.7	25-42	Psychosis and depression	Structured mini-international neuropsychiatric interview and community assessment of psychic experience	N/R	HRV was found for psychosis, with significant differences to healthy controls (all *P*≤.007) and to depression (all *P*≤.004),
2013	Minakuchi et al [46]	31	100	20-24	Mental stress	Profile of Mood States	N/R	LF/HF ratio significantly associated with Profile of Mood States with *P*<.01
2013	Lee et al [43]	5	N/R	N/R	State anxiety and trait anxiety	STAI	N/R	N/R
2013	Pham et al [52] (Second experiment)	35[ab]	N/R	N/R	Depression	N/R	N/R	Automated detection of mental illness is feasible using k-nearest neighbours with a sensitivity of 99.03%[aa] and specificity of 95.09%[aa]

Year	Study	Sample size, N	Sex ratio (% F[a])	Age range (years)	Mental health measures	Gold standard	Comorbidity or confounding factors	Evaluation metric (mental health vs control)
2012	Pham et al [51] (First experiment)	308	N/R	N/R	Depression	N/R	N/R	Automated detection of mental illness is feasible using support vector machines with a sensitivity of 99.78% and specificity of 99.28%

[a]F: female.

[b]4DSQ: four-dimensional symptom questionnaire.

[c]TST: total sleep time.

[d]MVPA: moderate-to-vigorous physical activity.

[e]HRV: heart rate variability.

[f]SCID-II: Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders–IV-Axis II.

[g]CGI-S:

[h]GAF: global level of functioning.

[i]HR: heart rate.

[j]PHQ-9: Patient Health Questionnaire–9.

[k]N/R: not reported.

[l]TSST: Trier Social Stress Test.

[m]MDD: major depressive disorder.

[n]SDS: Self-Rating Depression Scale.

[o]HDRS: Hamilton depression rating scale.

[p]BDI: Beck Depression Inventory.

[q]EPDS: Edinburgh Postnatal Depression Scale.

[r]STAI: State-Trait Anxiety Inventory for Adults.

[s]LF: low frequency.

[t]HF: high frequency.

[u]DASS-21: Depression Anxiety Stress Scales–21.

[v]STAI-Y: State-Trait Anxiety Inventory.

[w]SWLS: Satisfaction With Life Scale.

[x]SAD: social anxiety disorder.

[y]DSM-5: fifth edition of the Diagnostic and Statistical Manual of Mental Disorder.

[z]LSAS: Liebowitz Social Anxiety Scale.

[aa]These values correspond to the dataset that the same authors reported in [37]

[ab]The values presented here pertain to the second experiment that is reported in this journal paper. The first experiment, which the same authors had published earlier in a conference paper [51], is different from the one presented here.

Depression

Several studies have evaluated the use of PPG for diagnosing depression [32-34,37,38,40,41,44,45,48,49,52,54]. A study in Japan evaluated the use of finger PPG for major depressive disorder (MDD) and found that it could accurately detect it, with a very high correlation with electrocardiogram (ECG)-based HRV measures [33]. This included the high frequency (HF) component of HRV measures and the ratio of low frequency (LF) to HF, which is known as LF/HF HRV. The sensitivity and specificity for diagnosing MDD were 83% and 93%, diagnosing MDD [33].

In addition to finger-based PPG methods, face-based (video) methods have also been found to be accurate for diagnosing depression. A study in Japan found that a PPG web camera–based system that detects blood flow to the face (arteries in the sinus region) had a sensitivity of 73% and specificity of 85% for diagnosing MDD [54]. A considerable correlation was also found between depression scores (subjective self-report) and video-based PPG assessments of MDD. This technology has also been correlated with ECG findings [54]. Similarly, another study in Japan also found an 86% sensitivity in assessing MDD during exercise when compared with controls, although it was only 68% when exposed to a mental stress test [40].

The most advanced PPG technology integrates video- and finger-based PPG methods, using smartphones that simultaneously make use of both. A study in China found that smartphone-based PPG could be used to assess happiness, as subjectively reported that life satisfaction was positively correlated with an HRV index, specifically the root mean square of beat intervals (RMSDD) and negatively correlated with log HF and log LF. There was also a negative correlation between these HRV indices and depression scores [44]. Another study also found a considerably negative relationship between autonomic reactivity assessed via PPG and MDD and that this

relationship was considerable when compared with controls without MDD [42].

Interestingly, one study evaluated brain changes in depression and compared them with HRV. Researchers found a positive association between RMSDD, cortical thickness, self-reported depression, and HF HR [41]. Furthermore, a study in Germany found that patients who are depressed with psychosis had significantly lower RMSDD and SD of the N-N intervals than controls, whereas their HR and LF/HF ratio were significantly higher than in controls [32]. Similar findings with respect to increased RMSDD, SD of the N-N intervals, and LF/HR ratios were noted in patients who are depressed with MDD in Thailand relative to controls, and cortisol levels were also elevated [49]. Thus, there is a growing body of evidence suggesting that there are differences in HRs among patients with depression relative to healthy controls. In addition, these objective brain imaging data support the changes noted in patients with depression using HRV [41].

Anxiety Conditions

Most studies on anxiety used validated anxiety measures, such as the State-Trait Anxiety Scale [55], the Generalized Anxiety Disorder [56], and the Depression Anxiety Stress Scale [57,58], to validate their findings alongside PPG assessments. Most studies have involved comparisons between relaxation exercises and subsequent psychological stress-inducing exercises. In general, HR increased during stress-provoking tasks and declined during relaxation across the studies (Celka et al [10], Kontaxis et al[42], Lee et al [43], Chen et al [64], Panganiban et al [65]). Physiological distress responses in HR were similar for both anxiety and stress, with increases in HR across studies. A study in Canada found a strong association between self-reported anxiety and stress and subsequently elevated HR using smartphone-based PPG assessment [35]. Another study evaluated the Oura Ring, which is a commercially available wearable sleep and activity tracker (Oura Health, Oulu, Finland), and found that self-reported anxiety was positively associated with HRV [48]. There was also a high agreement between the Oura Ring PPG-based HRV assessment and ECG (r=0.98) [48].

A study in pregnant women found that they had significantly lower levels of LF if they had obsessive-compulsive disorder, whereas LF was elevated in panic disorders, social anxiety, and social phobias [39]. The LF/HF ratio was also elevated in pregnant women with phobias and panic problems [39]. Furthermore, the LF/HF ratio was elevated in participants who experienced stress after taking a stress-inducing test, most often involving arithmetic, with a few studies finding a strong positive relationship between the LF/HF ratio and stress [43,46,50]. One study specifically focused on anxiety and found a positive relationship between the LF/HF ratio and anxiety based on self-assessments using the State-Trait Anxiety Scale [50]. HR [34,35] and RMSDD increase during stress [46].

Validity

Several studies have found good agreement between ECG-based HRV indices and PPG and anxiety [33,36,48,50]. One study found that PPG had a sensitivity of 83% and a specificity of 93% for assessing anxiety. It was also found that ECG and PPG had perfect agreement (r=1.0) [33]. Another study found a high correlation (r=0.81) between self-reported anxiety and anxiety assessment using PPG [50]. A study on social anxiety disorder found that HR could predict the condition with 88.9% accuracy [36]. The study which evaluated the Oura Ring for anxiety found a high correlation between EEG and PPG findings (r=0.98) [48]. In addition, a study pointed out that PPG signaling is the most accurate when participants are relaxed rather than under stressful conditions [10].

Other Mental Health Problems and Symptoms

One study evaluated the use of PPG to assess the symptoms of suicidal ideation in patients with depression. By examining diastolic and systolic entropy lag scores, the researchers found that PPG has an accuracy of 96.7% in diagnosing suicidal ideation in patients with MDD by examining diastolic and systolic entropy lag scores [38]. The researchers also found that the scores of patients on the Hamilton Depression Scale, which they completed via self-reports, were highly correlated (r=0.96) with the entropy lag scores [38].

Besides suicidal thoughts, a few other studies have evaluated PPG for less prevalent mental health problems, including schizophrenia and borderline personality disorder. One study evaluated the use of PPG in schizophrenic patients and found that patients with schizophrenia had elevated LF/HF ratios compared with controls but reduced HF while resting (P<.05). However, HRV was reduced overall in patients during the stress and subsequent relaxation phases, with lower RMSDD and SD of the N-N intervals than in their counterparts without schizophrenia [31]. In addition, one study evaluated borderline personality disorders by using PPG. The researchers found that patients with a higher resting HR also had improved recovery when evaluated longitudinally over a 2-year period [53].

Study Quality

Most studies were conducted with small sample sizes. Ten out of 24 studies had a small sample size of <50 participants, ranging from 5 [43] to 40 [49] participants. The largest study was carried out in 2012 to assess depression using fingertip PPG–based technology, with a total of 308 participants [51,52]; however, sex (male or female) was not reported. A recent study [45] targeted the same research point, assessing depression using fingertip PPG in 36 participants. The first study on N=308 scored 99.53%, whereas the latter on N=36 achieved 85.3% accuracy.

Only 8 patients had a clinical diagnosis [33,36,42,47,51,52,54]. Only 3 studies compared patients with MDD with those who did not have clinical depression [42,52,54]. One study examined suicidal ideation in patients with MDD and in patients with MDD who did not have suicidal ideation [37]. There were no longitudinal studies. Randomization, allocation concealment, and blinding were not conducted in the interventions with the 2 arms.

Discussion

Principal Findings

This review aimed to evaluate the use of PPG in assessing and screening (including prescreening) mental health problems. We found an association between increased HR and mental health outcomes. The mental health literature indicates that depression and anxiety increase the risk of morbidity and mortality from heart disease [16,17], and the studies in this review confirmed changes in HRV in patients with mental health problems. However, none of the studies specifically used PPG-based technologies to screen patients for mental health problems or to monitor symptoms during treatment.

Although the studies were not equal in terms of quality and sample size, a few studies reported good accuracy in detecting mental health conditions; for example, Long et al [45] achieved an accuracy of 85.3% over 36 participants after using the Patient Health Questionnaire–9 as the gold standard for detecting depression. Ismail et al [36] included 34 participants, using social anxiety disorder as a gold standard, for detecting stress. All the developed detectors rely on changes in PPG-based HRV measurements in patients with depression and anxiety. This includes differences between stressful states or feelings of anxiety and relaxation states. However, we did not find any studies that monitored the progress, including symptoms, in patients undergoing treatment. Future studies should consider evaluating the use of PPG to monitor disease progression. However, it remains unclear whether PPG can distinguish between patients on and off medication or changes in mood without medication use.

PPG was also found to be comparable with ECG in assessing the changes in HRV indices associated with mental health outcomes. PPG was positively correlated with self-reported measures of mental health end points overall in this review. HR was elevated when participants were stressed, and PPG could accurately capture these physiological changes, including HR reduction during relaxing exercises [10,42,43,64,65]. In addition to stress, PPG also accurately detects patients who are depressed [32-34,37,38,40,41,44,45,48,49,52,54], including those with MDD. In other words, depression was associated with a lower HR in contrast to stress. This was the case when studies compared patients with MDD with those without MDD, highlighting the ability of PPG to distinguish between clinically depressive and nondepressive states. We were also able to differentiate between patients with clinical depression who had symptoms of suicidal ideation and those who did not.

These findings are relevant given that patients with clinical depression may be monitored for symptoms of suicidal ideation and may quickly receive emergency medical care when evaluated in remote and rural settings. It is also crucial that PPG can rapidly assess anxiety and depression because patients may be quickly referred to medical doctors and receive a timely diagnosis. In addition, access to mental health care is a problem for many people, especially those living in rural and remote areas [11,59]. Virtual PPG mental health screening and monitoring may enable access to medical care for those who do not typically have access to treatment. PPG can be used as a prescreening tool to screen and prioritize medical accessibility in rural patients [20. It is also important for medical practitioners to obtain HR data to obtain more objective mental health measures than subjective self-reports.

We found that the PPG technology has advanced in recent years. Not only are probe-based methods used, but PPG may also be integrated into smartphone technology, which has important implications for accessibility, given the ubiquitous nature of cell phones. However, we identified only a few studies that used smartphones for PPG, highlighting a gap in the research literature and the need for more studies. It appears to be a practical tool, given that patients can easily access technology. Face-based methods using mobile health to detect changes via facial imaging are simple, noninvasive, and practical for users.

General studies on psychological stress in the literature have also found good agreement between PPG- and ECG-based HRV findings, which is applicable to anxiety studies [60-67]. The accuracy in these studies ranged from 77.6% in smartphone-based (CorSense) PPG technology to 98% using a wrist-based method [64,65]. The correlation also varied according to HRV indices when compared with ECG including RMSDD (r=0.97) and LF/HF ratio (r=0.87) [66]. A PPG wrist-worn technological study found an overall accuracy of 83%, assessing calm states 67% of the time and stressed states 72% of the time [63]. However, one study that examined PPG in the earlobe did not find good agreement with the reference method [61]. Interestingly, a study that used smartphone-based PPG technology and a thermal camera found that the readings were the most accurate when both methods were combined [60]. One study validated their findings by comparing the Huawei PPG watch sensor with not only self-reports but also laboratory measures of cortisol [62]. Another study that used 3 LED lights on the wrist recommended utilizing light green instead of infrared for PPG-based stress detection [64].

Gold Standard

Currently, there is no widely accepted gold standard for assessing depression or anxiety using PPG. Although PPG has shown promise as a tool for measuring physiological arousal levels and has been used in a number of studies to investigate the physiological changes associated with depression and anxiety, its use as a diagnostic tool is still in its early stages of development. Therefore, a gold-standard tool must be used in combination with PPG.

The studies used self-reported questionnaires and interviews to assess mental health. For example, in terms of interviews, Clamor et al [32] used the Structured Mini-International Neuropsychiatric Interview, Khandoker et al [37] used Neuropsychiatric Interview (MINI) version 5, and Sigrist et al [53] used the Structured Clinical Interview for DSM-IV-Axis II. In terms of self-reported questionnaires, the following instruments were used: Four-Dimensional Symptom Questionnaire, Patient Health Questionnaire–9, Trier Social Stress Test, Self-Rating Depression Scale, Hamilton Depression Rating Scale and Beck Depression Inventory II, Edinburgh Postnatal Depression Scale and the State-Trait Anxiety Inventory for Adults, State-Trait Anxiety Inventory, Depression Anxiety Stress Scales, Satisfaction With Life Scale, DSM-5 and

Liebowitz Social Anxiety Scale, Wisconsin Card Sorting Test, Clinicians' Rating of Depression Severity and Beck Depression Inventory II, and Profile of Mood States. These self-report measures can provide valuable information about an individual's subjective experiences of depression or anxiety. A few studies [32,37,53] have used a combination of self-report measures and self-reported questionnaires, along with PPG. Five studies did not use the gold standard method, interviews, or questionnaires [33,40,47,51,52].

Limitations and Ethical Considerations

Despite the outcomes of this review, the results should be interpreted with caution. PPG is an emerging technology that cannot be used as a stand-alone method to assess or monitor mental health. It may also be used as an adjunct along with skilled mental health practitioners. Important factors to consider when diagnosing mental health problems include the severity of symptoms [68] and functional impairment [69], which cannot be detected presently according to the reviewed studies. Furthermore, many studies did not use a comparison group, which is another limitation of this study.

Additionally, it may be difficult to determine whether someone with a mental health problem experiences general life stress or whether their anxiety and stress are directly a result of their mental health condition given the variety in stressful exposures across the studies and similarity in results. Moreover, many studies involved a stressful task that included changes in physiology (getting up after sitting), mental tasks, or stress invoking exposures [42,43,46,47,50,60,64,65]. The studies did not discern between anxiety and anxious states in general. The key challenge is to discern between patients experiencing clinical anxiety, general personal stress, and feelings of anxiety or anxiety common and everyone, not only in clinical populations. Thus, it may also seem that the technology may be most useful in assessing mood changes in patients with depression, as the altered HRV in depression seen in the reviewed studies may be differentiated from stressed or anxious states.

Ethical issues must be considered when implementing PPG technology for mental health applications. Health ethics involves a consideration of harm and benefits [70]. Maximizing the benefits and potential harm is a prima facie ethical duty [70]. This could include the consideration of PPG-associated false positives or negatives when screening for suicidal symptoms or assessing changes in mood states in patients with depression. Thus, clinicians should be mindful of the limitations of this emerging technology and use it prudently along with their clinical practice, backing up findings when possible, through clinical assessment.

Design Bias

More studies with larger sample sizes and in clinical populations are needed before PPG can be used as an adjunct in assessing mental and psychological health. We found several limitations in the studies we analyzed, including the small number of studies in clinical populations and small sample sizes. There is also a need for randomized controlled trials that may follow up patients. There is a gap in the literature, and randomized controlled trials are needed to validate PPG technology. Patients should also be blinded in future studies to reduce the potential for placebo effect and bias when comparing subjective self-report measures with PPG. They could be blinded by providing them with a similar-looking application on their smartphones and by requiring them to look at the camera and place their fingers on their phones.

Ethnic Bias

It is worth noting that a recent study [71] revealed that PPG produces systematically higher saturation values in Black patients than in White patients. Participants with darker skin pigmentation may be at an increased risk of unrecognized health conditions. Therefore, collecting and testing PPG signals from Black participants is required to reduce bias [72]. None of the studies recognized or tested this bias in the assessment of mental health. Therefore, we strongly recommend creating publicly available databases with more balanced ethnicities (ie, participants with different pigmentation colors).

Confounding Bias

Confounds such as hospitalization or the use of pharmacological treatments are essential to control the study design and analysis. None of the included studies reported hospitalization or pharmacological treatments; however, 5 out of 24 studies [31,34,39,41,53] paid attention to possible confounding factors.

In fact, it is essential that all the studies used HR and HRV indices, which can be affected by a wide range of physical and psychological factors [73]. It is also crucial for PPG-based mental health studies to ensure that participants remain in a stable state before data collection. This may involve restricting data collection to a specific time of the day or requiring participants to avoid stressful or physically demanding activities before data collection [74].

Another factor to consider is that other conditions or exposures, such as COVID-19, may affect HRV [75,76]. Thus, clinicians should be mindful of this possibility and rule out COVID-19 or associated comorbidities when using PPG for mental health assessments.

By carefully controlling for these confounding factors and selecting study participants, PPG-based mental health studies can provide less bias about the physiological changes associated with depression and anxiety, helping to advance our understanding of these conditions and the potential of PPG as a diagnostic tool.

PPG-Based Technology Versus Traditional Mental-Health Assessment Tools

None of the articles included in this study discussed the use of PPG versus traditional mental health assessment tools. PPG is generally considered less invasive than answering questions via interviews or questionnaires, depending on the individual user's perspective.

We speculate that answering questions (via interview or filling out a questionnaire) about their mental health may make them feel more personal and intimate than having their physiological arousal levels measured through PPG. They may feel more

comfortable disclosing their thoughts and feelings through self-report measures, rather than monitoring their physiological responses.

PPG has the potential to offer new insights into the mechanisms underlying mental health disorders. PPG also has a considerable advantage as it provides real-time data acquisition, enabling mental health practitioners to monitor changes in physiological arousal levels. This can be particularly useful in situations where quick and effective interventions are needed. One major concern with this advantage is privacy; users may not feel comfortable when their physiological signals are monitored regularly [77]. They may worry about who has access to the data collected through PPG and how the data will be used.

At this stage, it is debatable what would be more effective and satisfy user preferences. The authors feel that this area is worth investigating to improve mental health assessments in the near future.

Conclusions

PPG for mental health assessment is an emerging technology with great potential. Currently, it cannot be used to diagnose mental health and psychological problems; however, future trials may provide sufficient evidence for the use of this technology as an adjunct or complementary tool for mental health assessment alongside conventional methods. With the increasing prevalence of telemedicine and remote patient monitoring, PPG has the potential to provide mental health practitioners with cost-effective and convenient means of monitoring their patients' physiological arousal levels. This can enable earlier detection of mental health issues and enable timely and effective interventions.

Conflicts of Interest

None declared.

References

1. Elgendi M. PPG Signal Analysis An Introduction Using MATLAB®. Boca Raton, Florida, United States: CRC Press; 2020.
2. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Feb 20;28(3):R1-39. [doi: 10.1088/0967-3334/28/3/r01]
3. Elgendi M, Fletcher R, Liang Y, Howard N, Lovell NH, Abbott D, et al. The use of photoplethysmography for assessing hypertension. NPJ Digit Med 2019 Jun 26;2(1):60 [FREE Full text] [doi: 10.1038/s41746-019-0136-7] [Medline: 31388564]
4. Frey L, Menon C, Elgendi M. Blood pressure measurement using only a smartphone. NPJ Digit Med 2022 Jul 06;5(1):86 [FREE Full text] [doi: 10.1038/s41746-022-00629-2] [Medline: 35794240]
5. Haugg F, Elgendi M, Menon C. Effectiveness of remote PPG construction methods: a preliminary analysis. Bioengineering (Basel) 2022 Sep 20;9(10):485 [FREE Full text] [doi: 10.3390/bioengineering9100485] [Medline: 36290452]
6. Allen J, Zheng D, Kyriacou P, Elgendi M. Photoplethysmography (PPG): state-of-the-art methods and applications. Physiol Meas 2021 Nov 24;42(10). [doi: 10.1088/1361-6579/ac2d82] [Medline: 34842179]
7. Elgendi M. On the analysis of fingertip photoplethysmogram signals. Curr Cardiol Rev 2012 Feb;8(1):14-25 [FREE Full text] [doi: 10.2174/157340312801215782] [Medline: 22845812]
8. Vescio B, Salsone M, Gambardella A, Quattrone A. Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating heart rate variability in healthy subjects in short- and long-term recordings. Sensors (Basel) 2018 Mar 13;18(3):844 [FREE Full text] [doi: 10.3390/s18030844] [Medline: 29533990]
9. De Ridder B, Van Rompaey B, Kampen JK, Haine S, Dilles T. Smartphone apps using photoplethysmography for heart rate monitoring: meta-analysis. JMIR Cardio 2018 Feb 27;2(1):e4 [FREE Full text] [doi: 10.2196/cardio.8802] [Medline: 31758768]
10. Celka P, Charlton PH, Farukh B, Chowienczyk P, Alastruey J. Influence of mental stress on the pulse wave features of photoplethysmograms. Healthc Technol Lett 2020 Feb 26;7(1):7-12 [FREE Full text] [doi: 10.1049/htl.2019.0001] [Medline: 32190335]
11. Morales DA, Barksdale CL, Beckel-Mitchener AC. A call to action to address rural mental health disparities. J Clin Transl Sci 2020 May 04;4(5):463-467 [FREE Full text] [doi: 10.1017/cts.2020.42] [Medline: 33244437]
12. Galea S, Merchant RM, Lurie N. The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. JAMA Intern Med 2020 Jun 01;180(6):817-818. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]
13. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
14. Barberis N, Cannavò M, Cuzzocrea F, Verrastro V. Suicidal behaviours during covid-19 pandemic: a review. Clin Neuropsychiatry 2022 Apr;19(2):84-96 [FREE Full text] [doi: 10.36131/cnfioritieditore20220202] [Medline: 35601250]
15. Costanza A, Amerio A, Aguglia A, Serafini G, Amore M, Macchiarulo E, et al. From "The Interpersonal Theory of Suicide" to "The Interpersonal Trust": an unexpected and effective resource to mitigate economic crisis-related suicide risk in times

of Covid-19? Acta Biomed 2021 Oct 01;92(S6):e2021417 [FREE Full text] [doi: 10.23750/abm.v92iS6.12249] [Medline: 34739460]
16. Olafiranye O, Jean-Louis G, Zizi F, Nunes J, Vincent M. Anxiety and cardiovascular risk: review of epidemiological and clinical evidence. Mind Brain 2011 Aug;2(1):32-37 [FREE Full text] [Medline: 21822473]
17. Saran R, Puri A, Agarwal M. Depression and the heart. Indian Heart J 2012 Jul;64(4):397-401 [FREE Full text] [doi: 10.1016/j.ihj.2012.06.004] [Medline: 22929824]
18. Rubio-Guerra AF, Rodriguez-Lopez L, Vargas-Ayala G, Huerta-Ramirez S, Serna DC, Lozano-Nuevo JJ. Depression increases the risk for uncontrolled hypertension. Exp Clin Cardiol 2013;18(1):10-12 [FREE Full text] [Medline: 24294029]
19. Pan Y, Cai W, Cheng Q, Dong W, An T, Yan J. Association between anxiety and hypertension: a systematic review and meta-analysis of epidemiological studies. Neuropsychiatr Dis Treat 2015;11:1121-1130 [FREE Full text] [doi: 10.2147/NDT.S77710] [Medline: 25960656]
20. Langarizadeh M, Tabatabaei MS, Tavakol K, Naghipour M, Rostami A, Moghbeli F. Telemental health care, an effective alternative to conventional mental care: a systematic review. Acta Inform Med 2017 Dec;25(4):240-246 [FREE Full text] [doi: 10.5455/aim.2017.25.240-246] [Medline: 29284913]
21. Aboraya A, Nasrallah H, Muvvala S, El-Missiry A, Mansour H, Hill C, et al. The standard for clinicians' interview in psychiatry (SCIP): a clinician-administered tool with categorical, dimensional, and numeric output-conceptual development, design, and description of the. Innov Clin Neurosci 2016;13(5-6):31-77 [FREE Full text] [Medline: 27800284]
22. Hoffmann M, Wensing M, Peters-Klimm F, Szecsenyi J, Hartmann M, Friederich H, et al. Perspectives of psychotherapists and psychiatrists on mental health care integration within primary care via video consultations: qualitative preimplementation study. J Med Internet Res 2020 Jun 18;22(6):e17569 [FREE Full text] [doi: 10.2196/17569] [Medline: 32554369]
23. Regier DA, Kuhl EA, Kupfer DJ. The DSM-5: classification and criteria changes. World Psychiatry 2013 Jun 04;12(2):92-98 [FREE Full text] [doi: 10.1002/wps.20050] [Medline: 23737408]
24. Santomauro DF, Mantilla Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov;398(10312):1700-1712. [doi: 10.1016/s0140-6736(21)02143-7]
25. Liu Z, Liu R, Zhang Y, Zhang R, Liang L, Wang Y, et al. Association between perceived stress and depression among medical students during the outbreak of COVID-19: the mediating role of insomnia. J Affect Disord 2021 Sep 01;292:89-94 [FREE Full text] [doi: 10.1016/j.jad.2021.05.028] [Medline: 34107425]
26. Bystritsky A, Khalsa SS, Cameron ME, Schiffman J. Current diagnosis and treatment of anxiety disorders. P T 2013 Jan;38(1):30-57 [FREE Full text] [Medline: 23599668]
27. Zorn JV, Schür RR, Boks MP, Kahn RS, Joëls M, Vinkers CH. Cortisol stress reactivity across psychiatric disorders: a systematic review and meta-analysis. Psychoneuroendocrinology 2017 Mar;77:25-36. [doi: 10.1016/j.psyneuen.2016.11.036] [Medline: 28012291]
28. Malinauskas R, Malinauskiene V. Meta-analysis of psychological interventions for reducing stress, anxiety, and depression among university students during the COVID-19 pandemic. Int J Environ Res Public Health 2022 Jul 27;19(15):9199. [doi: 10.3390/ijerph19159199] [Medline: 35954553]
29. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
30. Kavanagh BP. The GRADE system for rating clinical guidelines. PLoS Med 2009 Sep 15;6(9):e1000094 [FREE Full text] [doi: 10.1371/journal.pmed.1000094] [Medline: 19753107]
31. Akar SA, Kara S, Latifoğlu F, Bilgiç V. Analysis of heart rate variability during auditory stimulation periods in patients with schizophrenia. J Clin Monit Comput 2015 Feb 16;29(1):153-162. [doi: 10.1007/s10877-014-9580-8] [Medline: 24831932]
32. Clamor A, Hartmann MM, Köther U, Otte C, Moritz S, Lincoln TM. Altered autonomic arousal in psychosis: an analysis of vulnerability and specificity. Schizophr Res 2014 Apr;154(1-3):73-78. [doi: 10.1016/j.schres.2014.02.006] [Medline: 24582038]
33. Dagdanpurev S, Sun G, Shinba T, Kobayashi M, Kariya N, Choimaa L, et al. Development and clinical application of a novel autonomic transient response-based screening system for major depressive disorder using a fingertip photoplethysmographic sensor. Front Bioeng Biotechnol 2018 May 22;6:64 [FREE Full text] [doi: 10.3389/fbioe.2018.00064] [Medline: 29872656]
34. de Vries H, Kamphuis W, van der Schans C, Sanderman R, Oldenhuis H. Trends in daily heart rate variability fluctuations are associated with longitudinal changes in stress and somatisation in police officers. Healthcare (Basel) 2022 Jan 12;10(1):144 [FREE Full text] [doi: 10.3390/healthcare10010144] [Medline: 35052307]
35. Harvie HM, Jain B, Nelson BW, Knight EL, Roos LE, Giuliano RJ. Induction of acute stress through an internet-delivered trier social stress test as assessed by photoplethysmography on a smartphone. Stress 2021 Nov 02;24(6):1023-1032. [doi: 10.1080/10253890.2021.1995714] [Medline: 34726560]

36. Ismail NM, Airij AG, Sudirman R, Omar C. Early detection of social anxiety disorder by using screening tools and wearable sensors. In: Proceedings of the 6th International Conference on Computing Engineering and Design (ICCED). 2020 Presented at: 6th International Conference on Computing Engineering and Design (ICCED); Oct 15-16, 2020; Sukabumi, Indonesia.
37. Khandoker AH, Luthra V, Abouallaban Y, Saha S, Ahmed KI, Mostafa R, et al. Identifying depressed patients with and without suicidal ideation by finger photo-plethysmography. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:1842-1845. [doi: 10.1109/EMBC.2016.7591078] [Medline: 28268684]
38. Khandoker AH, Luthra V, Abouallaban Y, Saha S, Ahmed KI, Mostafa R, et al. Suicidal ideation is associated with altered variability of fingertip photo-plethysmogram signal in depressed patients. Front Physiol 2017 Jul 19;8:501 [FREE Full text] [doi: 10.3389/fphys.2017.00501] [Medline: 28769817]
39. Kimmel MC, Fransson E, Cunningham JL, Brann E, Grewen K, Boschiero D, et al. Heart rate variability in late pregnancy: exploration of distinctive patterns in relation to maternal mental health. Transl Psychiatry 2021 May 14;11(1):286 [FREE Full text] [doi: 10.1038/s41398-021-01401-y] [Medline: 33986246]
40. Kobayashi M, Sun G, Shinba T, Matsui T, Kirimoto T. Simple and objective screening of major depressive disorder by heart rate variability analysis during paced respiration and mental task conditions. Annu Int Conf IEEE Eng Med Biol Soc 2017 Jul;2017:1316-1319. [doi: 10.1109/EMBC.2017.8037074] [Medline: 29060118]
41. Koenig J, Westlund Schreiner M, Klimes-Dougan B, Ubani B, Mueller B, Kaess M, et al. Brain structural thickness and resting state autonomic function in adolescents with major depression. Soc Cogn Affect Neurosci 2018 Sep 04;13(7):741-753 [FREE Full text] [doi: 10.1093/scan/nsy046] [Medline: 29939340]
42. Kontaxis S, Gil E, Marozas V, Lazaro J, Garcia E, Posadas-de Miguel M, et al. Photoplethysmographic waveform analysis for autonomic reactivity assessment in depression. IEEE Trans Biomed Eng 2021 Apr;68(4):1273-1281. [doi: 10.1109/tbme.2020.3025908]
43. Lee Y, Shieh V, Lin C, Shiah Y. A stress evaluation and personal relaxation system based on measurement of photoplethysmography. In: Proceedings of the 2013 Second International Conference on Robot, Vision and Signal Processing. 2013 Presented at: 2013 Second International Conference on Robot, Vision and Signal Processing; Dec 10-12 , 2013; Kitakyushu, Japan.
44. Liu I, Ni S, Peng K. Happiness at your fingertips: assessing mental health with smartphone photoplethysmogram-based heart rate variability analysis. Telemed J E Health 2020 Dec 01;26(12):1483-1491. [doi: 10.1089/tmj.2019.0283] [Medline: 32101084]
45. Long Y, Lin Y, Zhang Z, Jiang R, Wang Z. Objective assessment of depression using multiple physiological signals. In: Proceedings of the 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 2021 Presented at: 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); Oct 23-25 , 2021; Shanghai, China.
46. Minakuchi E, Ohnishi E, Ohnishi J, Sakamoto S, Hori M, Motomura M, et al. Evaluation of mental stress by physiological indices derived from finger plethysmography. J Physiol Anthropol 2013 Oct 12;32(1):17 [FREE Full text] [doi: 10.1186/1880-6805-32-17] [Medline: 24119254]
47. Mitsuhashi R, Iuchi K, Goto T, Matsubara A, Hirayama T, Hashizume H. Video-based stress level measurement using imaging photoplethysmography. In: Proceedings of the 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). 2019 Presented at: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW); Jul 08 - 12 , 2019; Shanghai, China. [doi: 10.1109/icmew.2019.0-105]
48. Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. Front Psychiatry 2021;12:625247 [FREE Full text] [doi: 10.3389/fpsyt.2021.625247] [Medline: 33584388]
49. Ngampramuan S, Tungtong P, Mukda S, Jariyavilas A, Sakulisariyaporn C. Evaluation of autonomic nervous system, saliva cortisol levels, and cognitive function in major depressive disorder patients. Depress Res Treat 2018;2018:7343592-7343598 [FREE Full text] [doi: 10.1155/2018/7343592] [Medline: 29808118]
50. Perpetuini D, Chiarelli AM, Cardone D, Filippini C, Rinella S, Massimino S, et al. Prediction of state anxiety by machine learning applied to photoplethysmography data. PeerJ 2021;9:e10448 [FREE Full text] [doi: 10.7717/peerj.10448] [Medline: 33520434]
51. Pham T, Thang T, Oyama-Higa M. Toward the development of a cost-effective e-depression detection system. In: Proceedings of the 2012 International Conference on Computerized Healthcare (ICCH). 2012 Presented at: 2012 International Conference on Computerized Healthcare (ICCH); Dec 17 - 18 , 2012; Hong Kong, China.
52. Pham TD, Thang TC, Oyama-Higa M, Sugiyama M. Mental-disorder detection using chaos and nonlinear dynamical analysis of photoplethysmographic signals. Chaos Solitons Fractals 2013 Jun;51:64-74. [doi: 10.1016/j.chaos.2013.03.010]
53. Sigrist C, Reichl C, Schmidt SJ, Brunner R, Kaess M, Koenig J. Cardiac autonomic functioning and clinical outcome in adolescent borderline personality disorder over two years. Prog Neuropsychopharmacol Biol Psychiatry 2021 Dec 20;111:110336. [doi: 10.1016/j.pnpbp.2021.110336] [Medline: 33915219]
54. Unursaikhan B, Tanaka N, Sun G, Watanabe S, Yoshii M, Funahashi K, et al. Development of a novel web camera-based contact-free major depressive disorder screening system using autonomic nervous responses induced by a mental task and

its clinical application. Front Physiol 2021 May 14;12:642986 [FREE Full text] [doi: 10.3389/fphys.2021.642986] [Medline: 34054567]
55. Gustafson LW, Gabel P, Hammer A, Lauridsen HH, Petersen LK, Andersen B, et al. Validity and reliability of state-trait anxiety inventory in Danish women aged 45 years and older with abnormal cervical screening results. BMC Med Res Methodol 2020 Apr 23;20(1):89 [FREE Full text] [doi: 10.1186/s12874-020-00982-4] [Medline: 32326886]
56. Argyropoulos SV, Ploubidis GB, Wright TS, Palm ME, Hood SD, Nash JR, et al. Development and validation of the Generalized Anxiety Disorder Inventory (GADI). J Psychopharmacol 2007 Mar 16;21(2):145-152. [doi: 10.1177/0269881107069944] [Medline: 17329293]
57. González-Rivera JA, Pagán-Torres OM, Pérez-Torres EM. Depression, anxiety and stress scales (DASS-21): construct validity problem in hispanics. Eur J Investig Health Psychol Educ 2020 Jan 08;10(1):375-389 [FREE Full text] [doi: 10.3390/ejihpe10010028] [Medline: 34542491]
58. Parkitny L, McAuley J. The depression anxiety stress scale (DASS). J Physiother 2010;56(3):204 [FREE Full text] [doi: 10.1016/s1836-9553(10)70030-8] [Medline: 20795931]
59. Pass LE, Kennelty K, Carter BL. Self-identified barriers to rural mental health services in Iowa by older adults with multiple comorbidities: qualitative interview study. BMJ Open 2019 Nov 03;9(11):e029976 [FREE Full text] [doi: 10.1136/bmjopen-2019-029976] [Medline: 31685497]
60. Cho Y, Julier SJ, Bianchi-Berthouze N. Instant stress: detection of perceived mental stress through smartphone photoplethysmography and thermal imaging. JMIR Ment Health 2019 Apr 09;6(4):e10140 [FREE Full text] [doi: 10.2196/10140] [Medline: 30964440]
61. Correia B, Dias N, Costa P, Pêgo JM. Validation of a wireless bluetooth photoplethysmography sensor used on the earlobe for monitoring heart rate variability features during a stress-inducing mental task in healthy individuals. Sensors (Basel) 2020 Jul 13;20(14):3905 [FREE Full text] [doi: 10.3390/s20143905] [Medline: 32668810]
62. Li F, Xu P, Zheng S, Chen W, Yan Y, Lu S, et al. Photoplethysmography based psychological stress detection with pulse rate variability feature differences and elastic net. Int J Distributed Sensor Netw 2018 Sep 27;14(9):155014771880329. [doi: 10.1177/1550147718803298]
63. Zangróniz R, Martínez-Rodrigo A, López M, Pastor J, Fernández-Caballero A. Estimation of mental distress from photoplethysmography. Applied Sci 2018 Jan 05;8(1):69. [doi: 10.3390/app8010069]
64. Chen C, Li C, Tsai CW, Deng X. Evaluation of mental stress and heart rate variability derived from wrist-based photoplethysmography. In: Proceedings of the 2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS). 2019 Presented at: 2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS); May 31 - June 3 , 2019; Okinawa, Japan.
65. Panganiban F, Leon F. Stress detection using smartphone extracted photoplethysmography. In: Proceedings of the 2021 IEEE Region 10 Symposium (TENSYMP). 2021 Presented at: 2021 IEEE Region 10 Symposium (TENSYMP); Aug 23 - 25 , 2021; Jeju, Korea.
66. Salleh N. Heart rate variability derived from wrist photoplethysmography sensor for mental stress assessment. J Tomography Syst Sensor Appl 2019;2(2):7-11.
67. Yoo K, Lee W. Mental stress assessment based on pulse photoplethysmography. In: Proceedings of the 2011 IEEE 15th International Symposium on Consumer Electronics (ISCE). 2011 Presented at: 2011 IEEE 15th International Symposium on Consumer Electronics (ISCE); Jun 14 - 17 , 2011; Singapore.
68. Zimmerman M, Morgan TA, Stanton K. The severity of psychiatric disorders. World Psychiatry 2018 Oct 07;17(3):258-275 [FREE Full text] [doi: 10.1002/wps.20569] [Medline: 30192110]
69. Mauksch LB, Tucker SM, Katon WJ, Russo J, Cameron J, Walker E, et al. Mental illness, functional impairment, and patient preferences for collaborative care in an uninsured, primary care population. J Fam Pract 2001 Jan;50(1):41-47. [Medline: 11195480]
70. Varkey B. Principles of clinical ethics and their application to practice. Med Princ Pract 2021;30(1):17-28 [FREE Full text] [doi: 10.1159/000509119] [Medline: 32498071]
71. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Engl J Med 2020 Dec 17;383(25):2477-2478 [FREE Full text] [doi: 10.1056/NEJMc2029240] [Medline: 33326721]
72. Sinaki F, Ward R, Abbott D, Allen J, Fletcher R, Menon C, et al. Ethnic disparities in publicly-available pulse oximetry databases. Commun Med (Lond) 2022;2:59 [FREE Full text] [doi: 10.1038/s43856-022-00121-8] [Medline: 35637660]
73. Fatisson J, Oswald V, Lalonde F. Influence diagram of physiological and environmental factors affecting heart rate variability: an extended literature overview. Heart Int 2016 Sep 16;11(1):e32-e40 [FREE Full text] [doi: 10.5301/heartint.5000232] [Medline: 27924215]
74. Catai AM, Pastre CM, Godoy M, Silva E, Takahashi A, Vanderlei LCM. Heart rate variability: are you using it properly? Standardisation checklist of procedures. Braz J Phys Ther 2020 Mar;24(2):91-102 [FREE Full text] [doi: 10.1016/j.bjpt.2019.02.006] [Medline: 30852243]
75. Mol MB, Strous MT, van Osch FH, Vogelaar FJ, Barten DG, Farchi M, et al. Heart-rate-variability (HRV), predicts outcomes in COVID-19. PLoS One 2021 Oct 28;16(10):e0258841 [FREE Full text] [doi: 10.1371/journal.pone.0258841] [Medline: 34710127]

76. Asarcikli LD, Hayiroglu M, Osken A, Keskin K, Kolak Z, Aksu T. Heart rate variability and cardiac autonomic functions in post-COVID period. J Interv Card Electrophysiol 2022 Apr 01;63(3):715-721 [FREE Full text] [doi: 10.1007/s10840-022-01138-8] [Medline: 35106678]
77. Fairclough S. Physiological data must remain confidential. Nature 2014 Jan 16;505(7483):263. [doi: 10.1038/505263a] [Medline: 24429594]

Abbreviations

DSM-5: fifth edition of the Diagnostic and Statistical Manual of Mental Disorders
ECG: electrocardiogram
HF: high frequency
HR: heart rate
HRV: heart rate variability
LF: low frequency
MeSH: Medical Subject Headings
MDD: major depressive disorder
PPG: photoplethysmography
PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols
RMSDD: root mean square of beat intervals

Edited by J Torous; submitted 08.06.22; peer-reviewed by P Batterham, A Costanza; comments to author 04.10.22; revised version received 03.12.22; accepted 06.02.23; published 29.05.23.

Please cite as:
Lyzwinski LN, Elgendi M, Menon C
The Use of Photoplethysmography in the Assessment of Mental Health: Scoping Review
JMIR Ment Health 2023;10:e40163
URL: https://mental.jmir.org/2023/1/e40163
doi:10.2196/40163
PMID:37247209

Review

Psychiatric Treatment Conducted via Telemedicine Versus In-Person Modality in Posttraumatic Stress Disorder, Mood Disorders, and Anxiety Disorders: Systematic Review and Meta-Analysis

Ali Abbas Shaker[1,2], BEng, MD; Stephen F Austin[1,3], MSc, PhD; Ole Jakob Storebø[1,3], MSc, PhD; Julie Perrine Schaug[1], MSc; Alaa Ayad[1], MSc; John Aasted Sørensen[4], MSc, PhD; Kristine Tarp[5,6], MSc, PhD; Henrik Bechmann[4], MSc, PhD; Erik Simonsen[2,7], PhD, MD

[1]Psychiatric Department, Region Zealand Psychiatry, Psychiatric Research Unit, Slagelse, Denmark

[2]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

[3]Department of Psychology, University of Southern Denmark Faculty of Health Sciences, Odense, Denmark

[4]Department of Engineering Technology and Didactics, Research unit: AI, Mathematics and Software, Technical University of Denmark, Ballerup, Denmark

[5]Research Unit for Digital Psychiatry, Mental Health Services in the Region of Southern Denmark, Odense, Denmark

[6]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

[7]Mental Health Services East, Copenhagen University Hospital – Psychiatry Region Zealand, Roskilde, Denmark

Corresponding Author:
Ali Abbas Shaker, BEng, MD
Psychiatric Department, Region Zealand Psychiatry, Psychiatric Research Unit
Fælledvej 6
Slagelse, 4200
Denmark
Phone: 45 58536070
Email: ashak@regionsjaelland.dk

Related Article:

This is a corrected version. See correction statement: https://mental.jmir.org/2023/1/e52269

Abstract

Background: Telemedicine has played a vital role in providing psychiatric treatment to patients during the rapid transition of services during the COVID-19 pandemic. Furthermore, the use of telemedicine is expected to expand within the psychiatric field. The efficacy of telemedicine is well described in scientific literature. However, there is a need for a comprehensive quantitative review that analyzes and considers the different clinical outcomes and psychiatric diagnoses.

Objective: This paper aimed to assess whether individual psychiatric outpatient treatment for posttraumatic stress disorder, mood disorders, and anxiety disorders in adults using telemedicine is equivalent to in-person treatment.

Methods: A systematic search of randomized controlled trials was conducted using recognized databases for this review. Overall, 4 outcomes were assessed: treatment efficacy, levels of patient satisfaction, working alliance, and attrition rate. The inverse-variance method was used to summarize the effect size for each outcome.

Results: A total of 7414 records were identified, and 20 trials were included in the systematic review and meta-analysis. The trials included posttraumatic stress disorder (9 trials), depressive disorder (6 trials), a mix of different disorders (4 trials), and general anxiety disorder (1 trial). Overall, the analyses yielded evidence that telemedicine is comparable with in-person treatment regarding treatment efficacy (standardized mean difference −0.01, 95% CI −0.12 to 0.09; P=.84; I^2=19%, 17 trials, n=1814), patient satisfaction mean difference (−0.66, 95% CI −1.60 to 0.28; P=.17; I^2=44%, 6 trials, n=591), and attrition rates (risk ratio 1.07, 95% CI 0.94-1.21; P=.32; I^2=0%, 20 trials, n=2804). The results also indicated that the working alliance between telemedicine

and in-person modalities was comparable, but the heterogeneity was substantial to considerable (mean difference 0.95, 95% CI −0.47 to 2.38; P=.19; I^2=75%, 6 trials, n=539).

Conclusions: This meta-analysis provided new knowledge on individual telemedicine interventions that were considered equivalent to in-person treatment regarding efficacy, patient satisfaction, working alliance, and attrition rates across diagnoses. The certainty of the evidence regarding efficacy was rated as moderate. Furthermore, high-quality randomized controlled trials are needed to strengthen the evidence base for treatment provided via telemedicine in psychiatry, particularly for personality disorders and a range of anxiety disorders where there is a lack of studies. Individual patient data meta-analysis is suggested for future studies to personalize telemedicine.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42021256357; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=256357

(JMIR Ment Health 2023;10:e44790) doi:10.2196/44790

KEYWORDS

telemedicine; telepsychiatry; video consultation; mobile health; mHealth; eHealth; COVID-19; synchronous technology; anxiety; psychiatry; patient satisfaction; depression; posttraumatic stress disorder; PTSD

Introduction

Background

During the last 2 decades, there has been increasing interest in and publication of research studies addressing the effect of telemedicine on psychiatric patients [1-5]. Research has highlighted several potential advantages of using telemedicine in mental health services. Some of the most apparent advantages of integrating telemedicine in mental health services are enabling clinicians to reach out to patients living in rural areas and patients with mental health difficulties who find it challenging to attend treatment in person [1,2]. However, several studies have also reported challenges regarding the use of telemedicine for mental health services, including concerns about establishing a good patient-therapist alliance and the underuse of telemedicine by clinicians in resource-constrained clinics [3-5].

The experience of COVID-19 has placed an increased focus on the provision of interventions using telemedicine. This unique world situation, coupled with continual advances in technology, means that a regular synthesis of evidence for psychiatric interventions using telemedicine is warranted [6-10].

In recent years, several meta-analyses have compared the efficacy of psychiatric treatment provided using telemedicine with in-person treatment [11-13]. Drago et al [11] reviewed the evidence of psychiatric counseling (but not specific psychiatric or psychotherapeutic interventions) using telemedicine compared with in-person treatment. They included 24 randomized controlled trials (RCTs) primarily for posttraumatic stress disorder (PTSD) and major depression and found no difference in treatment effects between the 2 modalities. Their review did not examine the satisfaction, alliance, or attrition rates between the 2 modes of treatment. Batastini et al [12] conducted a large meta-analysis with broad inclusion criteria and combined data from a variety of study designs (RCTs and within subjects), reported outcomes (observer rated and self-report), and treatment format (individual and group). Their analysis included 43 studies and found that treatment effects were largely comparable between telemedicine and in-person modalities; however, they did not examine satisfaction, alliance, or attrition rates between the 2 modes of treatment. Giovanetti et al [13] focused exclusively on comparing psychotherapeutic interventions for depression. They included 11 RCTs and found that telemedicine-based psychotherapy had comparable efficacy with in-person psychotherapy [13]. They also found no differences in the attrition rates between the 2 modalities in patients diagnosed with depression.

Objective

This meta-analysis builds on the results of previous reviews by addressing some of the deficiencies of earlier meta-analyses and providing a comprehensive and updated overview of the evidence for telemedicine in psychiatric settings. Thus, the primary aim of this systematic review and meta-analysis is to examine whether individual psychiatric outpatient interventions for adults using telemedicine are equivalent to the in-person format regarding treatment efficacy. As part of this comprehensive meta-analysis comparing the treatment effects between telemedicine and in person, we examined different diagnostic disorders and analyzed a range of moderators. Second, the meta-analysis addressed several gaps in current scientific research using standard and valid measures to examine the satisfaction, working alliance, and attrition rates between telemedicine and in-person modalities across a range of psychiatric diagnoses.

Methods

Overview

The methods section of this systematic review and meta-analysis is described in a published peer-reviewed protocol [14]. This systematic review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15]. The PRISMA checklist can be found on the web (Multimedia Appendix 1).

Registration and Protocol

This systematic review was registered in PROSPERO (CRD42021256357). A peer-reviewed protocol has been published for this study. First, a post hoc analysis was conducted to evaluate attrition outcomes based on diagnosis (PTSD and Depression). Second, subgroup analysis for the moderators, “settings,” and “vulnerable populations” were poorly described

in the included studies, and subgroup analysis for these moderators was therefore not applicable. Third, the title names have been adjusted. Apart from the listed amendments, no significant amendments were made compared with the published protocol.

Inclusion Criteria

The eligibility criteria were based and restricted on the type of study, population, intervention, comparator, and outcomes of the studies.

Types of Studies

RCTs were considered.

Types of Participants

The participants were (1) adults (aged >18 years), (2) receiving individual psychiatric outpatient treatment, and (3) diagnosed with PTSD, mood disorders, anxiety, or personality disorders according to both the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders III-V and the World Health Organization's International Statistical Classification of Diseases 9 or 10. Participants with comorbid diagnoses were also included, with the exception of those diagnoses covered in the exclusion criteria.

Types of Interventions

Individual treatment through synchronous real-time video consultations in outpatient settings. Treatment was defined as an intervention that involved psychotherapy, pharmacological treatment, or psychoeducation.

Types of Comparators or Controls

The comparator was individual treatment in person with the same active treatment that the intervention group (telemedicine) received.

Types of Outcomes

The primary outcome was studies that assessed psychopathology (efficacy) after using a mental health service. The secondary outcomes of interest were (1) patient satisfaction, (2) working alliance, and (3) attrition rate.

Exclusion Criteria

The following criteria were considered the reason for exclusion:

- Participant aged <18 years
- Group therapy
- Different psychotherapeutic (treatments) approaches used in the telemedicine and in-person modalities
- Trials involving populations primarily treating psychotic disorders, mental retardation, bipolar disorder, alcohol abuse, and substance use disorders
- Trials using asynchronous communication systems as an intervention (eg, emails and static websites without video function) and telephones with only audio function as an intervention

Information Sources and Search Strategy

The first step in the systematic review was a comprehensive search in electronic databases. The database search strategy was developed using input from the project team. A search was conducted for studies published between 1967 and October 2022.

The following databases were used: MEDLINE (PubMed interface, 1986 onward), APA PsycINFO (Ovid interface, 1967 onward), Embase (Ovid interface, 1974 onward), Web of Science (Clarivate interface, 2001 onward), and CINAHL (EBSCOhost interface, 1981 onward).

Medical Subject Headings (MeSH) and text words related to the search terms "psychiatry" and "telemedicine" were used to develop the search string in MEDLINE.

Examples of MeSH and text words related to the term "psychiatry" included: ("Psychiatry"[MeSH Terms] or "Mental Disorders"[MeSH Terms] or "Mental Health Services"[MeSH Terms]) and ("mental health counseling"[Title/Abstract] or "mental health care"[Title/Abstract] or "psychiatric home care"[Title/Abstract] or "psychiatric outpatient*"[Title/Abstract]).

Examples of MeSH and text words related to the term "telemedicine" included: ("Telemedicine"[MeSH Terms] or "Videoconferencing"[MeSH Terms] or "Remote Consultation"[MeSH Terms] and "telecare"[Title/Abstract] or "teleconsultation*"[Title/Abstract] or "telemedic*"[Title/Abstract] or "telepsychiatr*") Both search terms "psychiatry" and "telemedicine" were combined with (AND).

Specific syntax and subject headings were subsequently adapted individually to the different databases.

No language or date restrictions were implemented in the search process. Owing to the preliminary search's unmanageable results (>20.000 hits), the highly sensitive search strategy filters of Cochrane identifying randomized trials were applied. Unpublished studies were not sought.

The second step in the search strategy was a manual literature search to identify additional primary studies for systematic review. The third step involved scanning the reference lists of the included studies or relevant reviews identified in the first and second steps, respectively.

Data Management

Records from the literature search were exported to the reference manager Endnote X9 (Clarivate Analytics) [16]. From Endnote, records were exported to Covidence (Veritas Health Innovation), a web app tool that facilitates collaboration among review team members during the study selection and data extraction process [17]. Data extracted in Covidence were exported to RevMan 5.4 (Cochrane Collaboration) for data analysis [18].

Selection Process

AAS and AA were responsible for the selection process. The 2 authors independently screened the titles and abstracts of the records in Covidence to identify potentially eligible records. The second step involved screening full-text reports to assess whether the reports met the eligibility criteria. Three authors (pairwise) were responsible for the second step (AAS, SFA, or JPS). Disagreements in the full-text screening process were resolved through discussions between the authors. A fourth

reviewer, OJS, was consulted in case of continued disagreement despite discussion. The selection process was documented in the PRISMA flow diagram, including reasons for exclusion. Interrater reliability was measured using Cohen κ coefficient for the title and abstract screening and full-text review processes.

Data Collection Process

AAS, SFA, and JPS were responsible for the data collection process. Data extraction was performed independently by 2 authors using a standardized electronic data extraction form in Covidence. The data extraction form was pilot tested on 5 reports, and the reviewers met and discussed the form before starting the review. Disagreements in the data collection process were resolved through discussions between the authors. A third reviewer (OJS) was consulted when disagreements could not be resolved between the independent authors. If multiple reports of the same study were encountered, data from all reports were extracted into a single data collection form in Covidence [19]. Missing data were obtained by contacting and requesting these data from the study authors.

Data Items

We extracted the following data items for each study: (1) study characteristics (authors, author contact details, aim of the study, trial design, location, trial size, sample size calculation, year of publication, and country); (2) population characteristics (remote or rural area or urban, country, diagnosis or condition, mean age, and sex); (3) intervention or control (internet connection speed, bandwidth, therapy type, number of consultation sessions, and duration of consultation); and (4) clinical outcome (assessment tools, psychopathology [efficacy outcome], patient satisfaction, working alliance, and attrition rate). When reported in the studies, we collected data from the intention-to-treat analysis; otherwise, we collected data from the per-protocol analysis.

Outcomes and Prioritization

The primary outcome was efficacy, as assessed by clinician or patient-rated scales. As we expected that different assessment tools had been used for measuring the primary outcome, we prioritized clinician-rated scales over patient-rated scales, should both be available.

The secondary outcomes were (1) patient satisfaction, (2) working alliance, and (3) attrition rate. The patient satisfaction measure was restricted to the Client Satisfaction Questionnaire-8 (CSQ-8) [20], and the working alliance was restricted to the Working Alliance Inventory-Client version (WAI-C) [21]. The attrition rate was defined as the proportion of individuals who withdrew after being randomized to a modality to the total number of participants randomized to a modality.

Risk of Bias in Individual Studies

Authors AAS, SFA, and JPS performed (pairwise) the risk of bias (quality) assessment for the primary outcome (treatment efficacy) in each individual study, using the revised Cochrane risk-of-bias tool for randomized trials (RoB 2) [19]. The bias domains assessed included (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in the measurement of the outcome, and (5) bias in the selection of the reported result. The overall risk of bias for each study was marked as (1) "low risk of bias" if all domains were judged to be at low risk of bias, (2) "some concerns" if at least one domain was judged to raise some concerns but not to be at high risk of bias for any domain, or (3) "high risk of bias" if any domain was judged to be at a high risk of bias. Disagreements between the mentioned researchers regarding the risk of bias were resolved through consensus or by a third researcher (OJS). The Covidence tool was used to assess the risk of bias.

Data Synthesis (Statistical Methods)

The general strategy for data synthesis was to perform a quantitative synthesis (meta-analysis). Heterogeneity (I^2) values were judged as follows: 0%-40% may represent little heterogeneity, 30%-60% may represent moderate heterogeneity, 50%-90% may represent substantial heterogeneity, and 75%-100% may represent considerable heterogeneity. Heterogeneity, which is the percentage of variation across studies owing to heterogeneity rather than chance, was evaluated for clinical, methodological, and statistical heterogeneity [19].

Quantitative Synthesis

We expected clinical and methodological heterogeneity in the pooled studies; therefore, we applied a random effects model to obtain the overall effect size estimate. The inverse-variance method was used to perform the meta-analysis. Larger studies with less variance were given more weight in the meta-analysis owing to more precise effect size estimates than smaller studies.

Continuous Outcome Measures

The standardized mean difference (SMD) effect size was calculated for the primary outcome using the Hedges g formula. Different assessment tools were used to calculate the effect size of the primary outcome in each study. Therefore, SMD was statistically suitable for estimating the effect size for each study. Forest plots were used to present study-specific effect sizes and overall effect sizes, including 95% CIs. Furthermore, we calculated the I^2 statistic to quantify heterogeneity and the χ^2 statistic to test for heterogeneity ($P \leq .10$ significance level).

For the secondary outcomes—patient satisfaction and working alliance—the mean difference (MD) effect size was calculated as these secondary outcomes were assessed using a single standardized tool (CSQ-8 and WAI-C). Therefore, standardization was not needed to calculate the effect size across studies. The same statistical approach used for the primary outcome was applied to the secondary outcomes of patient satisfaction and working alliance.

Postintervention data (sample size, mean, and SD) for each treatment modality (in person and telemedicine) were used to calculate the effect size of the continuous outcome measures (efficacy, satisfaction, and alliance), which is considered a valid approach [22].

Dichotomous Outcome Measures

The risk ratio effect size and its 95% CI was calculated for the secondary outcome attrition rate. A forest plot was created to present the effect size for each study and the overall effect size

for pooled analysis and was supplemented with I^2 and χ^2 statistics.

Additional Primary Outcome Analyses (Investigating Heterogeneity)

Moderator Analysis

For the primary outcome, subgroup analyses for different patient groups were performed based on (1) participant diagnosis, as specified in the eligibility criteria; (2) age; (3) length of treatment course or program; and (4) therapy type.

The year of study publication was evaluated through a meta-regression, and the *P* value for the regression was computed ($P \leq .05$ significance level).

Sensitivity Analysis

A sensitivity analysis was also performed to determine the robustness of the meta-analysis and included (1) sensitivity testing for only high-quality trials and (2) testing for whether the findings were sensitive to random effects or fixed effects models.

Meta-Bias

Publication bias was assessed visually using a funnel plot and tested statistically using the Egger test [19,23].

Certainty of the Evidence

The GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach as recommended by the Cochrane Collaboration was used to assess the certainty (confidence) of the evidence [19,24]. The certainty of the evidence for the primary outcome was evaluated for 5 domains and included an evaluation of the risk of bias (Rob 2), inconsistency, indirectness, imprecision, and publication bias. Each domain was graded as having a "serious," "very serious," or "not serious" impact on the certainty of the evidence and was downgraded with 1 level, 2 levels, or no downgrading, respectively. The GRADE approach was conducted using GRADEpro GDT software [25]. Two independent authors performed the GRADE approach (AAS and OJS), and a third author (SA) was consulted when disagreements occurred. The certainty of the evidence will be presented in the GRADE summary of the findings table.

Ethical Considerations and Dissemination

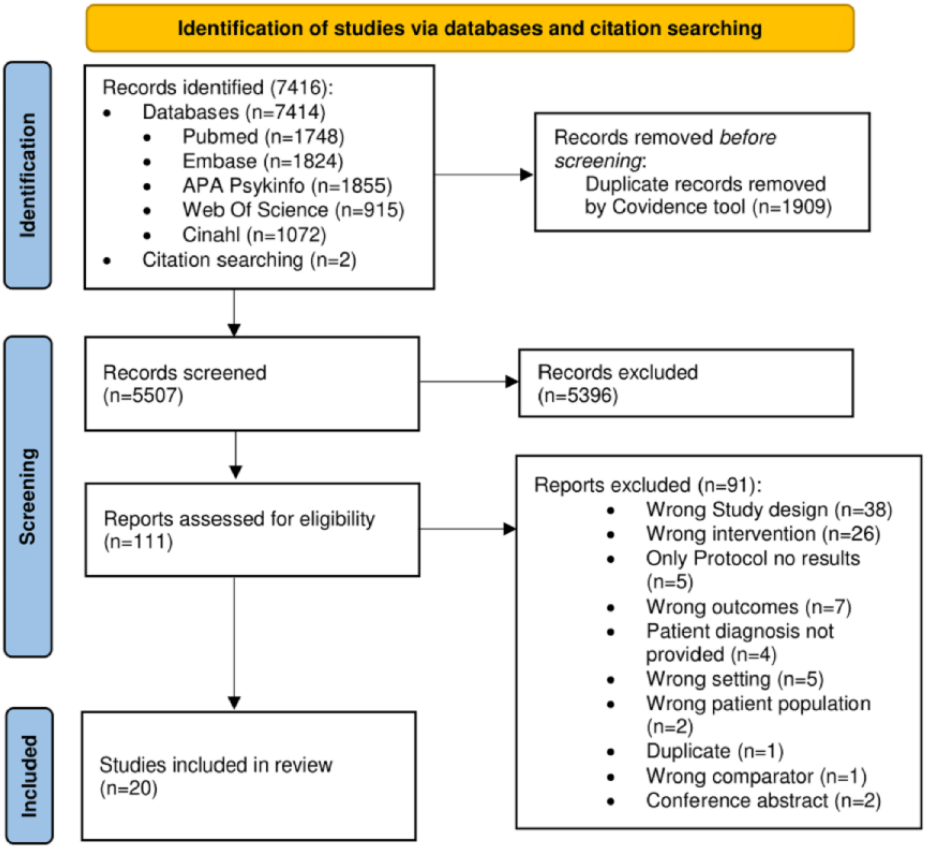
Ethical approval was not required for this systematic review and meta-analysis. The data sets (extraction) are deposited in the Zenodo repository (DOI:10.5281/zenodo.7339263). This study will be disseminated at scientific conferences.

Results

Study Selection

This study focused on outpatient psychiatric treatment conducted via telemedicine (video) or in person for PTSD, mood disorders, and anxiety disorders. The PRISMA flow diagram shows all stages of the article identification, screening, inclusion, and exclusion processes (Figure 1). Searches generated 7414 records. Two records were identified through manual literature search and reference list scanning. After removing the duplicates and applying inclusion and exclusion criteria to the titles and abstracts, 111 reports emerged as candidates for full-text review. A total of 20 studies were included in the final review. Cohen κ coefficient indicated fair interrater reliability for the title and abstract screening process (AAS and AA: Cohen κ=0.27), whereas it was moderate for the full-text review process (AAS and SFA: Cohen κ=0.52; AAS and JS: Cohen κ=0.5).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram across all stages of article identification.



Study Characteristics

Overview

Table 1 presents the descriptive characteristics of the 20 included studies.

Table 1. Overview of included studies.

Study, year	Sample size, n	Diagnosis	Age (years), mean (SD)	Female, n (%)	Intervention type	Number of sessions (interval in weeks)	Outcome or outcomes of interest (assessment instrument)	Bias (RoB 2[a]) assessed for primary outcome
Peterson et al [26], 2022	120	PTSD[b]	39.95 (10.36)	14 (11)	Cognitive processing therapy	12 (2 sessions per week)	• Psychopathology (PCL-5[c]) • Attrition	High risk
Acierno et al [27], 2021	136	PTSD	43.4 (11.5)	136 (100)	Prolonged exposure	14 (1)	• Psychopathology (PCL-5) • Attrition	High risk
Liu et al [28], 2020	207	PTSD	48.4 (14.1)	47 (23)	Cognitive processing therapy	12 (1)	• Psychopathology (CAPS[d]) • Attrition	Some concerns
Morland et al [29], 2020	175	PTSD	46.5 (14.11)	43 (24)	Prolonged exposure	6-15 (1)	• Psychopathology (CAPS-5) • Attrition	High risk
Watts et al [30], 2020	115	GAD[e]	41 (15.7)	95 (83)	Cognitive behavioral therapy	15 (1)	• Psychopathology (ADIS[f]-IV) • Working alliance (WAI-C[g]) • Attrition	High risk
Haghnia et al [31], 2019	71	PTSD	Range: 45-60 year	0 (0)	Case management	9 (1)	• Attrition	High risk
Acierno et al [32], 2016	265	PTSD	45.6 (14.9)	16 (6)	Behavioral activation	8 (1)	• Psychopathology (PCL-Military version) • Attrition	Some concerns
Hungerbuehler et al [33], 2016	107	Depression	35.64 (8.33)	76 (71)	Case management	5 (4)	• Psychopathology (HDRS[h]-17) • Satisfaction (CSQ[i]-8) • Working alliance (WAI-C) • Attrition	Some concerns
Luxton et al [34], 2016	121	Depression	Range: 19-65 year	22 (18)	Behavioral activation	8 (1)	• Psychopathology (BDI[j]-II) • Satisfaction (CSQ-8) • Attrition	Low risk
Maieritsch et al [35], 2016	90	PTSD	30.93 (6.05)	6 (7)	Cognitive processing therapy	A minimum of 10 (1)	• Psychopathology (CAPS) • Working alliance (WAI-C) • Attrition	Some concerns
Egede et al [36], 2015	241	Depression	63.9 (5.1)	5 (2)	Behavioral activation	8 (1)	• Psychopathology (BDI) • Attrition	Some concerns
Morland et al [37], 2015	124	PTSD	46.4 (11.9)	100 (100)	Cognitive processing therapy	12 (1)	• Psychopathology (CAPS) • Working alliance (WAI-C) • Attrition	Low risk
Yuen et al [38], 2015	52	PTSD	43.98 (15.18)	1 (2)	Prolonged exposure	8-12 (1)	• Psychopathology (CAPS) • Attrition	Low risk
Choi et al [39], 2014	85	Depression	65.21 (9.22)	66 (78)	Case management	6	• Psychopathology (HAM-Dh) • Attrition	High risk

Study, year	Sample size, n	Diagnosis	Age (years), mean (SD)	Female, n (%)	Intervention type	Number of sessions (interval in weeks)	Outcome or outcomes of interest (assessment instrument)	Bias (RoB 2[a]) assessed for primary outcome
Stubbings et al [40], 2013	26	Mixed Diagnosis[k]	20 (11)	15 (58)	Cognitive behavioral therapy	12 (1)	• Psychopathology (DASS[l] subscales) • Working alliance (WAI-C) • Satisfaction (CSQ-8) • Attrition	High risk
Chong et al [41], 2012	167	Depression	N/A[m]	148 (89)	Case management	6 (4)	• Psychopathology (PHQ[n]-9) • Working alliance (WAI-C) • Attrition	Some concerns
O'Reilly et al [42], 2007	495	Mixed Diagnosis	Range: 18-65 year	312 (63)	Case management	Up to 4 (4)	• Psychopathology (GSI[o]) • Satisfaction (CSQ-8) • Attrition	High risk
De Las Cuevas et al [43], 2006	140	Mixed Diagnosis	Range: 25-65 year	93 (66)	Cognitive behavioral therapy	8 (3)	• Psychopathology (SCL[p]-90R) • Attrition	Some concerns
Ruskin et al [44], 2004	119	Depression	49.7 (12.8)	14 (12)	Case management	8 (up to 7)	• Attrition	High risk
Bishop et al [45], 2002	24	Mixed Diagnosis	Range: 18-75 year	17 (71)	Case management	8	• Satisfaction (CSQ-8) • Attrition	Some concerns

[a]RoB 2: revised Cochrane risk-of-bias tool for randomized trials.

[b]PTSD: posttraumatic stress disorder.

[c]PCL: posttraumatic stress disorder checklist.

[d]CAPS: Clinician-Administered Posttraumatic Stress Disorder Scale.

[e]GAD: general anxiety disorder.

[f]ADIS: anxiety disorders interview schedule.

[g]WAI-C: Working Alliance Inventory-Client version.

[h]HDRS or HAM-D: Hamilton Depression Rating Scale.

[i]CSQ: Client Satisfaction Questionnaire.

[j]BDI: Beck Depression Inventory.

[k]The 4 studies with "mixed diagnosis" included mainly patients with PTSD and depressive disorders.

[l]DASS: Depression Anxiety Stress Scales.

[m]N/A: not applicable.

[n]PHQ: Patient Health Questionnaire.

[o]GSI: Global Severity Index.

[p]SCL: symptom checklist.

Demographics

The number of participants included in each study ranged from 24 to 495 (mean 144, SD 101.40). Out of 20 studies, 14 (70%) studies reported mean ages ranging from 20 to 65 years (mean 44.32, SD 11.19), 5 (25%) studies reported different ranges and percentages of age, and 1 (5%) study did not provide information about participant age. Most studies included a mix of males and females (17/20, 85%); 10% (2/20) of studies consisted of only females, and 5% (1/20) of only males. The diagnoses included PTSD (9/20, 45% studies), depressive disorders (6/20, 30% studies), a mix of different diagnoses (4/20, 20% studies), and general anxiety disorder (1/20, 5% studies).

Interventions

A range of treatment interventions via telemedicine and in person were offered, including prolonged exposure (3/20, 15% studies), cognitive processing therapy (4/20, 20% studies), behavioral activation treatment (3/20, 15% studies), cognitive behavioral therapy (3/20, 15% studies), and case management (7/20, 35% studies). The overall number of sessions varied between 4 and 15, with a median of around 8 sessions.

Outcomes

This study had 4 outcomes of interest: treatment efficacy, working alliance, treatment satisfaction, and attrition rates. Attrition rates were the only outcome reported in every study (20/20, 100%), whereas efficacy was measured in 85% (17/20) of studies. Out of 20 studies, 6 (30%) studies measured treatment satisfaction using the CSQ-8, and 6 (30%) studies measured the working alliance using a version of the WAI-C.

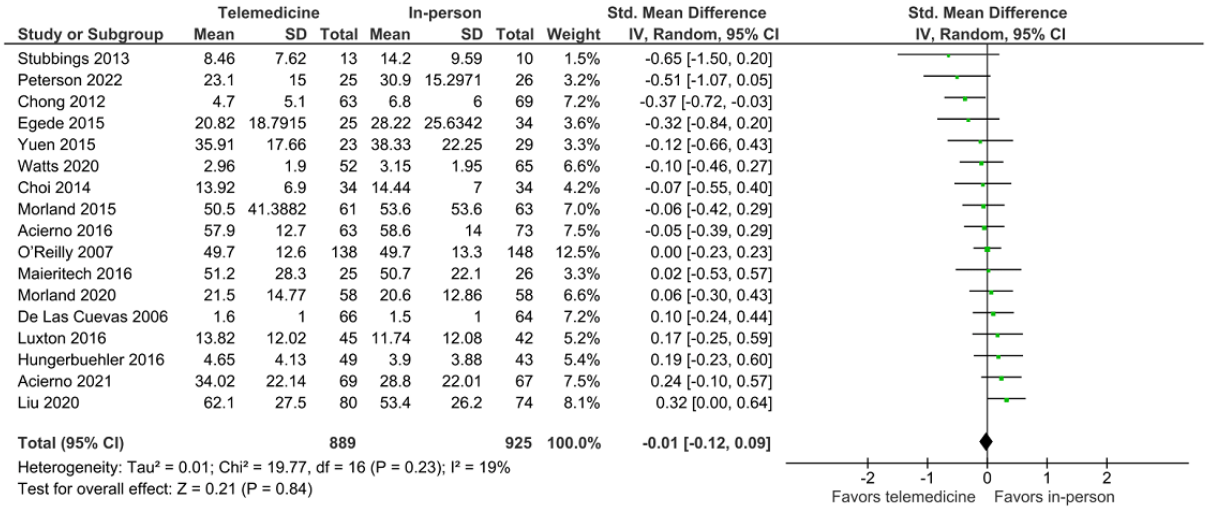
Treatment Efficacy (Psychopathology)

Overview

Data from 17 RCTs were pooled in the random-effect meta-analysis to examine the efficacy of treatment delivered via telemedicine and in person (Figure 2).

The effect size for each study was calculated and pooled. A negative effect size favors telemedicine treatment, whereas a positive effect size favors in-person treatment. The analysis shows that the CI for the overall estimated effect size include 0 (SMD=−0.01, 95% CI −0.12 to 0.09; *P*=.84; I^2=19%, 17 trials, n=1814). Therefore, the result indicates no statistical difference in the treatment effect between the in-person and telemedicine modalities. The estimated total heterogeneity or I^2 was 19%, indicating little heterogeneity. The 3 nonpooled studies, which did not provide enough information to be included in the quantitative analysis, reported the same results, that is, no difference between in-person and telemedicine treatment regarding efficacy [31,44,45].

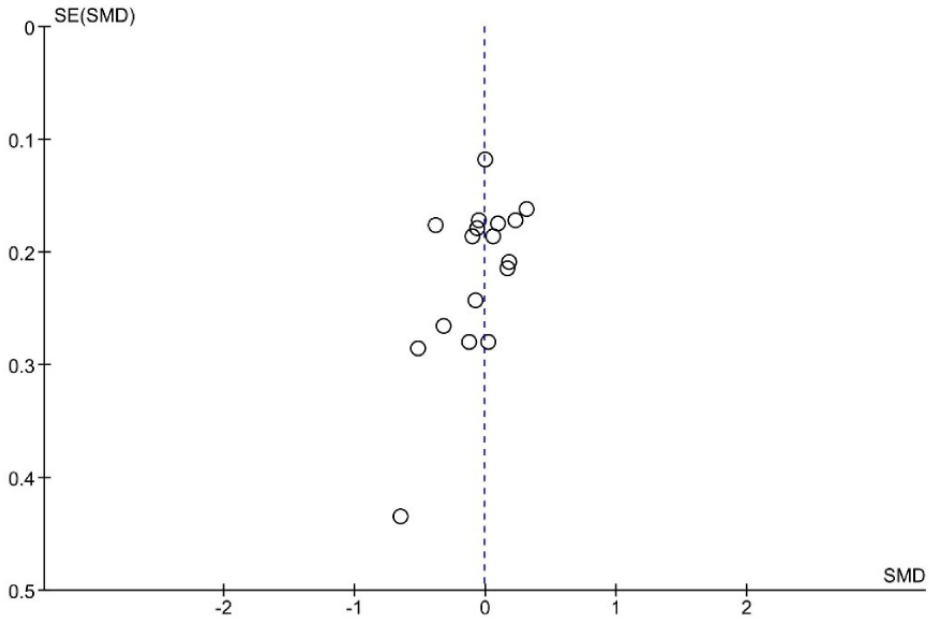
Figure 2. Forest plot (treatment efficacy) [26-30,32-43].



Meta-Bias

Publication bias was visually evaluated using a funnel plot (Figure 3). The funnel plot shows symmetry, indicating no risk of publication bias. The risk of publication bias was also evaluated statistically using the Egger test, which did not reveal the presence of funnel plot asymmetry (intercept=−1.718, 95% CI −3.57 to −1.818, t_{15}=−1.818; *P*=.09, 17 trials, n=1814).

Figure 3. Funnel plot (efficacy). SMD: standardized mean difference.



Certainty of the Evidence

The certainty of the evidence was rated as moderate according to GRADE (Figure 4). The downgrading was due to the risk of bias in multiple trials (poor reporting of how studies were planned and conducted). Of the 17 included trials, only 3 (18%) studies were rated as having a low risk of bias, 7 (41%) studies were rated as having some concerns of bias, and 7 (41%) studies were rated as having a high risk of bias.

Figure 4. GRADE (Grading of Recommendations Assessment, Development, and Evaluation) summary of findings (treatment efficacy). *Evidence limited due to risk of bias.

Is individual psychiatric outpatient treatment for adults using telemedicine as good as in-person modality regarding efficacy?

Patient or population: Mainly PTSD and depressive disorders

Setting: Psychiatric outpatients

Intervention: Telemedicine

Comparison: In-person

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with In-person	Risk with Telemedicine				
Treatment efficacy	-	SMD **0.01 SD lower** (0.12 lower to 0.9 higher)	-	1814 (17 RCTs)	⊕⊕⊕◯ Moderate*	

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: confidence interval; **SMD:** standardised mean difference

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Additional Primary Outcome Analyses (Investigating Heterogeneity)

Moderator Analyses

Moderator analyses were performed on 5 moderators: 4 performed in the subgroup analyses (age, diagnosis, number of sessions, and type of treatment) and 1 performed in a meta-regression analysis (publication year). The subgroup analyses presented in Table 2 indicate no statistical differences between the aggregated subgroups for any potential moderators. Most subgroup analyses showed little to moderate heterogeneity. In general, the results of the subgroup analyses must be interpreted with caution because of the small number of trials (<10) included in each subgroup analysis.

The meta-regression analysis included 17 trials and revealed no association between the estimated effect size and year of publication ($F_{1,15}$=0.355; P=.56; I^2=19.84, 17 trials, n=1814).

Table 2. Subgroup analyses.

Moderator	Included studies, n	SMD[a] (95% CI)	I^2	χ^2 (*df*)	Sample size, n	*P* value	$P_{subgroup}$ value
Age							.87
Age (range values: 18-75 years)	4	–0.03 (0.25 to 0.18)	0.43	5.3 (3)	635	.15	
Mean age (20-44 years)	7	–0.04 (–0.25 to 0.17)	0.29	8.5 (6)	522	.20	
Mean age (45-65 years)	6	0.02 (–0.14 to 0.19)	0.12	5.7 (5)	657	.34	
Diagnosis							.60
Mixed	3	–0.01 (–0.25 to 0.22)	0.22	2.6 (2)	439	.28	
PTSD[b]	8	0.04 (–0.12 to 0.20)	0.21	8.9 (7)	820	.26	
Depression	5	–0.08 (–0.33 to 0.16)	0.39	6.5 (4)	438	.16	
Number of sessions							.72
>8	9	0.00 (–0.17 to 0.18)	0.32	11.7 (8)	824	.16	
≤8	8	–0.04 (–0.17 to 0.10)	0.08	7.6 (7)	990	.37	
Treatment							.82
Prolonged exposure	3	0.11 (–0.12 to 0.34)	0	1.3 (2)	304	.54	
Cognitive Processing Therapy	4	–0.01 (–0.34 to 0.31)	0.57	7.0 (3)	380	.07	
Behavioral activation	3	–0.04 (–0.28 to 0,20)	0.03	2.1 (2)	282	.36	
Cognitive Behavioral Therapy	3	–0.07 (–0.36 to 0.23)	0.26	2.7 (2)	270	.26	
Case Management	4	–0.07 (–0.29 to 0.15)	0.38	4.8 (3)	578	.19	

[a]SMD: standardized mean difference.

[b]PTSD: posttraumatic stress disorder.

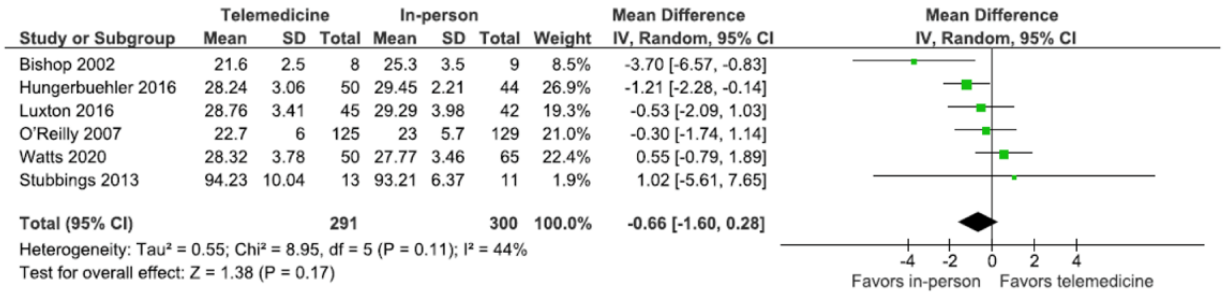
Sensitivity Analyses

Sensitivity analyses were performed to determine the robustness of the meta-analysis. It included (1) removing low-quality studies (ie, studies rated as "some concerns" and "high concerns" assessed using RoB 2) and (2) testing whether the findings are sensitive to fixed effects models. Three studies, judged to be high-quality studies, were pooled together and yielded results supporting the robustness of the meta-analysis, that is, no difference in the efficacy between in-person and telemedicine modalities (SMD=0.00, 95% CI −0.24 to 0.25; *P*=.98; I^2=0%, 3 trials, n=263). The overall estimated effect size was not sensitive to the fixed effects model (SMD=0.00, 95% CI −0.10 to 0.09; *P*=.92; I^2=19%, 17 trials, n=1814), further supporting the robustness of the meta-analysis.

Patient Satisfaction

Patient satisfaction was assessed using CSQ-8, which was used in 6 trials. The forest plot shows the overall estimated effect size for the pooled trials (Figure 5). The analysis shows that the overall estimated effect size include 0 (MD=−0.66, 95% CI −1.60 to 0.28; *P*=.17; I^2=44%, 6 trials, n=591). Therefore, the results indicate no statistical difference in patient satisfaction between the in-person and telemedicine treatment modalities. The estimated total heterogeneity for the analysis is 44%, indicating moderate heterogeneity.

Figure 5. Forest plot (patient satisfaction) [30,33,34,40,42,45].

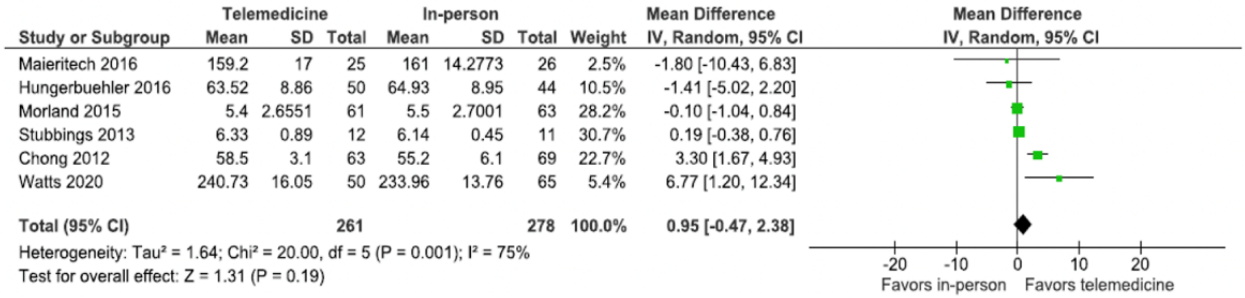


Working Alliance

The working alliance, as rated by patients, was assessed using WAI-C, which was applied in 6 trials. The forest plot shows the overall estimated effect size for the pooled trials (Figure 6). The analysis shows that the overall estimated effect size include 0 (MD=0.95, 95% CI −0.47 to 2.38; *P*=.19; I^2=75%, 6 trials, n=539). Thus, the results indicate no statistical difference in the

working alliance between the in-person and telemedicine treatment modalities, and the levels of the working alliance are comparable for the 2 treatment modalities. However, the estimated total heterogeneity for the analysis is 75%, indicating substantial to considerable heterogeneity and is statistically significant (χ^2_5=20; *P*=.001).

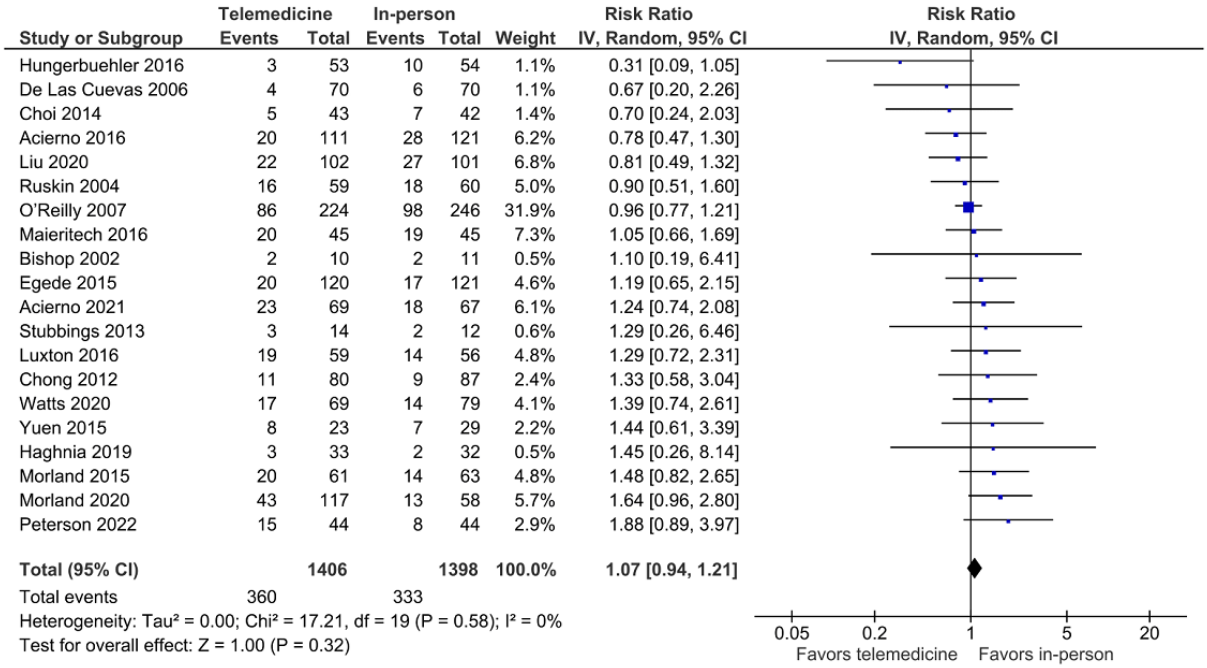
Figure 6. Forest plot (working alliance) [30,33,35,37,40].



Attrition Rate

All trials included in the meta-analysis (N=20) reported data on attrition rates. The number of attrition events for each modality was either directly extracted from the studies or calculated by subtracting the number of completers from the number of those randomized to a treatment modality. The forest plot shows the overall estimated effect size for the pooled trials assessing the attrition ratio between the modalities (Figure 7). The analysis shows that the overall estimated effect size include 1; therefore, the results indicate no statistical difference in patient attrition rates between the in-person and telemedicine treatment modalities (risk ratio 1.07, 95% CI 0.94-1.21; *P*=.32; I^2=0%, 20 trials, n=2804). The estimated total heterogeneity for the analysis is 0%, indicating no heterogeneity. The funnel plot and the Egger test was conducted and did not reveal the presence of asymmetry, indicating no risk of publication bias. Post hoc analysis was conducted to evaluate attrition outcomes based on diagnosis (PTSD and Depression), which did not influence the overall findings.

Figure 7. Forest plot (attrition rate).



Discussion

Principal Findings

The primary objective of this systematic review and meta-analysis was to evaluate whether individual psychiatric outpatient treatment for adults with PTSD, mood disorders, and anxiety disorders using telemedicine is comparable with in-person treatment. Second, the study evaluated whether patient satisfaction, working alliance using standard measures, and attrition rates were comparable between telemedicine and in-person treatment modalities. A comprehensive literature search for RCTs comparing telemedicine with in-person modalities was conducted. Using stringent eligibility criteria, 20 RCTs that met prespecified eligibility criteria were identified [14]. Overall, the study results indicate that treatment delivered through telemedicine is comparable with in-person treatment modality regarding treatment efficacy, patient satisfaction, working alliance, and attrition rate.

The primary outcome evaluated in this systematic review was efficacy and included 17 trials assessing treatment effects between the telemedicine and in-person treatment modalities. The results did not indicate a risk of publication bias. To test the robustness of the overall finding, that is, no difference in the treatment effect between telemedicine and in-person treatment modality regarding efficacy, prespecified sensitivity analyses were conducted, which supported the overall finding. Different moderators were tested for their potential influence on efficacy and included different age groups, diagnoses, number of sessions, types of psychotherapy, and publication year. None of the evaluated moderators affected the overall findings, strengthening the evidence for the equality between telemedicine and in-person treatment on various clinical and methodological characteristics. The results of this systematic review and meta-analysis are consistent with those of previously published meta-analyses, indicating the nonsuperiority of in person to telemedicine across psychiatric diagnoses [11-13]. However, when interpreting the results for the moderators in the subgroup analysis, it is important to note that relatively few studies (<10) were included in each moderator analysis, which is a considerable limitation [46]. The meta-analysis focused on PTSD, mood disorders, and anxiety disorders in which most trials included were focused on PTSD (n=8) and depression (n=5). There were no studies included that compared the treatment effects between telemedicine and in person for patients with personality disorders, social phobia, and agoraphobia. Further studies are required to determine the efficacy of psychiatric treatment using telemedicine for these disorders.

The secondary outcomes investigated between the 2 treatment modalities included patient satisfaction, working alliance, and attrition rates.

Satisfaction measurement was limited to CSQ-8, resulting in 6 RCTs in the meta-analysis. The results indicate that satisfaction between treatment modalities (in person and telemedicine) is comparable, and heterogeneity is low to moderate. To our knowledge, satisfaction has only been evaluated in a single meta-analysis by Hyler et al [47], who concluded the equivalence in satisfaction between telemedicine and in-person treatment modalities regarding psychiatric assessment. However, the authors also mention limitations with the study owing to the ad hoc and untested instruments applied for measuring satisfaction and pooling the satisfaction measure for both the patients and therapists in the same analysis. This meta-analysis on satisfaction was restricted to a single validated questionnaire (CSQ-8), strengthening the evidence for equivalence in treatment satisfaction between telemedicine and in-person modalities.

The assessment of working alliance was limited to alliance reported by patients (WAI-C), resulting in 6 RCTs in the meta-analysis. The results indicate that the working alliance between treatment modalities was comparable, although heterogeneity was substantial to considerable (I^2=75%; χ^2_5=20; P=.001), weakening the finding. Further analyses investigating heterogeneity were not possible because of the small number of included studies applying WAI-C, and this was not prespecified in the protocol. A working alliance is considered an important factor in psychotherapy outcomes, and clinicians have shown some concern that the telemedicine format of treatment may negatively impact the working alliance [48,49]. Previous research on alliance in telemedicine interventions has shown varied results but generally shows that a good therapeutic alliance can be established in telemedicine interventions and is comparable with in-person interventions [50]. Although patients consistently rate working alliance as good in telemedicine interventions, therapists have shown a tendency to rate the alliance as lower than in-person interventions. Norwood et al [3] evaluated the working alliance in a noninferiority meta-analysis and found that alliance in telemedicine is inferior to in person. Although this study combined ratings of patient and therapist alliance, the authors acknowledged that this procedure may have reduced the overall levels of alliance in the telemedicine condition. To our knowledge, this meta-analysis is the first study to examine therapeutic alliance as rated only by the patients. The comparable levels of alliance reported by patients are further supported by the similar levels of attrition and patient satisfaction rates between the telemedicine and in-person interventions found in this meta-analysis. Future studies should aim to examine how the working alliance is established over time and its implications for outcomes when treatment is delivered through telemedicine.

The attrition rate was comparable between treatment modalities (telemedicine and in person) across diagnoses (PTSD and depression) and is consistent with the results published by Giovanetti et al [13]. The authors included 11 RCTs and found equivalent attrition rates between the telemedicine and in-person treatment modalities; however, the authors only assessed attrition rates in patients with depression. Thus, the telemedicine modality of psychiatric treatment did not appear to negatively impact sustained engagement when compared with treatment via the in-person modality.

Study Limitations

First, the stringent eligibility criteria regarding population, intervention, control, and outcome measures limited the total number of studies available for analysis. Second, the inadequate and poor reporting of the included studies led to only 3 studies being evaluated as high-quality based on the RoB 2 criteria. Owing to the limited number of high-quality studies, the certainty of the evidence (according to GRADE) was rated as moderate, and a high certainty of the evidence was therefore not achievable for the efficacy outcome. Third, data were nonuniformly reported in a number of the included studies. For example, age was reported as ranges (percentages) in some studies, whereas other studies reported the mean age, making it challenging to aggregate subgroups for the moderator analysis. Finally, most studies did not provide separate results for male and female participants, so it was not possible to determine whether there was a differential effect of sex.

Implications for Clinical Practice and Research

Rapid digitalization within mental health care is changing clinical practice. This study provides a comprehensive and up-to-date meta-analytic overview of the use of telemedicine in individual psychiatric treatment across a range of diagnoses and clinical outcomes. Evidence generated by this meta-analysis

can guide clinical practice regarding which disorders can be effectively treated using individual psychiatric treatment conducted via telemedicine and which psychotherapy approaches are effective.

Importantly, the results of this review revealed a complete absence of high-quality studies examining the efficacy of psychiatric treatment via telemedicine for personality disorders and a range of anxiety disorders. This knowledge gap must be addressed in future studies. Future studies should also examine how to identify the most suitable treatment modality (telemedicine, in person, telephone, etc) for psychiatric patients, thereby matching patient needs to the treatment mode to optimize outcomes. Using individual patient data, meta-analysis within telemedicine applications could potentially address this challenge.

Conclusions

In summary, the results of this meta-analysis indicate that psychiatric treatment via telemedicine for PTSD, mood disorders, and anxiety disorders were equivalent to in-person treatment in terms of treatment efficacy, satisfaction, and attrition rate. Although working alliance as rated by patients was also deemed to be comparable between the 2 modalities, heterogeneity in the analysis was substantial. Thus, there is a need for further high-quality controlled trials to fully understand the complex issue of working alliance for interventions conducted via telemedicine.

Authors' Contributions

AAS is the guarantor of the systematic review and wrote the first draft. OJS provided Statistical Consultation and Rating the Certainty in Evidence (Grading of Recommendations Assessment, Development, and Evaluation). JPS contributed to data extraction, risk of bias assessment, and the table overview of included studies. AA contributed to records screening. JAS and HB contributed with scientific knowledge on information technology and health care technology. KT contributed with general scientific knowledge regarding telemedicine. ES developed the idea and rationale for the systematic review. SFA designed the data extraction form, data extraction, and risk of bias assessment. All authors revised and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 74 KB - mental_v10i1e44790_app1.pdf]

References

1. Chakrabarti S. Usefulness of telepsychiatry: a critical evaluation of videoconferencing-based approaches. World J Psychiatry 2015 Sep 22;5(3):286-304 [FREE Full text] [doi: 10.5498/wjp.v5.i3.286] [Medline: 26425443]
2. Hubley S, Lynch SB, Schneck C, Thomas M, Shore J. Review of key telepsychiatry outcomes. World J Psychiatry 2016 Jun 22;6(2):269-282 [FREE Full text] [doi: 10.5498/wjp.v6.i2.269] [Medline: 27354970]
3. Norwood C, Moghaddam NG, Malins S, Sabin-Farrell R. Working alliance and outcome effectiveness in videoconferencing psychotherapy: a systematic review and noninferiority meta-analysis. Clin Psychol Psychother 2018 Nov;25(6):797-808. [doi: 10.1002/cpp.2315] [Medline: 30014606]
4. Cowan KE, McKean AJ, Gentry MT, Hilty DM. Barriers to use of telepsychiatry: clinicians as gatekeepers. Mayo Clin Proc 2019 Dec;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
5. Hailey D, Ohinmaa A, Roine R. Limitations in the routine use of telepsychiatry. J Telemed Telecare 2009;15(1):28-31. [doi: 10.1258/jtt.2008.080609] [Medline: 19139217]
6. Bennett CB, Ruggero CJ, Sever AC, Yanouri L. eHealth to redress psychotherapy access barriers both new and old: a review of reviews and meta-analyses. J Psychother Integr 2020 Jun;30(2):188-207 [FREE Full text] [doi: 10.1037/int0000217]
7. Wright JH, Caudill R. Remote treatment delivery in response to the COVID-19 pandemic. Psychother Psychosom 2020;89(3):130-132 [FREE Full text] [doi: 10.1159/000507376] [Medline: 32213775]
8. Wells SY, Morland LA, Wilhite ER, Grubbs KM, Rauch SA, Acierno R, et al. Delivering prolonged exposure therapy via videoconferencing during the COVID-19 pandemic: an overview of the research and special considerations for providers. J Trauma Stress 2020 Aug;33(4):380-390 [FREE Full text] [doi: 10.1002/jts.22573] [Medline: 32881116]
9. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of telehealth during the COVID-19 pandemic: scoping review. J Med Internet Res 2020 Dec 01;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]
10. Reeves JJ, Ayers JW, Longhurst CA. Telehealth in the COVID-19 era: a balancing act to avoid harm. J Med Internet Res 2021 Feb 01;23(2):e24785 [FREE Full text] [doi: 10.2196/24785] [Medline: 33477104]
11. Drago A, Winding TN, Antypa N. Videoconferencing in psychiatry, a meta-analysis of assessment and treatment. Eur Psychiatry 2016 Aug;36:29-37. [doi: 10.1016/j.eurpsy.2016.03.007] [Medline: 27311105]

12. Batastini AB, Paprzycki P, Jones AC, MacLean N. Are videoconferenced mental and behavioral health services just as good as in-person? A meta-analysis of a fast-growing practice. Clin Psychol Rev 2021 Feb;83:101944. [doi: 10.1016/j.cpr.2020.101944] [Medline: 33227560]
13. Giovanetti AK, Punt SE, Nelson EL, Ilardi SS. Teletherapy versus in-person psychotherapy for depression: a meta-analysis of randomized controlled trials. Telemed J E Health 2022 Aug;28(8):1077-1089. [doi: 10.1089/tmj.2021.0294] [Medline: 35007437]
14. Shaker AA, Austin SF, Sørensen JA, Storebø OJ, Simonsen E. Psychiatric treatment conducted via telemedicine versus in-person consultations in mood, anxiety and personality disorders: a protocol for a systematic review and meta-analysis. BMJ Open 2022 Sep 28;12(9):e060690 [FREE Full text] [doi: 10.1136/bmjopen-2021-060690] [Medline: 36171023]
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
16. Clarivate version: EndNote X9 type: 64 bit. EndNote. 2013. URL: https://endnote.com/downloads [accessed 2023-04-15]
17. Veritas Health Innovation. Covidence. URL: https://www.covidence.org/ [accessed 2023-04-15]
18. Review manager (RevMan). Cochrane Collaboration. 2020. URL: https://training.cochrane.org/online-learning/core-software/revman [accessed 2023-04-15]
19. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M. Cochrane handbook for systematic reviews of interventions. Version 6.2. Cochrane. 2021 Feb. URL: https://training.cochrane.org/handbook [accessed 2023-04-15]
20. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
21. Horvath AO, Greenberg LS. Development and validation of the working alliance inventory. J Couns Psychol 1989;36(2):223-233 [FREE Full text] [doi: 10.1037/0022-0167.36.2.223]
22. da Costa BR, Nüesch E, Rutjes AW, Johnston BC, Reichenbach S, Trelle S, et al. Combining follow-up and change data is valid in meta-analyses of continuous outcomes: a meta-epidemiological study. J Clin Epidemiol 2013 Aug;66(8):847-855. [doi: 10.1016/j.jclinepi.2013.03.009] [Medline: 23747228]
23. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
24. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. The GRADE Working Group. URL: https://www.gradepro.org/terms/cite [accessed 2022-11-19]
25. GDT: GRADEpro guideline development tool. GRADEpro. URL: https://www.gradepro.org/terms/cite [accessed 2022-11-19]
26. Peterson AL, Mintz J, Moring JC, Straud CL, Young-McCaughan S, McGeary CA, et al. In-office, in-home, and telehealth cognitive processing therapy for posttraumatic stress disorder in veterans: a randomized clinical trial. BMC Psychiatry 2022 Jan 17;22(1):41 [FREE Full text] [doi: 10.1186/s12888-022-03699-4] [Medline: 35038985]
27. Acierno R, Jaffe AE, Gilmore AK, Birks A, Denier C, Muzzy W, et al. A randomized clinical trial of in-person vs. home-based telemedicine delivery of prolonged exposure for PTSD in military sexual trauma survivors. J Anxiety Disord 2021 Oct;83:102461. [doi: 10.1016/j.janxdis.2021.102461] [Medline: 34391978]
28. Liu L, Thorp SR, Moreno L, Wells SY, Glassman LH, Busch AC, et al. Videoconferencing psychotherapy for veterans with PTSD: results from a randomized controlled non-inferiority trial. J Telemed Telecare 2020 Oct;26(9):507-519. [doi: 10.1177/1357633X19853947] [Medline: 31216210]
29. Morland LA, Mackintosh MA, Glassman LH, Wells SY, Thorp SR, Rauch SA, et al. Home-based delivery of variable length prolonged exposure therapy: a comparison of clinical efficacy between service modalities. Depress Anxiety 2020 Apr;37(4):346-355. [doi: 10.1002/da.22979] [Medline: 31872563]
30. Watts S, Marchand A, Bouchard S, Gosselin P, Langlois F, Belleville G, et al. Telepsychotherapy for generalized anxiety disorder: impact on the working alliance. J Psychother Integr 2020 Jun;30(2):208-225 [FREE Full text] [doi: 10.1037/int0000223]
31. Haghnia Y, Samad-Soltani T, Yousefi M, Sadr H, Rezaei-Hachesu P. Telepsychiatry-based care for the treatment follow-up of Iranian war veterans with post-traumatic stress disorder: a randomized controlled trial. Iran J Med Sci 2019 Jul;44(4):291-298 [FREE Full text] [doi: 10.30476/IJMS.2019.44944] [Medline: 31439972]
32. Acierno R, Gros DF, Ruggiero KJ, Hernandez-Tejada BM, Knapp RG, Lejuez CW, et al. Behavioral activation and therapeutic exposure for posttraumatic stress disorder: a noninferiority trial of treatment delivered in person versus home-based telehealth. Depress Anxiety 2016 May;33(5):415-423. [doi: 10.1002/da.22476] [Medline: 26864655]
33. Hungerbuehler I, Valiengo L, Loch AA, Rössler W, Gattaz WF. Home-based psychiatric outpatient care through videoconferencing for depression: a randomized controlled follow-up trial. JMIR Ment Health 2016 Aug 03;3(3):e36 [FREE Full text] [doi: 10.2196/mental.5675] [Medline: 27489204]
34. Luxton DD, Pruitt LD, Wagner A, Smolenski DJ, Jenkins-Guarnieri MA, Gahm G. Home-based telebehavioral health for U.S. military personnel and veterans with depression: a randomized controlled trial. J Consult Clin Psychol 2016 Nov;84(11):923-934. [doi: 10.1037/ccp0000135] [Medline: 27599225]

35. Maieritsch KP, Smith TL, Hessinger JD, Ahearn EP, Eickhoff JC, Zhao Q. Randomized controlled equivalence trial comparing videoconference and in person delivery of cognitive processing therapy for PTSD. J Telemed Telecare 2016 Jun;22(4):238-243. [doi: 10.1177/1357633X15596109] [Medline: 26231819]
36. Egede LE, Acierno R, Knapp RG, Walker RJ, Payne EH, Frueh BC. Psychotherapy for depression in older veterans via telemedicine: effect on quality of life, satisfaction, treatment credibility, and service delivery perception. J Clin Psychiatry 2016 Dec;77(12):1704-1711. [doi: 10.4088/JCP.16m10951] [Medline: 27835713]
37. Morland LA, Mackintosh M, Rosen CS, Willis E, Resick P, Chard K, et al. Telemedicine versus in-person delivery of cognitive processing therapy for women with posttraumatic stress disorder: a randomized noninferiority trial. Depress Anxiety 2015 Nov;32(11):811-820. [doi: 10.1002/da.22397] [Medline: 26243685]
38. Yuen EK, Gros DF, Price M, Zeigler S, Tuerk PW, Foa EB, et al. Randomized controlled trial of home-based telehealth versus in-person prolonged exposure for combat-related PTSD in veterans: preliminary results. J Clin Psychol 2015 Jun;71(6):500-512. [doi: 10.1002/jclp.22168] [Medline: 25809565]
39. Choi NG, Hegel MT, Marti CN, Marinucci ML, Sirrianni L, Bruce ML. Telehealth problem-solving therapy for depressed low-income homebound older adults. Am J Geriatr Psychiatry 2014 Mar;22(3):263-271 [FREE Full text] [doi: 10.1016/j.jagp.2013.01.037]
40. Stubbings DR, Rees CS, Roberts LD, Kane RT. Comparing in-person to videoconference-based cognitive behavioral therapy for mood and anxiety disorders: randomized controlled trial. J Med Internet Res 2013 Nov 19;15(11):e258 [FREE Full text] [doi: 10.2196/jmir.2564] [Medline: 24252663]
41. Chong J, Moreno F. Feasibility and acceptability of clinic-based telepsychiatry for low-income Hispanic primary care patients. Telemed J E Health 2012 May;18(4):297-304. [doi: 10.1089/tmj.2011.0126] [Medline: 22424078]
42. O'Reilly R, Bishop J, Maddox K, Hutchinson L, Fisman M, Takhar J. Is telepsychiatry equivalent to face-to-face psychiatry? Results from a randomized controlled equivalence trial. Psychiatr Serv 2007 Jun;58(6):836-843. [doi: 10.1176/ps.2007.58.6.836] [Medline: 17535945]
43. De Las Cuevas C, Arredondo MT, Cabrera MF, Sulzenbacher H, Meise U. Randomized clinical trial of telepsychiatry through videoconference versus face-to-face conventional psychiatric treatment. Telemed J E Health 2006 Jun;12(3):341-350. [doi: 10.1089/tmj.2006.12.341] [Medline: 16796502]
44. Ruskin P, Silver-Aylaian M, Kling MA, Reed SA, Bradham DD, Hebel JR, et al. Treatment outcomes in depression: comparison of remote treatment through telepsychiatry to in-person treatment. Am J Psychiatry 2004 Aug;161(8):1471-1476. [doi: 10.1176/appi.ajp.161.8.1471] [Medline: 15285975]
45. Bishop JE, O'Reilly RL, Maddox K, Hutchinson LJ. Client satisfaction in a feasibility study comparing face-to-face interviews with telepsychiatry. J Telemed Telecare 2002;8(4):217-221. [doi: 10.1258/135763302320272185] [Medline: 12217104]
46. Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. Cochrane Training. URL: https://training.cochrane.org/handbook/current/chapter-10 [accessed 2021-11-15]
47. Hyler SE, Gangure DP, Batchelder ST. Can telepsychiatry replace in-person psychiatric assessments? A review and meta-analysis of comparison studies. CNS Spectr 2005 May;10(5):403-413. [doi: 10.1017/s109285290002277x] [Medline: 15858458]
48. Rees CS, Stone S. Therapeutic alliance in face-to-face versus videoconferenced psychotherapy. Prof Psychol Res Pr 2005;36(6):649-653 [FREE Full text] [doi: 10.1037/0735-7028.36.6.649]
49. Simpson SG, Reid CL. Therapeutic alliance in videoconferencing psychotherapy: a review. Aust J Rural Health 2014 Dec;22(6):280-299. [doi: 10.1111/ajr.12149] [Medline: 25495622]
50. Berger T. The therapeutic alliance in internet interventions: a narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]

Abbreviations

CSQ-8: Client Satisfaction Questionnaire-8
GRADE: Grading of Recommendations Assessment, Development, and Evaluation
MD: mean difference
MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
RoB 2: revised Cochrane risk-of-bias tool for randomized trials
SMD: standardized mean difference
WAI-C: Working Alliance Inventory-Client version

Edited by J Torous; submitted 05.12.22; peer-reviewed by Z Niu, Y Chu; comments to author 06.03.23; revised version received 28.04.23; accepted 22.05.23; published 05.07.23.

Please cite as:
Shaker AA, Austin SF, Storebø OJ, Schaug JP, Ayad A, Sørensen JA, Tarp K, Bechmann H, Simonsen E
Psychiatric Treatment Conducted via Telemedicine Versus In-Person Modality in Posttraumatic Stress Disorder, Mood Disorders, and Anxiety Disorders: Systematic Review and Meta-Analysis
JMIR Ment Health 2023;10:e44790
URL: https://mental.jmir.org/2023/1/e44790
doi:10.2196/44790
PMID:37277113

Review

The Use of Virtual Reality Interventions to Promote Positive Mental Health: Systematic Literature Review

Giorgio Li Pira[1], MD; Beatrice Aquilini[1], MD; Alessandro Davoli[1], MD; Silvana Grandi[1], MD; Chiara Ruini[2], PhD

[1]Department of Psychology, University of Bologna, Bologna, Italy

[2]Department for Life Quality Studies, University of Bologna, Rimini, Italy

Corresponding Author:
Chiara Ruini, PhD
Department for Life Quality Studies
University of Bologna
Corso d'Augusto 237
Rimini, 47921
Italy
Phone: 39 0547 338533
Email: chiara.ruini@unibo.it

Abstract

Background: A large body of research has documented the efficacy of psychological interventions integrated with virtual reality (VR) therapies in treating psychiatric disorders. However, the concept of positive mental health calls for a 2-fold approach in which both symptoms and positive functioning should be addressed by modern interventions.

Objective: This review aimed to summarize studies that applied VR therapies by embracing the positive mental health perspective.

Methods: A literature search was conducted by entering the following keywords—"virtual reality" AND "intervention" OR "treatment" OR "therapy" AND "mental health" NOT "systematic review or meta-analysis"—and limiting it to "journal article" and the English language. To be included in this review, articles had to present at least one quantitative measure of positive functioning and one quantitative measure of symptoms or distress and had to investigate adult populations, including populations with psychiatric disorders.

Results: A total of 20 articles were included. They described various VR protocols that were applied for the treatment of anxiety disorders (5/20, 25%), depression (2/20, 10%), posttraumatic stress disorder (3/20, 15%), psychosis (3/20, 15%), and stress (7/20, 35%). Most of the studies (13/20, 65%) showed the beneficial effects of VR therapies in improving stress and negative symptoms. However, 35% (7/20) of the studies showed no or a small effect on the various dimensions of positivity, particularly in clinical samples.

Conclusions: VR interventions might be cost-effective and largely scalable, but further research is needed to develop existing VR software and treatments according to the modern positive mental health approach.

(JMIR Ment Health 2023;10:e44998) doi:10.2196/44998

KEYWORDS

positive mental health; well-being; virtual reality; interventions; psychopathology; mobile phone

Introduction

Background

The implementation of virtual reality (VR) programs in psychotherapy has been increasing over the past 3 decades [1-3], especially after the past COVID-19 pandemic (Table 1). Since the early 1990s, VR has been used to treat disorders such as specific phobias, panic disorder, and posttraumatic stress disorder (PTSD). Although the technology was quite raw, it was sufficiently capable of producing convincing 3D images inside a head-mounted display (HMD) or inside a room where the images were projected on the walls (cave automatic virtual environment) [4]. Today, many protocols are available for the treatment of different psychopathologies, from VR exposure therapy (VRET) to protocols dedicated to eating disorders, depression, obsessive-compulsive disorder (OCD), and psychosis.

Table 1. Cumulative record of publications in the PsycInfo database under the keywords phrase "VR intervention" from the years 2012 to 2022.

Year	Cumulative record of publications, n
2012	7
2013	9
2014	11
2015	6
2016	4
2017	10
2018	17
2019	28
2020	52
2021	63
2022	31

However, the current definition of mental health, which emerged from the positive psychology perspective, entails 2 related components: symptoms or psychopathology and well-being and positive functioning [5,6]. According to this model, psychopathology and well-being coexist along a continuum where the states of mental illness and languishing are the negative components in contrast to the state of flourishing, which is the positive component. Languishing is defined as a state of emptiness in which the individual experiences few instances of well-being and is considered vulnerable to the development of psychiatric disorders. Flourishing, in contrast, is also labeled as optimal human functioning and is characterized by the presence of high levels of emotional, psychological, and social well-being. The treatment of symptoms and distress has been the main focus of the mental health agenda over the past decades. However, according to this complete mental health model, the promotion of well-being and optimal human functioning should receive the same attention. According to the positive psychology perspective, positive psychological interventions can be conceived as "treatment methods" or intentional activities that aim to cultivate positive feelings, behaviors, or cognitions [7]. Longitudinal investigations have shown that cross-time gains in well-being predicted cross-time declines in mental illnesses and vice versa [8]. Various authors have observed that positive and negative characteristics can be considered at opposite ends of a single dimension and that their valence can vary according to the specific context [6,9,10]. For example, anger can become adaptive when it helps individuals reach their goals or fight for their ideas, depression can exist in the continuum with happiness, and it may become adaptive in encouraging people to change suboptimal aspects of their lives. Similarly, anxiety and calmness exist on the same continuum and are often assessed using similar items but with opposite coding procedures. As a consequence, mental health involves a complex balance of positive and negative psychological characteristics, and modern interventions should be able to address both components. Wood and Tarrier [9] proposed a positive clinical psychology framework in which clinical interventions may help individuals move away from symptoms and implicitly have an equal and opposite effect on well-being. Thus, the authors call for an integration between positive and clinical psychology in which psychological interventions should treat distress as well as they should promote well-being. According to Keyes [5], the promotion of well-being in the long run can be considered the most cost-effective mental health policy as it allows for the maintenance of optimal functioning in the population and prevents the onset of psychiatric disorders and the economic burden of expensive treatments.

Digital technologies can be considered valid tools to pursue this goal. Recently, with the advent of the pandemic, digital technologies have been used to deliver web-based psychotherapeutic interventions for addressing the psychological effects of COVID-19 or promoting positive functioning and well-being [11,12]. The latter applications can be subsumed under the umbrella of "Positive Technology" research [13].

As stated by Riva [13], "Positive Technology consists in the scientific and applied approach to the use of technology for improving the quality of our personal experience." The core theoretical framework of this approach comes from the field of positive psychology and aims to use technology to manipulate the quality of experience, increase well-being, and generate strengths and resilience at the level of individuals, organizations, and society. Therefore, owing to "Positive Technology," it is possible to study and understand how digital technologies can be used to promote health and well-being [14]. VR, owing to its ability to create controlled and tailored experiences that enhance the user's sense of presence [13,15], has proven to be an important tool for increasing well-being. In particular, VR can affect 3 characteristics of personal experience that can promote personal well-being. VR can induce positive and pleasurable experiences; foster engagement and self-realization; and support and enhance connection and interpersonal relationships among individuals, groups, and organizations [16]. Studies conducted thus far have demonstrated the effectiveness of VR in several areas of application [17]. However, a limited body of research has used the complete mental health approach and tested whether modern, digitalized interventions were able to address symptoms and distress and promote well-being. A notable exception is a review published recently in this journal [18] that focused on the promotion of well-being through digital technologies. It found that self-help interventions improved

well-being in young people aged 9 to 25 years. However, the review targeted the young population only. This population has very specific psychological features that change according to the stage of development of the participants. The results of this review cannot be generalized to the rest of the population. Similarly, other research groups have developed digital interventions for older adults and documented encouraging results for the promotion of positive mental health in the aging population [19-21].

Objectives

This systematic review aimed to fill this literature gap by summarizing the most recent (last 10 years) research in the field. To the authors' knowledge, the current literature lacks a systematic review of studies that provide a global assessment of the benefits of VR in treating symptoms and promoting well-being in adult populations. Therefore, the primary goal of this work was to collect and summarize scientific literature on the application of VR therapies according to the positive mental health framework, which includes not only data regarding the efficacy of VR treatments in improving symptoms but also data on the impact of these technologically advanced interventions in promoting positive functioning. In doing so, we hope to provide the reader with a broader understanding of the impact of these emerging digital therapies on the global mental health of adult individuals.

Considering the large number of private and public investments in digital technologies for mental health, this review may provide important initial data on their worthiness for promoting complete mental health in the adult population through the use of VR.

Methods

Design

A systematic review was conducted to extract recently published scientific papers that dealt with measures of positive functioning and symptoms and distress in adult populations. We focused on the last 10 years (from 2012 onward) as during that period, the Positive Technology research agenda was scientifically recognized and, since then, many VR protocols have been developed. This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [22].

Search Method

A systematic literature search was conducted on the following electronic databases: EBSCOhost (PsycArticles and PsycInfo), PubMed, and Scopus.

The literature search was conducted by entering the following keywords: "virtual reality" AND "intervention" OR "treatment" OR "therapy" AND "mental health" NOT "systematic review or meta-analysis."

Selection Criteria

Inclusion

To be included in this review, the studies should involve (1) measures of positive functioning, (2) measures of symptoms and psychological distress, (3) adult populations, and (4) a description of a psychological intervention delivered using the VR system.

Exclusion

Studies were excluded if they (1) did not provide an assessment of positive functioning; (2) did not provide an assessment of symptoms; (3) were not experimental studies (eg, systematic reviews, protocols, or book chapters); (4) included only older adult or child populations; and (5) included adult populations with eating disorders, gambling, or substance use disorders. The latter clinical conditions were excluded as the source of positive emotions and well-being could be directly related to the exposure in the virtual environment to the objects of their addiction and could be a manifestation of the disorder itself [6,23,24]. Thus, individuals diagnosed with these clinical conditions may have a more complex balance between symptoms and positive functioning [23,25,26] and may require complex treatments with medications or further treatment ingredients [6,24,27,28].

Data Collection

The titles and abstracts of the articles assessed for potential inclusion were identified and independently inspected by 1 reviewer, who excluded duplicates and articles that did not meet the inclusion criteria. All potentially relevant articles were then fully assessed by other 2 reviewers, who decided on inclusion. Uncertainty was resolved by coming to a consensus. Data were extracted using a predesigned template (Multimedia Appendix 1 [29-48]) with the following specified headings: study, sample size, outcome measures, treatment conditions, follow-up, main findings, and limitations.

Risk of Bias and Quality of the Articles

To assess the bias of the individual studies, the following tools were used to sample a selection of quantitative studies: the Cochrane risk-of-bias tool [49] for between-group studies and the Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields [50] for single-arm or within-subject studies. The risk of bias was calculated by one of the authors for all the included articles.

Results

Search Outcome

The first screening identified 682 articles. Most of them (626/682, 91.8%) were extracted from the EBSCOhost and Scopus databases. Only a few articles (56/682, 8.2%) were extracted from PubMed. After removing duplicates, 676 publications were identified and individually assessed based on the study title and the information provided in the abstract (see the review flowchart in Figure 1). Of these 676 papers, 501 (74.1%) were excluded according to the aforementioned inclusion and exclusion criteria, whereas the full texts of 174 (25.7%) articles were examined before the decision was made on whether to include them. Of these 174 studies, 76 (43.7%) did not assess well-being in combination with psychopathology or symptoms, 51 (29.3%) were not empirical studies, 26 (14.9%) did not include our target adult population, and it was not

possible to obtain the full text of 1 (0.6%) paper; therefore, they were excluded. A total of 20 papers met all the inclusion criteria and were included in this review.

Multimedia Appendix 1 presents a summary of the findings from the 20 studies included in the review. Of these 20 studies, only 8 (40%) were randomized controlled studies, of which 3 (38%) used a waiting list as a control condition; 6 (30%) were pilot studies; 4 (20%) were case series or single-arm conditions; 1 (5%) had a within-subject design; and 1 (5%) had a between-subject design with 3 intervention groups. Hence, the quality of the included papers was low considering the high percentage of pilot and non–randomized controlled trial (RCT) studies.

The risk of bias was found to be low for both between-group and single-arm or within-subject studies. A total of 20% (2/10) [29,30] of the studies assessed using the Cochrane risk-of-bias tool showed some concern relative to the overall risk of bias, and 10% (1/10) of the studies [31] assessed using the Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields showed some concern related to the quality of the study (see Table 2 and Figure 2 for a summary).

A total of 900 participants received an intervention integrated with the use of VR technologies. The main outcome measures used to evaluate the effects of the VR therapies on symptoms were questionnaires and interviews considered the gold standard in the evaluation of those specific psychopathologies (ie, the Beck Depression Inventory for depression or the State-Trait Anxiety Inventory for anxiety). Some investigations also used general indicators of distress such as the Depression, Anxiety, and Stress Scale or the Patient Health Questionnaire (Multimedia Appendix 1). Conversely, when considering positive functioning, the authors referred to different domains of positivity, including positive emotions, self-compassion, social functioning, and relaxation (see Textbox 1 for a summary). The Positive and Negative Affect Schedule (PANAS) and measures of quality of life were the most commonly used indicators.

Figure 1. Flowchart depicting the identification and selection of the articles.

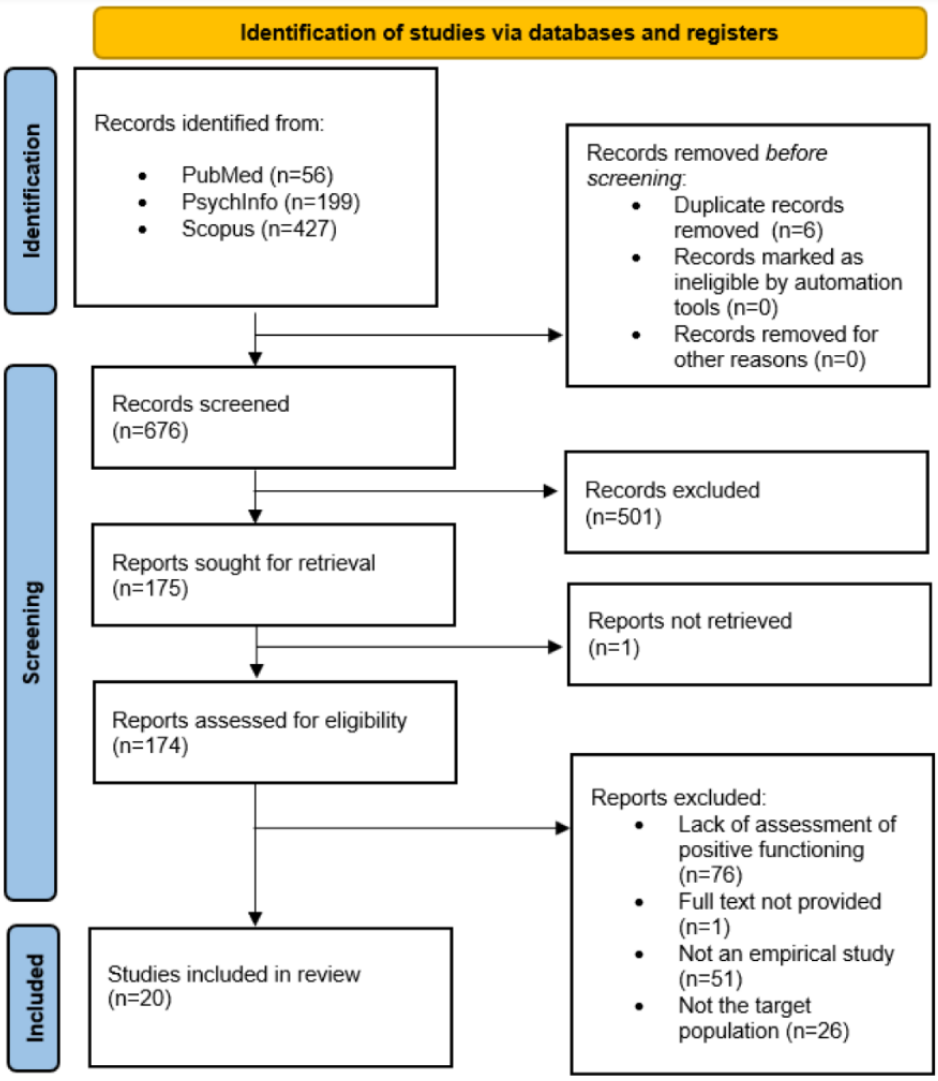


Table 2. Quality assessment carried out using the Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields.

Study	Quality assessment
Lindner et al [32], 2020	0.95
Geraets et al [35], 2019	0.86
Jones et al [31], 2020	0.5[a]
Tang et al [36], 2021	0.77
Falconer et al [38], 2016	0.95
Habak et al [39], 2020	0.86
Thompson et al [42], 2020	0.95
Chan et al [46], 2021	1
Desai et al [47], 2021	0.9
Riva et al [48], 2021	1

[a]High risk of bias.

Figure 2. Risk of bias carried out using the Cochrane Risk of Bias Tool [29,30,33,34,37,40,41,43-45].

Textbox 1. Summary of scales for positive functioning used in the included studies.

Specific phobia

- Rus-Calafell et al [33]: Life Interference Scale
- Lindner et al [32]: Brunnsviken Brief Quality of Life Scale

Social anxiety disorder

- Kampmann et al [34]: EQ-5D
- Geraets et al [35]: Manchester Short Assessment of Quality of Life (MSAQ)

Generalized anxiety disorder

- Malbos et al [30]: 12-item Short Form Health Survey

Posttraumatic stress disorder

- Jones et al [31]: Difficulties in Emotion Regulation Scale (DERS)
- Tang et al [36]: DERS and qualitative analysis
- Vlake et al [37]: 36-item Short Form Health Survey, EQ-5D, and health-related quality of life

Depression

- Falconer et al [38]: Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form (QLS-SF)
- Habak et al [39]: Short Warwick-Edinburgh Mental Well-being Scale

Psychosis and schizophrenia

- du Sert et al [40]: QLS-SF
- Pot-Kolder et al [41]: MSAQ
- Thompson et al [42]: EQ-5D

Stress and adjustment disorder

- Chan et al [46]: Positive and Negative Affect Schedule (PANAS)
- Desai et al [47]: Pittsburgh Sleep Quality Index and qualitative and thematic analysis
- Riva et al [48]: Social Connectedness Scale and Smith Relaxation States Inventory 3
- Gaggioli et al [43]: Satisfaction with Life Scale
- Cebolla et al [44]: PANAS, State Mindfulness Scale, Self-Other Four Immeasurables Scale, and Mindful Self-Care Scale
- Matsumoto et al [29]: psychological check test
- Richesin et al [45]: PANAS

In the following sections, we briefly illustrate the various studies included in this review, each organized under the clinical conditions for which the VR interventions were applied.

Specific Phobias and Anxiety Disorders

A total of 25% (5/20) of the studies, with 142 participants, were included in this category, ranging from the treatment of specific phobias to generalized anxiety disorders and social anxiety.

Specific Phobias

Regarding the treatment of specific phobias, Lindner et al [32] studied the effectiveness of VRET in the treatment of arachnophobia. A total of 25 participants (19 female) underwent automated VRET composed of 8 sequential levels with increasingly realistic and frightening spiders. The results showed a significant effect of treatment on phobia symptoms (P<.001; Cohen d=1.26) and a small effect on positive functioning (P<.001; Cohen d=0.49). The results were maintained at the 6-month follow-up. These results suggest that automated VRET applications could be promising automated treatments for this disorder, with a positive effect also on patients' well-being.

The efficacy of VRET was further supported by Rus-Calafell et al [33], who compared VRET with a classic mental imagery exposure therapy to treat fear of flying. In this case, 15 patients (13 female and 2 male) were randomly assigned to 1 of 2 groups and received either VRET (mean age 37.14, SD 14.28 years; sex, 6/7, 86% female) or mental imagery therapy (mean age 36.13, SD 12.59 years). VRET comprised 3 different scenarios (room, airport, and plane). Participants completed 6 individual treatment sessions of 60 to 75 minutes over a period of 3 weeks and were asked to buy a plane ticket to use during the 15 days following the end of treatment. The results indicated that both groups improved similarly in posttreatment and follow-up

assessments. Analysis of the scores obtained immediately after the real flight and 6 months after treatment revealed that the VR group continued to improve on some of the measures (Fear of Flight Questionnaire), whereas the mental imagery group did not; moreover, participants in the VR group experienced less anxiety during the real flight. Regarding positive functioning, the participants in the VR group and the therapists agreed that the degree of severity and interference in the patient's daily life had decreased, whereas in the mental imagery group, patients reported a significant improvement just in the severity of symptoms and not in interference. In summary, both mental imagery exposure therapy and VRET were effective in treating fear of flying. However, VRET was shown to perform better in the maintenance of outcomes and perceived interference of fear in the participants' lives.

Social Anxiety Disorder

A total of 10% (2/20) of the studies included in this review assessed the potential benefits of VRET in the treatment of social anxiety disorder (SAD). In total, 50% (1/2) of these studies adopted a 3-arm randomized controlled approach with 3 different groups. Kampmann et al [34] randomly assigned patients diagnosed with an SAD to individual VRET (n=20), individual in vivo exposure therapy (iVET; n=20), or a waiting list (n=20). The treatments proposed by Kampmann et al [34] comprised ten 90-minute sessions twice a week, and the virtual situations provided for the VRET group covered one-to-one and group scenarios designed to provoke anxiety in individuals with SAD. The results of the study showed an improvement from the pre- to postassessment time points for both VRET and iVET but with a greater decrease in symptoms for iVET than for VRET. This trend was also confirmed after the 3-month follow-up. iVET but not VRET improved the positive functioning of the participants as measured using the EUROHIS Quality of Life Scale; however, this difference was not significant after the 3-month follow-up.

The second study [35] assessed the feasibility and potential effect of VR-based cognitive behavioral therapy (VR-CBT) in patients with SAD without a control group. The authors recruited 15 patients with SAD, who underwent up to 16 VR-CBT sessions. Questionnaires regarding clinical and functional outcomes, as well as diary assessments of social activity, social anxiety, and paranoia, were completed at baseline, posttreatment assessment, and the 6-month follow-up. The treatment comprised a 40-minute VR session, and patients were able to explore different virtual scenarios that the therapist could manipulate in terms of crowdedness, ethnicity, intensity, frequency of hostile looks, interpersonal distances, and watching or staring behavior. The VR exercise allowed the patients to test their beliefs and approach or avoidance behaviors while the therapist provided feedback on cognitions and behaviors. The results showed a general improvement in all the outcome measures—social interaction anxiety was significantly reduced at posttreatment compared with baseline assessment (P=.008; Cohen d=0.9), depression was significantly reduced at posttreatment assessment (P=.01; Cohen d=1.1), and positive functioning (measured using the Manchester Short Assessment of Quality of Life) increased between baseline and posttreatment assessment (P=.02; Cohen d=0.5). This improvement was maintained at follow-up.

Generalized Anxiety Disorder

Malbos et al [30] assessed the efficacy of VR combined with relaxation in patients with generalized anxiety disorder by comparing VR relaxation therapy with a standard mental imagery exposure therapy. A total of 27 participants (13 female) were randomly assigned to VR relaxation therapy or mental imagery. The therapy was delivered in 6 weekly sessions of 30 minutes—the first 3 sessions were used to teach relaxation techniques that the participants were free to choose and repeat in the remaining 3 sessions. At the beginning of each session, VR relaxation therapy participants could select 6 different relaxing virtual situations to exercise the relaxation techniques. The results showed significant improvements in anxiety, worry mood, and positive functioning in both groups, although the 2 groups were not statistically different.

In summary, VRET has proven to be effective, useful, and user-friendly for patients with phobias and anxiety disorders in reducing the related symptomatology. However, further research is needed to draw inferences about the effectiveness of this advanced technology in improving positive functioning as the data regarding this domain are still weak, with 20% (1/5) of the studies included in this category not showing any improvement or additional advantage in promoting well-being for VRET compared with standard in vivo therapies.

PTSD Symptoms

A total of 15% (3/20) of the studies examined the effect of a VR program on posttraumatic stress symptoms, as well as indicators of positive functioning. The first 67% (2/3) of these articles described a single research trial, not published yet. The third article (1/3, 33%) described an intervention aimed at preventing the long-term posttraumatic psychological consequences of being hospitalized in the intensive care unit (ICU).

Jones et al [31] presented the initial data of a new computer-assisted rehabilitation virtual environment developed to treat combat-related PTSD symptoms in the first 11 patients recruited for this RCT study. The multimodal motion-assisted memory desensitization and reconsolidation (3MDR) therapy consists of six 90-minute sessions where participants initially walk on a treadmill (placed inside a virtual room) while listening to self-selected music reminiscent of their military deployment. Then, 1 to 7 images are projected, and the participant has to describe the traumatic scenario as well as the associated physical sensations, emotional words, and thoughts. A ball displaying a series of numbers (which the participants read out loud) briefly appears, moving back and forth horizontally across the screen. This cycle is repeated for all 7 images. Jones et al [31] demonstrated that 3MDR significantly improved PTSD symptoms and emotion regulation strategies (assessed using the Difficulties in Emotion Regulation Scale).

A further pilot study with the same VR intervention was published [36], with a specific focus on the issue of emotion regulation. In this study, 9 participants were interviewed, and the qualitative analyses of these interviews were correlated with

their scores on the Difficulties in Emotion Regulation Scale. After the intervention, participants reported better emotion regulation, with an increased ability to recognize, accept, and cope with emotions (positive and negative).

The third study, conducted by Vlake et al [37], explored the effects of an ICU-specific VR (ICU-VR) program on the psychological distress and quality of life of 89 patients hospitalized in the ICU because of COVID-19 infection. After hospital discharge, they were assigned to the ICU-VR intervention to prevent the onset of possible posttraumatic symptoms. The intervention consisted of watching a 14-minute–long informational video in VR, where the person was welcomed into the ICU and every machine and procedure was described in detail. The results of this investigation showed that ICU-VR improved patients' perceived quality of, satisfaction with, and rating of ICU aftercare and decreased psychological distress up to 6 months after hospital discharge. However, the intervention did not significantly improve psychological recovery or quality of life.

Depression

In total, 10% (2/20) of studies addressing depression were included in this review. They involved 94 patients and applied 2 different VR interventions.

In the first study, Falconer et al [38] addressed self-criticism in a sample of 15 individuals with depression (10 female and 5 male; mean age 32 years) by promoting self-compassion. The authors developed an immersive VR scenario (a virtual room) in which participants could interact compassionately with a crying virtual child while in a virtual adult body. In the second phase of the protocol, participants were in the child's body and could experience a recording of their compassionate gestures and words being delivered to them from this first-person perspective in the body of the child. The same protocol was repeated for a total of 3 weekly sessions. Depressive symptoms, self-compassion and self-criticism were also assessed (Multimedia Appendix 1). By having participants in an adult and then a child virtual body in succession, the authors documented that the VR scenario effectively provided a self-to-self situation, enabling participants to deliver compassionate sentiments and statements to themselves. After each session, patients increased their recognition of the self in the body of the adult and the child, and they reported feeling comforted while in the body of the child. Statistical analyses revealed a significant linear decrease in depressive symptoms (measured using the Patient Health Questionnaire–9) from baseline to follow-up, with over half of the patients reporting reliable levels of improvement. The self-compassion scale demonstrated improvements at postintervention measurement, whereas self-criticism decreased significantly. Considering its brevity (3 sessions), the authors concluded that this VR scenario could be easily integrated into traditional psychotherapies to treat individuals with depression [38].

In the second study included [39], a new VR program—Edge of the Present—was developed and applied to promote optimism and future thinking in 79 individuals with depression (53 female, 23 male, and 3 intersex), with an average age range of 25 to 34 years. They were assessed with a pretest-posttest research design using the PANAS, Short Warwick-Edinburgh Mental Well-being Scale, and Beck Hopelessness Scale. The VR protocol included a single 10-minute session with the Edge of the Present software. It consisted of a sparsely furnished room with doors and windows that participants explored freely. Edge of the Present is designed to reward such exploration with positively experienced imagery. Hence, the door opens onto a series of spectacular immersive landscapes (7 different vistas, including alpine scenes, lush rainforest, tropical beaches, and a desert) accompanied by environmental effects such as a warm breeze, intensifying the sensory experience. The greater the engagement with the room (ie, opening and closing a door or window), the more increasingly enriched the bare room becomes by the outside landscape (ie, ferns growing inside the room). Edge of the Present provokes a sense of hopeful anticipation—each time the door is opened, there is a new landscape for the user to experience and be incorporated into their world (the room). Thus, through their virtual explorations, the user learns both that openness and curiosity lead to positively reinforcing experiences and that elements of these experiences and environments become integrated into the room they are inhabiting (ie, to enrich the internal world of the user). The primary outcome measure in this study was hopelessness as measured using the Beck Hopelessness Scale, which decreased significantly from pre- to postintervention measurement. Changes in positive and negative mood and increases in well-being were measured using the PANAS and the Short Warwick-Edinburgh Mental Well-being Scale. These results suggest that 10 minutes within the immersive virtual environment can have a positive impact on mood and a significant increase in well-being following the participant's involvement in Edge of the Present.

These 2 investigations showed that the VR protocol developed to treat depressive symptoms can also have a beneficial effect on different areas of positive functioning, from self-compassion to positive emotions and subjective well-being, in line with the positive mental health approach.

Psychosis and Schizophrenia

In total, 15% (3/20) of the studies evaluated the benefits of VR treatment in patients with schizophrenia or other psychotic disorders. A total of 150 patients were included, and 3 different VR programs were tested.

du Sert et al [40] used an RCT design to test the effect of a VR intervention using an avatar to treat patients with drug-resistant schizophrenia and compared them with a treatment-as-usual condition. The intervention consisted of 7 weekly sessions; the first one was dedicated to the creation of a specific avatar for each patient that represented their most recurrent persecutor. In sessions 1 to 3, the therapist induced a dialogue between the patients and their avatars to improve emotional regulation and assertiveness. Self-esteem was emphasized in session 4, reinforced by enabling the patients to express themselves and consider their personal qualities. In the final consolidation sessions, patients were encouraged to apply what they had previously learned. Over the course of the therapy, the avatar's interaction with the patient became gradually less abusive and more supportive. The authors assessed both symptoms and

positive functioning before and after the intervention and at follow-up and concluded that the beneficial effect of the VR intervention consisted in changing the way patients relate and respond to their voices by tackling emotional regulation, enhancing self-esteem, and promoting acceptance rather than directly challenging beliefs about the voices [40].

In another RCT, Pot-Kolder et al [41] addressed the issues of paranoid ideation and social functioning in a sample of 116 patients with psychosis; 58 of them were randomized to receive VR-CBT treatment, and 58 were randomized to the waiting list. The VR-CBT consisted of 4 virtual social environments (a street, bus, café, and supermarket) where various avatars were placed to interact with the patients. The therapist could vary the number of human avatars (0-40), the characteristics of the avatars (including sex and ethnicity), and the avatars' responses to the patient (neutral or hostile and eye contact) to match the paranoid fears of the patient. Patients and therapists communicated during VR sessions to explore and challenge suspicious thoughts during social situations, drop safety behaviors during social situations, and test harm expectancies. No homework exercises were given between sessions. The treatment included 16 sessions. The primary outcome of this RCT was patients' social participation, which did not change significantly after treatment. The same nonsignificant change was also observed for positive functioning, but social functioning and paranoid ideation significantly improved in the VR-CBT condition compared with the waiting list [41].

Similarly, the third investigation addressed improving social-cognitive functioning in 19 patients in the early stages of psychosis. It was a pilot study in which a virtual world environment platform (*Second Life*) was used to adapt a traditionally face-to-face–delivered social cognition and interactional training. The social cognition and interactional training–VR intervention consisted of 10 sessions (2 individual and 8 group sessions with 3-5 participants). The first 3 sessions focused primarily on emotion recognition, the next 3 sessions focused on attribution bias and paranoia as an emotion, and the last 2 sessions focused on "skills acquisition" using a cognitive behavioral therapy (CBT) framework to discuss examples of social difficulties faced by the participant. After the intervention, a significant increase in emotion recognition and a significant decrease in the anxiety and depression subscale of the EQ-5D were observed. This is the first study to use a virtual world to deliver structured group therapy in early psychosis with a specific focus on social interactions. It documented a positive effect with good feasibility and acceptability from participants [42].

In summary, the 3 VR protocols applied in the treatment of patients with psychosis demonstrated to have a beneficial effect on their paranoid symptoms and also a positive effect on self-esteem, acceptance, and emotion regulation strategy. However, the positive effect on social interactions in the real world has yet to be confirmed.

Stress and Adjustment Disorder

A total of 35% (7/20) of protocols using VR interventions for addressing stress were included in this review, with a total of 403 participants.

The first protocol is the one used by Gaggioli et al [43] in the context of stress-associated disorders. The authors evaluated the effectiveness of an interreality protocol for the prevention and management of psychological stress compared with stress management training based on CBT and a waitlist control group in a sample of 121 workers (61 high school teachers and 60 nurses) with high levels of perceived stress and low levels of self-efficacy. The interreality protocol included virtual experiences—controlled by the therapist and focused on learning coping skills and improving self-efficacy—and specific real-world experiences where the person's behaviors and emotions were constantly monitored through the use of wearable biosensors and smartphones to assess the situation and improve the coping skills used in real time. After a 5-week treatment (10 sessions, 2 times per week), reductions in perceived stress and improvements in coping skills were observed in both conditions (interreality and CBT); however, the use of interreality resulted in a significant reduction in chronic "trait" anxiety and a significantly greater increase in emotional support skills compared with CBT (for more details on the outcome measures, see Multimedia Appendix 1).

The second protocol identified is the one used in the study by Cebolla et al [44] in compassion-based interventions. The study involved 16 college students (mean age 30.56, SD 10.86 years) who participated in a self-compassion meditation supported or not (control condition) by an embodied VR system. The Machine To Be Another (TMTBA) is a VR system based on a body swap, and this study involved the use of a visor (VR Oculus Rift) that remotely controlled a camera placed in front of the torso of an actor trained to imitate the performer's precise movements. The system allows the person to identify with another person's body, whose torso, legs, and arms they can see. In the TMTBA-VR condition, the self-compassion meditation was played to the participant for 15 minutes, allowing a third-person perspective of oneself. The study showed that following a 2-week meditative practice yielded increased awareness and attention to mental events and bodily sensations, with no differences between the groups. Furthermore, the VR condition yielded an increase in positive affect toward oneself and self-care behaviors, which were significantly higher than those in the traditional meditation condition. Finally, in the TMTBA condition, adherence to meditative practice after 2 weeks was higher in participants with lower visual imagery. According to the authors, these data emphasize the important role that VR could play in psychological interventions where mental imagery is often used (see Multimedia Appendix 1 for more details on the outcomes measured). The results obtained by Cebolla et al [44] showed the positive role of VR in compassion-based interventions, emphasizing the need for future studies on this topic.

VR was also found to be effective in managing stress without the guidance of a counselor, as demonstrated by the Digital–Structured Association Technique (SAT) method developed and evaluated by Matsumoto et al [29]. The Digital-SAT method is an adaptation of a stress management technique used in SAT therapy and consists of re-enacting a stressful physiological response and then counterbalancing it with the visualization of pleasant images using VR technology,

which leads to a reduction in discomfort. The intervention protocol involved the combined use of 2 apps to implement the Digital-SAT method: the VR app, aimed at reducing emotional stress by displaying pleasant images through an HMD, and the chat bot app, aimed at improving the continuity of the intervention through an automated chatbot. Matsumoto et al [29] found that the combined use of the VR app and the chat bot app produced a better emotional stress reduction effect after 4 weeks, and they encouraged the continuous implementation of the intervention in a sample of 70 nurses exposed to mental and physical strain and pressure (see Multimedia Appendix 1 for detailed outcome measures). The Digital-SAT method can be considered a promising tool for improving self-care autonomously even in work and daily life settings as it does not necessarily require the involvement of a therapist.

Another area where the application of VR aims to promote positive functioning and manage stress is the context of art. In this regard, Richesin et al [45] compared 2D and 3D art making on measures of stress, anxiety, and mood in a sample of 44 psychology students. Specifically, the 3D intervention involved the use of an HMD (Oculus Quest) and a drawing app (Google Tilt Brush), which participants used to draw freely for a period of 15 minutes using any available tools. This condition was compared with a traditional art-making intervention, and the main outcomes of this study were physiological and self-report measures of stress, anxiety, and well-being (see Multimedia Appendix 1 for more details on the outcome measurements). The study showed that drawing freely using VR can lead to a reduction in negative affect and anxiety levels (state and trait) similar to the effect of the traditional method of art making, but only the 3D group showed a significantly greater decrease in heart rate than the control group. The authors suggested that VR could be a useful tool in art therapy, but the results may be a consequence of the novelty effect.

Another 15% (3/20) of studies examined the effectiveness of VR in stress management during the COVID-19 pandemic. The study by Chan et al [46] involved 2 experiments that explored and compared the effects of VR nature and VR urban environments on the mental condition of individuals who experienced isolation because of COVID-19. The protocol consisted of 7-minute sessions of exposure to both urban and natural environments once per week. A total of 30 students participated in the first study, and 20 older adults participated in the second one. Their positive and negative affect, as well as their connectedness with nature, were assessed before and after treatment. The results from the student and older adult groups showed that walking in a virtual forest decreased stress and negative affect owing to its connection with nature, although no effect was observed on positive affect.

The study designed by Desai et al [47] explored whether using a virtual heart-based meditation program was associated with the improvement of stress levels and quality of sleep during the COVID-19 pandemic. A total of 63 participants underwent 1 weekly virtual trainer-guided group heartfulness relaxation and meditation session for 8 weeks while self-practice was recommended. The results showed a decrease in stress, assessed using the Perceived Stress Scale, and improvements in sleep quality, assessed using the Pittsburgh Sleep Quality Index.

The aim of the last study (Riva et al [48]) was to evaluate the effectiveness of a VR protocol (*the "Secret Garden" 360-degree VR experience*) to assist individuals in coping with the psychological burden related to the COVID-19 pandemic, improving their well-being, and reinforcing social connectedness. In each session, 40 participants who had experienced at least 2 months of quarantine could travel through a virtual garden for 10 minutes for a total of 7 sessions once per week. The results showed that the intervention was associated with improvements in depression and stress but not in perceived hopelessness. There was also a significant increase in social connectedness.

In summary, VR protocols were effective in reducing stress and promoting self-care and positive coping skills also without the presence of a therapist. However, the VR protocols and assessment measures used were very heterogeneous.

Discussion

Principal Findings

This review aimed to collect and summarize the literature on the application of positive mental health interventions integrated with VR that yielded beneficial effects for both symptoms and indicators of positive functioning. The first observation derived from this review is the paucity of studies that assessed both distress and well-being despite the increasing number of published articles in recent years (Table 1) and the increasing use of digital technologies in mental health practice. A possible explanation for this phenomenon is that positive outcomes are not widely evaluated in research and clinical practice compared with the assessment of symptoms or indicators of distress. This may represent an existing imbalance between the traditional and the positive psychology perspective [7,9] or simply the need to limit the survey burden and time dedicated to assessment in current research. In any case, we included only 20 investigations over 10 years of research.

Starting with anxiety and phobias, the use of VR in this field mainly involves traditional exposure therapy delivered through VR devices (VRET). Although exposure therapy is a traditional and effective method to treat anxiety disorders, we included studies that applied this technique supported by VR devices as exposure implies behavior changes and opportunities for learning new skills, which are strongly associated with increases in self-efficacy and mastery [51]. Various authors have suggested that exposure therapy may indirectly promote positive characteristics such as courage, persistence, goal setting, and planning and could trigger positive affect, a sense of pride, and satisfaction once the exposure activity is successfully performed [6,24,52]. Thus, it can be included under the wide umbrella of positive interventions [7,9]. Exposure therapy is generally delivered in the presence of a clinician, but 20% (1/5) of the included studies [32] used an automated exposure protocol. All the included protocols (5/5, 100%) were effective in reducing the symptomatology of the patients, but they were not equally effective in improving their positive functioning. The results of the study by Rus-Calafell et al [33] showed an improvement in the interference of the phobic object in life only for the VR therapy compared with the imagery control condition, whereas

Kampmann et al [34] found that VR therapy did not improve positive functioning at all as compared with a waitlist control. Finally, Malbos et al [30] showed that VR relaxation improved positive functioning but not more than a classic imagery intervention. From this picture, VRET seems to be effective and useful for patients in reducing the symptomatology related to phobias and anxiety disorders (see also the studies by Baghaei et al [1], Schroeder et al [2], and van Loenen et al [3]). However, further research is needed to draw inferences about the effectiveness of this advanced technology in improving positive functioning as the data regarding this domain are still inconclusive.

The same inferences can also be drawn for other mental health problems analyzed in this review. For instance, for PTSD, we only found 15% (3/20) of studies that assessed positive functioning (ie, emotional regulation) together with symptomatology. In total, 67% (2/3) of these studies referred to the same intervention protocol, consisting of exploring a virtual room on a treadmill followed by the presentation of 7 images associated with 3MDR, which can be considered a technologically advanced form of traditional trauma intervention—exposure therapy and eye movement desensitization and reconsolidation [53]. The results of this VR intervention were promising—PTSD symptoms improved, together with better use of emotion regulation strategies. Participants reported a better ability to recognize, accept, and cope with emotions after treatment. However, these studies involved only a few patients who were treated with one specific VR protocol. Similarly, the results of the study by Vlake et al [37] showed a decrease in distress among COVID-19 ICU survivors but not a significant improvement in the positive functioning of the patients.

The studies on depression were based on different types of protocols, but also in this case, the findings regarding improvements in positive functioning are inconclusive. One of the studies focused on self-compassion, and participants were required to compassionately interact with an avatar first and then re-experience the effect of their interaction by taking on the role of the child avatar. In the second study, participants had to explore a virtual room, and such exploration was rewarded with positive experiences as the virtual doors opened onto a series of spectacular immersive landscapes. Both studies showed promising results and highlighted the benefits of the protocols in decreasing the severity of depressive symptoms. However, concerning positive functioning, the results of the first treatment showed that the protocol did not yield any change in self-compassion, only in self-criticism, whereas the second intervention yielded an improvement in individuals' positive affect and a significant increase in well-being following the participants' involvement in the virtual procedure. The first investigation contributes important data on the crucial differentiation between symptom improvements (ie, a decrease in self-criticism) and the promotion of positive functioning (ie, an increase in self-compassion). These findings emphasize the need to implement interventions specifically focused on well-being promotion and not merely on symptom reduction [6,7,54,55].

The second VR intervention, in contrast, shares the same theoretical framework of traditional behavior activation therapy, where positive and rewarding activities are prescribed to help patients with depression experience pleasure and engagement in their lives [56,57]. Hence, both investigations confirm the observation that depressive disorders can be addressed by changing the complex balance between positive and negative affect [58,59]. VR may play an important role in this regard, but further studies are needed to develop the most suitable protocols for enhancing well-being in individuals with depression.

VR treatments for individuals with psychosis and schizophrenia were focused on social functioning and positive relationships. A total of 15% (3/20) of the studies included in this review were based on social interaction with one or more avatars to improve social interactions and cognitive functioning. du Sert et al [40] showed that the VR intervention was effective in changing the way patients related and responded to their voices by tackling emotional regulation, enhancing self-esteem, and promoting acceptance as compared with treatment as usual. These results are in line with those of Thompson et al [42], who found a significant increase in emotion recognition and a significant decrease in anxiety and depression in their pilot study. In contrast with these results, Pot-Kolder et al [41] did not register a significant change in patients' social participation and positive social functioning compared with the waitlist condition. Only the dimensions of social functioning and paranoid ideation improved when compared with the control group. However, it should be noted that the authors used the same VR intervention that Geraets et al [35] used in the treatment of individuals with SAD, and in this case, the intervention was able to show an improvement in social functioning. When the protocol was delivered to patients with psychosis, findings showed that the intervention based on a one-to-one or group interaction with an avatar yielded general beneficial effects on paranoia and other psychotic symptoms but not necessarily on the social participation and positive interpersonal functioning of patients (see the study by Schroeder et al [2] for a review). Further research is needed to clarify whether the beneficial effect of the VR intervention might be linked to the severity of the clinical condition of participants, their cognitive bias, their preexisting social skills, or other peculiar issues yet to be investigated [60,61].

Outside the domain of psychiatric conditions, this systematic review analyzed VR interventions aimed at promoting well-being and reducing stress, including COVID-19 pandemic distress. As many protocols involving the use of VR are being developed in this field, a summary of their findings might be useful to understand whether these interventions were beneficial in decreasing the stress and anxiety related to the virus and social isolation and in restoring well-being after the pandemic. In total, 10% (2/20) of the included studies [46,48] were based on a virtual natural scenario that the participant could explore, whereas another protocol reproduced an ICU in which patients were hospitalized because of COVID-19. The interventions with virtual natural scenarios were shown to be effective in reducing stress, negative affect, and depression and increasing social connectedness, in line with the well-established

association between feelings of well-being and being in contact with nature (in this case, through a virtual landscape) [62-64]. However, the protocol by Chan et al [46] did not yield an improvement in positive affect, and the one by Riva et al [48] was not effective in improving perceived hopelessness. Finally, Desai et al [47] proposed a heart-based relaxation protocol that showed a decrease in stress and a better quality of sleep. These data support the use of VR in addressing the stress and negative feelings associated with the pandemic, but its beneficial effects for restoring well-being have not been confirmed, with the sole exceptions of social connectedness and sleep quality.

Other than stress related to COVID-19 and the pandemic, VR interventions also showed their utility in relieving and preventing stress more generally. Gaggioli et al [43] showed that a VR exposure-based therapy prompted a significant reduction in perceived stress and improvements in coping skills in workers. Moreover, the group that underwent the VR intervention showed a reduction in chronic "trait" anxiety and a significantly greater increase in emotional support skills. Compassion-based interventions have also been developed and delivered using VR, and the study by Cebolla et al [44] showed that this intervention was effective in increasing positive attitudes toward the self and others and decreasing negative qualities toward the self. However, the data do not fully support an additional benefit of using VR compared with traditional meditation techniques. Finally, 10% (2/20) of the studies included in this review showed that VR interventions were effective in reducing emotional stress also without the guidance of a counselor [29] and in reducing negative affect and anxiety levels (state and trait) using virtual art therapy [45]. In summary, when considering nonclinical populations such as workers or college students, VR interventions showed their utility and effectiveness in reducing stress and promoting positive function (ie, emotional support skills and compassion toward others and the self) even without the guidance of a therapist or counselor. These data are very promising considering the large scalability of these VR protocols after the pandemic and among the general population [65]. However, solid evidence that VR interventions are more effective compared with traditional ones (ie, mindfulness, traditional CBT, relaxation, and art making) is still to be completely demonstrated.

Limitations

One of the main limitations of this review was its heterogeneous nature. It included different populations (clinical and nonclinical) treated with many different VR protocols and evaluated using different outcome measures. This is particularly true for the assessment of positive functioning, where no consensus emerged among investigations with the sole exceptions of the PANAS and the EQ-5 used EUROHIS Quality of Life Scale in 15% (3/20) of the studies. Hence, there were not enough quantitative data to support a meta-analysis. Accordingly, this work aims to provide a summary of the progress made in the last decade and delineate future directions to follow for researchers and clinicians interested in promoting positive mental health through VR interventions.

Another limitation of this study was the absence or exclusions of specific clinical domains (such as OCD, eating disorders, and addiction-related disorders), where many other VR protocols have been recently developed [11,12]. For OCD, we found no studies that assessed positive functioning along with negative symptoms. We decided not to include eating disorders and addiction-related disorders as the positive functioning of these clinical populations was found to be intertwined with their disorders [23,25,26]. Thus, when analyzing the effects of VR interventions, these clinical conditions might need to be considered from a different perspective.

Finally, the studies included in this review account only for adult populations, leaving out children and older individuals. The reason for this choice lies in the fact that these populations present specific features in terms of mental health; in older adults, core issues are correlated with memory loss and physical and cognitive decline, whereas in children and adolescents, issues such as self-esteem, identity, and personality development are the main challenges associated with distress and well-being. Moreover, the features of psychological distress and well-being in these populations are strongly influenced by age and the stage of development or aging [66,67]. We found that many studies that used VR to promote the mental health of children and older adults [18-21] focused on aspects that differed from those in studies on adults, such as neurological and cognitive development or decline, respectively.

Conclusions and Future Directions

In conclusion, this review provides robust evidence supporting the beneficial effect of VR therapy in improving stress and negative symptoms. However, in participants, the impact of VR treatments on positive functioning remains unclear, with 35% (7/20) of the studies showing no or a small effect on various dimensions of positivity, ranging from positive affect to relaxation, self-compassion, and social interaction. The variety of outcome measures included and their different sensitivities to clinical changes might be the reason why the results on the beneficial effect of VR interventions in promoting positive functioning are still inconclusive [68]. It is of crucial importance that future research systematically addresses the impact of VR interventions also on individuals' positive functioning to identify the most suitable protocols for enhancing well-being in individuals with different mental health conditions (from those with psychiatric disorders to those dealing with stress or the negative consequences of the recent COVID-19 pandemic). Indeed, our review showed that VR interventions applied in the general population were effective both in addressing stress and promoting positive mood, social connectedness, and better sleep quality. These data suggest that VR interventions might be cost-effective and largely scalable, particularly when they do not require the guidance of a counselor or therapist.

However, this review also suggested that the beneficial effect of VR interventions might be linked to the severity of the clinical condition of participants (ie, the same VR protocol with the interaction with an avatar was able to improve social functioning in individuals with social anxiety but not in patients with paranoia). Previous work has documented the peculiar combinations of personal resources and vulnerabilities that characterize clinical populations [23,25,66]. In particular, it was found that clinical populations generally present impairments

in well - being, and positive interventions developed within the positive psychology perspective were found to be effective in addressing those impairments [7,28]. Owing to its flexibility and capacity to engage participants, VR interventions might play a crucial role in promoting better mental health in the population, including clinical samples. Considering the large number of private and public investments in digital technologies for mental health, this review suggests further research to develop existing VR software and treatments to align them with the modern positive mental health approach.

Data Availability

Data sharing does not apply to this paper as no new data were created or analyzed in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of the studies included in the review.
[DOCX File , 33 KB - mental_v10i1e44998_app1.docx]

References

1. Baghaei N, Stemmet L, Khaliq I, Ahmadi A, Halim I, Liang HN, et al. Designing individualised virtual reality applications for supporting depression: a feasibility study. In: Proceedings of the Companion of the 2021 ACM SIGCHI Symposium on Engineering Interactive Computing Systems. 2021 Presented at: EICS '21; Jun 8-11, 2021; Virtual Event p. 6-11 URL: https://dl.acm.org/doi/abs/10.1145/3459926.3464761 [doi: 10.1145/3459926.3464761]
2. Schroeder AH, Bogie BJ, Rahman TT, Thérond A, Matheson H, Guimond S. Feasibility and efficacy of virtual reality interventions to improve psychosocial functioning in psychosis: systematic review. JMIR Ment Health 2022 Feb 18;9(2):e28502. [doi: 10.2196/28502] [Medline: 35179501]
3. van Loenen I, Scholten W, Muntingh A, Smit J, Batelaan N. The effectiveness of virtual reality exposure-based cognitive behavioral therapy for severe anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder: meta-analysis. J Med Internet Res 2022 Feb 10;24(2):e26736 [FREE Full text] [doi: 10.2196/26736] [Medline: 35142632]
4. Lindner P. Better, virtually: the past, present, and future of virtual reality cognitive behavior therapy. Int J Cogn Ther 2021;14(1):23-46 [FREE Full text] [doi: 10.1007/s41811-020-00090-7]
5. Keyes CL. The mental health continuum: from languishing to flourishing in life. J Health Soc Behav 2002 Jun;43(2):207-222 [FREE Full text] [doi: 10.2307/3090197]
6. Ruini C. Positive Psychology in the Clinical Domains Research and Practice. Cham, Switzerland: Springer; 2017.
7. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
8. Keyes CL, Dhingra SS, Simoes EJ. Change in level of positive mental health as a predictor of future risk of mental illness. Am J Public Health 2010 Dec;100(12):2366-2371. [doi: 10.2105/AJPH.2010.192245] [Medline: 20966364]
9. Wood AM, Tarrier N. Positive clinical psychology: a new vision and strategy for integrated research and practice. Clin Psychol Rev 2010 Nov;30(7):819-829. [doi: 10.1016/j.cpr.2010.06.003] [Medline: 20655136]
10. Kashdan TB, Rottenberg J. Psychological flexibility as a fundamental aspect of health. Clin Psychol Rev 2010 Nov;30(7):865-878 [FREE Full text] [doi: 10.1016/j.cpr.2010.03.001] [Medline: 21151705]
11. Asiain J, Braun M, Roussos AJ. Virtual reality as a psychotherapeutic tool: current uses and limitations. Br J Guid Couns 2021 Apr 28;50(1):1-28 [FREE Full text] [doi: 10.1080/03069885.2021.1885008]
12. Cieślik B, Mazurek J, Rutkowski S, Kiper P, Turolla A, Szczepańska-Gieracha J. Virtual reality in psychiatric disorders: a systematic review of reviews. Complement Ther Med 2020 Aug;52:102480 [FREE Full text] [doi: 10.1016/j.ctim.2020.102480] [Medline: 32951730]
13. Riva G. What is Positive Technology and its impact on cyberpsychology. Stud Health Technol Inform 2012;181:37-41. [Medline: 22954824]
14. Gorini A, Pallavicini F, Algeri D, Repetto C, Gaggioli A, Riva G. Virtual reality in the treatment of generalized anxiety disorders. Stud Health Technol Inform 2010;154:39-43. [Medline: 20543266]
15. Riva G. Virtual reality in psychotherapy: review. Cyberpsychol Behav 2005 Jun;8(3):220-240. [doi: 10.1089/cpb.2005.8.220] [Medline: 15971972]
16. Botella C, Riva G, Gaggioli A, Wiederhold BK, Alcaniz M, Baños RM. The present and future of positive technologies. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):78-84 [FREE Full text] [doi: 10.1089/cyber.2011.0140] [Medline: 22149078]
17. Kim J, Jung YH, Shin YB, Kim MK, Eom H, Kim E, et al. Development and validation of a virtual reality-based training program for promoting subjective well-being. Psychiatry Investig 2020 Dec;17(12):1207-1215 [FREE Full text] [doi: 10.30773/pi.2020.0311] [Medline: 33301665]

18. Babbage CM, Jackson GM, Davies EB, Nixon E. Self-help digital interventions targeted at improving psychological well-being in young people with perceived or clinically diagnosed reduced well-being: systematic review. JMIR Ment Health 2022 Aug 26;9(8):e25716 [FREE Full text] [doi: 10.2196/25716] [Medline: 36018675]
19. Botella C, Etchemendy E, Castilla D, Baños RM, García-Palacios A, Quero S, et al. An e-health system for the elderly (Butler Project): a pilot study on acceptance and satisfaction. Cyberpsychol Behav 2009 Jun;12(3):255-262. [doi: 10.1089/cpb.2008.0325] [Medline: 19445633]
20. Thangavel G, Memedi M, Hedström K. Customized information and communication technology for reducing social isolation and loneliness among older adults: scoping review. JMIR Ment Health 2022 Mar 07;9(3):e34221 [FREE Full text] [doi: 10.2196/34221] [Medline: 35254273]
21. Lluesma-Vidal M, Carcelén González R, García-Garcés L, Sánchez-López MI, Peyro L, Ruiz-Zaldibar C. Effect of virtual reality on pediatric pain and fear during procedures involving needles: systematic review and meta-analysis. JMIR Serious Games 2022 Aug 09;10(3):e35008 [FREE Full text] [doi: 10.2196/35008] [Medline: 35943776]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement (Chinese edition). J Chin Integr Med 2009;7(9):889-896 [FREE Full text] [doi: 10.3736/jcim20090918]
23. Hall-Simmonds A, McGrath RE. Character strengths and clinical presentation. J Posit Psychol 2017 Aug 20;14(1):51-60. [doi: 10.1080/17439760.2017.1365160]
24. Johnson J, Wood AM. Integrating positive and clinical psychology: viewing human functioning as continua from positive to negative can benefit clinical assessment, interventions and understandings of resilience. Cogn Ther Res 2015 Oct 6;41(3):335-349. [doi: 10.1007/s10608-015-9728-y]
25. Brandel M, Vescovelli F, Ruini C. Beyond Ryff's scale: comprehensive measures of eudaimonic well-being in clinical populations. A systematic review. Clin Psychol Psychother 2017 Nov 02;24(6):O1524-O1546. [doi: 10.1002/cpp.2104] [Medline: 28770559]
26. Lamers SM, Westerhof GJ, Glas CA, Bohlmeijer ET. The bidirectional relation between positive mental health and psychopathology in a longitudinal representative panel study. J Posit Psychol 2015 May 27;10(6):553-560. [doi: 10.1080/17439760.2015.1015156]
27. Bohart AC. Focusing on the positive, focusing on the negative: implications for psychotherapy. J Clin Psychol 2002 Sep;58(9):1037-1043. [doi: 10.1002/jclp.10097] [Medline: 12209863]
28. Weiss LA, Westerhof GJ, Bohlmeijer ET. Can we increase psychological well-being? The effects of interventions on psychological well-being: a meta-analysis of randomized controlled trials. PLoS One 2016 Jun 21;11(6):e0158092 [FREE Full text] [doi: 10.1371/journal.pone.0158092] [Medline: 27328124]
29. Matsumoto A, Kamita T, Tawaratsumida Y, Nakamura A, Fukuchimoto H, Mitamura Y, et al. Combined use of virtual reality and a chatbot reduces emotional stress more than using them separately. J Univ Comput Sci 2021 Dec 28;27(12):1371-1389. [doi: 10.3897/jucs.77237]
30. Malbos E, Borwell B, Lançon C, Weindel G, Chichery N, Seimandi J. Virtual reality for relaxation in the treatment of generalized anxiety disorder: a comparative trial. Annual Rev CyberTherapy Telemed 2020 Jun;18:206 [FREE Full text]
31. Jones C, Bremault-Phillips S, Smith-MacDonald L, Brown M. Tackling trauma with technology: treating chronic combat-related PTSD in Canadian Armed Forces personnel and veterans with 3MDR. In: Proceedings of the 25th Annual International CyberPsychology, CyberTherapy & Social Networking Conference (CYPSY25). 2021 Presented at: CYPSY '21; September 13-15, 2021; Virtual Event.
32. Lindner P, Miloff A, Bergman C, Andersson G, Hamilton W, Carlbring P. Gamified, automated virtual reality exposure therapy for fear of spiders: a single-subject trial under simulated real-world conditions. Front Psychiatry 2020 Mar 3;11:116 [FREE Full text] [doi: 10.3389/fpsyt.2020.00116] [Medline: 32210850]
33. Rus-Calafell M, Gutiérrez-Maldonado J, Botella C, Baños RM. Virtual reality exposure and imaginal exposure in the treatment of fear of flying: a pilot study. Behav Modif 2013 Jul;37(4):568-590. [doi: 10.1177/0145445513482969] [Medline: 23585557]
34. Kampmann IL, Emmelkamp PM, Hartanto D, Brinkman WP, Zijlstra BJ, Morina N. Exposure to virtual social interactions in the treatment of social anxiety disorder: a randomized controlled trial. Behav Res Ther 2016 Feb;77:147-156. [doi: 10.1016/j.brat.2015.12.016] [Medline: 26752328]
35. Geraets CN, Veling W, Witlox M, Staring AB, Matthijssen SJ, Cath D. Virtual reality-based cognitive behavioural therapy for patients with generalized social anxiety disorder: a pilot study. Behav Cogn Psychother 2019 Mar 27;47(6):745-750. [doi: 10.1017/s1352465819000225]
36. Tang E, Jones C, Smith-MacDonald L, Brown MR, Vermetten EH, Brémault-Phillips S. Decreased emotional dysregulation following multi-modal motion-assisted memory desensitization and reconsolidation therapy (3MDR): identifying possible driving factors in remediation of treatment-resistant PTSD. Int J Environ Res Public Health 2021 Nov 22;18(22):12243 [FREE Full text] [doi: 10.3390/ijerph182212243] [Medline: 34831999]
37. Vlake JH, van Bommel J, Wils EJ, Bienvenu J, Hellemons ME, Korevaar TI, et al. Intensive care unit-specific virtual reality for critically ill patients with COVID-19: multicenter randomized controlled trial. J Med Internet Res 2022 Jan 31;24(1):e32368 [FREE Full text] [doi: 10.2196/32368] [Medline: 34978530]

38. Falconer CJ, Rovira A, King JA, Gilbert P, Antley A, Fearon P, et al. Embodying self-compassion within virtual reality and its effects on patients with depression. BJPsych Open 2016 Feb 15;2(1):74-80 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002147] [Medline: 27703757]
39. Habak S, Bennett J, Davies A, Davies M, Christensen H, Boydell KM. Edge of the present: a virtual reality tool to cultivate future thinking, positive mood and wellbeing. Int J Environ Res Public Health 2020 Dec 28;18(1):140 [FREE Full text] [doi: 10.3390/ijerph18010140] [Medline: 33379156]
40. du Sert OP, Potvin S, Lipp O, Dellazizzo L, Laurelli M, Breton R, et al. Virtual reality therapy for refractory auditory verbal hallucinations in schizophrenia: a pilot clinical trial. Schizophr Res 2018 Jul;197:176-181. [doi: 10.1016/j.schres.2018.02.031] [Medline: 29486956]
41. Pot-Kolder RM, Geraets CN, Veling W, van Beilen M, Staring AB, Gijsman HJ, et al. Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: a single-blind randomised controlled trial. Lancet Psychiatry 2018 Mar;5(3):217-226. [doi: 10.1016/S2215-0366(18)30053-1] [Medline: 29429948]
42. Thompson A, Elahi F, Realpe A, Birchwood M, Taylor D, Vlaev I, et al. A feasibility and acceptability trial of social cognitive therapy in early psychosis delivered through a virtual world: the VEEP study. Front Psychiatry 2020 Mar 25;11:219 [FREE Full text] [doi: 10.3389/fpsyt.2020.00219] [Medline: 32269534]
43. Gaggioli A, Pallavicini F, Morganti L, Serino S, Scaratti C, Briguglio M, et al. Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. J Med Internet Res 2014 Jul 08;16(7):e167 [FREE Full text] [doi: 10.2196/jmir.3235] [Medline: 25004803]
44. Cebolla A, Herrero R, Ventura S, Miragall M, Bellosta-Batalla M, Llorens R, et al. Putting oneself in the body of others: a pilot study on the efficacy of an embodied virtual reality system to generate self-compassion. Front Psychol 2019 Jul 2;10:1521 [FREE Full text] [doi: 10.3389/fpsyg.2019.01521] [Medline: 31338048]
45. Richesin MT, Baldwin DR, Wicks LA. Art making and virtual reality: a comparison study of physiological and psychological outcomes. Arts Psychother 2021 Sep;75:101823 [FREE Full text] [doi: 10.1016/j.aip.2021.101823]
46. Chan SH, Qiu L, Esposito G, Mai KP, Tam KP, Cui J. Nature in virtual reality improves mood and reduces stress: evidence from young adults and senior citizens. Virtual Real 2021 Nov 26:1-16 [FREE Full text] [doi: 10.1007/s10055-021-00604-4] [Medline: 34849087]
47. Desai K, Gupta P, Parikh P, Desai A. Impact of virtual heartfulness meditation program on stress, quality of sleep, and psychological wellbeing during the COVID-19 pandemic: a mixed-method study. Int J Environ Res Public Health 2021 Oct 22;18(21):11114 [FREE Full text] [doi: 10.3390/ijerph182111114] [Medline: 34769634]
48. Riva G, Bernardelli L, Castelnuovo G, Di Lernia D, Tuena C, Clementi A, et al. A virtual reality-based self-help intervention for dealing with the psychological distress associated with the COVID-19 lockdown: an effectiveness study with a two-week follow-up. Int J Environ Res Public Health 2021 Aug 02;18(15):8188 [FREE Full text] [doi: 10.3390/ijerph18158188] [Medline: 34360479]
49. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
50. Kmet LM, Lee RC, Alberta Heritage Foundation for Medical Research, Cook LS. Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields. Edmonton, Canada: Alberta Heritage Foundation for Medical Research; 2004.
51. Maddux JE, editor. Self-Efficacy, Adaptation, and Adjustment Theory, Research, and Application. Cham, Switzerland: Springer; 1995.
52. Breuninger C, Tuschen-Caffier B, Svaldi J. Dysfunctional cognition and self-efficacy as mediators of symptom change in exposure therapy for agoraphobia - Systematic review and meta-analysis. Behav Res Ther 2019 Sep;120:103443. [doi: 10.1016/j.brat.2019.103443] [Medline: 31374484]
53. Foster SL. Integrating positive psychology applications into the EMDR Peak Performance protocol. Eur Rev Appl Psychol 2012 Oct;62(4):213-217. [doi: 10.1016/j.erap.2012.04.005]
54. Lyubomirsky S, Sheldon KM, Schkade D. Pursuing happiness: the architecture of sustainable change. Rev General Psychol 2005 Jun 01;9(2):111-131. [doi: 10.1037/1089-2680.9.2.111]
55. Sin NL, Della Porta MD, Lyubomirsky S. Tailoring positive psychology interventions to treat depressed individuals. In: Donaldson SI, Csikszentmihalyi M, Nakamura J, editors. Applied Positive Psychology: Improving Everyday Life, Health, Schools, Work, and Society. Oxfordshire, UK: Routledge; 2011.
56. Layous K, Chancellor J, Lyubomirsky S. Positive activities as protective factors against mental health conditions. J Abnorm Psychol 2014 Feb;123(1):3-12 [FREE Full text] [doi: 10.1037/a0034709] [Medline: 24661154]
57. MacLeod AK, Luzon O. The place of psychological well-being in cognitive therapy. In: Fava GA, Ruini C, editors. Increasing Psychological Well-being in Clinical and Educational Settings Interventions and Cultural Contexts. Dordrecht, the Netherlands: Springer; 2014.
58. Fava GA, Ruini C. Development and characteristics of a well-being enhancing psychotherapeutic strategy: well-being therapy. J Behav Ther Exp Psychiatry 2003 Mar;34(1):45-63. [doi: 10.1016/s0005-7916(03)00019-3] [Medline: 12763392]

59. Ruini C, Cesetti G. Spotlight on eudaimonia and depression. A systematic review of the literature over the past 5 years. Psychol Res Behav Manag 2019 Aug;Volume 12:767-792. [doi: 10.2147/prbm.s178255]
60. Broyd A, Jolley S, Johns L. Determinants of subjective well-being in people with psychosis referred for psychological therapy in South London. Br J Clin Psychol 2016 Nov 29;55(4):429-440 [FREE Full text] [doi: 10.1111/bjc.12112] [Medline: 27126303]
61. Riches S, Schrank B, Rashid T, Slade M. WELLFOCUS PPT: modifying positive psychotherapy for psychosis. Psychotherapy (Chic) 2016 Mar;53(1):68-77. [doi: 10.1037/pst0000013] [Medline: 25961372]
62. Baños RM, Etchemendy E, Castilla D, García-Palacios A, Quero S, Botella C. Positive mood induction procedures for virtual environments designed for elderly people. Interact Comput 2012 May;24(3):131-138. [doi: 10.1016/j.intcom.2012.04.002]
63. Brooks AM, Ottley KM, Arbuthnott KD, Sevigny P. Nature-related mood effects: season and type of nature contact. J Environ Psychol 2017 Dec;54:91-102. [doi: 10.1016/j.jenvp.2017.10.004]
64. Browning MH, Mimnaugh KJ, van Riper CJ, Laurent HK, LaValle SM. Can simulated nature support mental health? Comparing short, single-doses of 360-degree nature videos in virtual reality with the outdoors. Front Psychol 2020 Jan 15;10:2667 [FREE Full text] [doi: 10.3389/fpsyg.2019.02667] [Medline: 32010003]
65. Bell IH, Nicholas J, Alvarez-Jimenez M, Thompson A, Valmaggia L. Virtual reality as a clinical tool in mental health research and practice. Dialogues Clin Neurosci 2020 Jun;22(2):169-177 [FREE Full text] [doi: 10.31887/DCNS.2020.22.2/lvalmaggia] [Medline: 32699517]
66. Ryff CD. Psychological well-being revisited: advances in the science and practice of eudaimonia. Psychother Psychosom 2014;83(1):10-28 [FREE Full text] [doi: 10.1159/000353263] [Medline: 24281296]
67. Owens RL, Waters L. What does positive psychology tell us about early intervention and prevention with children and adolescents? A review of positive psychological interventions with young people. J Posit Psychol 2020 Jul 07;15(5):588-597. [doi: 10.1080/17439760.2020.1789706]
68. Fava GA, Ruini C, Rafanelli C. Psychometric theory is an obstacle to the progress of clinical research. Psychother Psychosom 2004 Mar 29;73(3):145-148. [doi: 10.1159/000076451] [Medline: 15031586]

Abbreviations

3MDR: multimodal motion-assisted memory desensitization and reconsolidation
CBT: cognitive behavioral therapy
HMD: head-mounted display
ICU: intensive care unit
ICU-VR: intensive care unit–specific virtual reality
iVET: individual in vivo exposure therapy
OCD: obsessive-compulsive disorder
PANAS: Positive and Negative Affect Schedule
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
SAD: social anxiety disorder
SAT: Structured Association Technique
TMTBA: The Machine To Be Another
VR: virtual reality
VR-CBT: virtual reality–based cognitive behavioral therapy
VRET: virtual reality exposure therapy

Edited by J Torous; submitted 15.12.22; peer-reviewed by B Birckhead, T Ong; comments to author 22.01.23; revised version received 09.03.23; accepted 13.03.23; published 06.07.23.

Please cite as:
Li Pira G, Aquilini B, Davoli A, Grandi S, Ruini C
The Use of Virtual Reality Interventions to Promote Positive Mental Health: Systematic Literature Review
JMIR Ment Health 2023;10:e44998
URL: https://mental.jmir.org/2023/1/e44998
doi:10.2196/44998
PMID:37410520

Review

Wearable and Mobile Technologies for the Evaluation and Treatment of Obsessive-Compulsive Disorder: Scoping Review

Adam C Frank[1], MD, PhD; Ruibei Li[1], MD; Bradley S Peterson[1,2], MD; Shrikanth S Narayanan[3], PhD

[1]Department of Psychiatry and Behavioral Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

[2]Division of Child and Adolescent Psychiatry, Children's Hospital Los Angeles, Los Angeles, CA, United States

[3]Ming Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Adam C Frank, MD, PhD
Department of Psychiatry and Behavioral Sciences
Keck School of Medicine
University of Southern California
2250 Alcazar St Suite 2200
Los Angeles, CA, 90033
United States
Phone: 1 323 442 6000
Email: adam.frank@med.usc.edu

Abstract

Background: Smartphones and wearable biosensors can continuously and passively measure aspects of behavior and physiology while also collecting data that require user input. These devices can potentially be used to monitor symptom burden; estimate diagnosis and risk for relapse; predict treatment response; and deliver digital interventions in patients with obsessive-compulsive disorder (OCD), a prevalent and disabling psychiatric condition that often follows a chronic and fluctuating course and may uniquely benefit from these technologies.

Objective: Given the speed at which mobile and wearable technologies are being developed and implemented in clinical settings, a continual reappraisal of this field is needed. In this scoping review, we map the literature on the use of wearable devices and smartphone-based devices or apps in the assessment, monitoring, or treatment of OCD.

Methods: In July 2022 and April 2023, we conducted an initial search and an updated search, respectively, of multiple databases, including PubMed, Embase, APA PsycINFO, and Web of Science, with no restriction on publication period, using the following search strategy: (“OCD” OR “obsessive” OR “obsessive-compulsive”) AND (“smartphone” OR “phone” OR “wearable” OR “sensing” OR “biofeedback” OR “neurofeedback” OR “neuro feedback” OR “digital” OR “phenotyping” OR “mobile” OR “heart rate variability” OR “actigraphy” OR “actimetry” OR “biosignals” OR “biomarker” OR “signals” OR “mobile health”).

Results: We analyzed 2748 articles, reviewed the full text of 77 articles, and extracted data from the 25 articles included in this review. We divided our review into the following three parts: studies without digital or mobile intervention and with passive data collection, studies without digital or mobile intervention and with active or mixed data collection, and studies with a digital or mobile intervention.

Conclusions: Use of mobile and wearable technologies for OCD has developed primarily in the past 15 years, with an increasing pace of related publications. Passive measures from actigraphy generally match subjective reports. Ecological momentary assessment is well tolerated for the naturalistic assessment of symptoms, may capture novel OCD symptoms, and may also document lower symptom burden than retrospective recall. Digital or mobile treatments are diverse; however, they generally provide some improvement in OCD symptom burden. Finally, ongoing work is needed for a safe and trusted uptake of technology by patients and providers.

(JMIR Ment Health 2023;10:e45572) doi:10.2196/45572

KEYWORDS

wearable; smartphone; obsessive-compulsive disorder; OCD; digital; phenotype; biomarker; mobile phone

Introduction

Background

The use of smartphones and wearable devices has increased recently, with an estimated 87% of adults in the United States carrying a smartphone [1] and 1 in 5 Americans using a wearable device [2]. Smartphones provide a near-constant connection to the internet and contain a suite of sensors for estimating parameters such as location, movement, and ambient sound levels [3]. Further, apps are continually being developed for smartphones that range in design from games to music and video streaming services to social media to health monitoring. Data collection from smartphones can be passive, occurring without user awareness or input (eg, accelerometry measurements), or active, where the user is engaged and directly contributes to data collection (eg, answering questionnaires or prompts). Wearable devices are a technology that directly connect to the human body and can sense aspects of physiology (eg, heart rate, oxygen saturation, glucose levels, and lactate levels) or behavior (eg, step count, time of sleep onset, and amount and type of exercise completed) and include items such as wrist-based monitors (eg, fitness trackers and smart watches), smart clothes (eg, shirts and shoes), skin patches, eyeglasses, and contact lenses (see Chan et al [4], Kim et al [5], and Zhang et al [6] for review of this technology). These devices and their associated apps are increasingly finding applications in health and medicine [3-8]. Examples include glucose monitoring in patients with diabetes [5], activity sensing in patients with heart failure [9], and lung function monitoring in patients with chronic obstructive pulmonary disease [10].

In psychiatry, use cases for wearable sensors and smartphone-based apps range in type, design, function, and objective [11,12]. Broadly, these technologies have been used to (1) detect and monitor symptoms [13-16]; (2) estimate diagnostic class, illness severity, and risk for relapse [17-20]; (3) predict response to treatment [21-23]; (4) and deliver digital interventions [24,25]. Populations in which wearable and smartphone-based technologies have been investigated and implemented vary: studies in nonclinical populations have sought to broadly promote and track mental health [26,27]. Within clinical samples, a multitude of psychiatric conditions have been explored, including depression [15,20-22], anxiety [28], schizophrenia [24], bipolar disorder [13,14,17], social anxiety [19], and obsessive-compulsive disorder (OCD) [29]. The extent of the uptake of these technologies in psychiatry varies across conditions, and we are interested in understanding the landscape of the literature covering this within the domain of OCD.

OCD is a chronic and prevalent psychiatric disorder characterized by intrusive and distressing thoughts, images, impulses, and repetitive or ritualistic behaviors [30]. OCD is considered one of the most disabling psychiatric disorders [31] and exacts a significant personal [32] and societal economic toll [33]. The course of OCD is chronic and fluctuating for many individuals [34,35], and treatment response typically hovers near 50% [36,37]. The fluctuating nature and limited treatment responsiveness of OCD present a unique opportunity for wearable and smartphone-based technologies to impact the care for and treatment of individuals with OCD. The use of technology in the treatment of anxiety and obsessive-compulsive spectrum disorders [38], specifically the use of technology in assessing and treating OCD [29], has recently been reviewed. In both reviews, the authors found heterogeneity in the implementation of technology in the care of individuals with OCD as well as an opportunity for advancing research and clinical care.

Objectives of This Review

Given the speed at which new technologies are developed and implemented in clinical settings, a continual reappraisal of this field is needed. In this scoping review, we sought to map the literature on the use of wearable devices and smartphone-based devices or apps in the assessment or monitoring of OCD symptoms and treatment of OCD. Regarding treatment, we focused on novel interventions and excluded studies on the mobile implementation of standard psychotherapy (such as cognitive behavioral therapy [CBT] with exposure-response prevention [ERP]). We aimed to assess domains in which wearable and mobile technologies have had an impact on OCD care while also identifying areas for continued improvement and innovation within this realm.

Methods

Study Design

Given our objectives, the known heterogeneity in implementing technology in individuals with OCD, and the focus on emerging innovative interventions, a scoping review is the most appropriate synthesis approach. The purpose of a scoping review is to identify all available evidence to assess the breadth, depth, and nature of research activity in a topic of interest, and it is particularly useful in rapidly mapping evidence in emerging topics while maintaining rigorous search and study selection processes [39].

The protocol for this review was preregistered at the Open Science Foundation on August 9, 2022 [40]. We consulted a research librarian at the University of Southern California regarding scoping review protocols, topic development, search strategies, and data management. Keywords were initially identified from recent literature reviews relevant to the topic and preliminarily tested using the University of Southern California library database and Google Scholar (Google LLC). We included additional search terms yielded from discussion between the authors. We conducted our initial search in July 2022, with an update in April 2023; we searched multiple databases, including PubMed, Embase, APA PsycInfo, and Web of Science, with no restriction on original study design or publication period, using the following search strategy: (“OCD” OR “obsessive” OR “obsessive-compulsive”) AND (“smartphone” OR “phone” OR “wearable” OR “sensing” OR “biofeedback” OR “neurofeedback” OR “neuro feedback” OR “digital” OR “phenotyping” OR “mobile” OR “heart rate variability” OR “actigraphy” OR “actimetry” OR “biosignals” OR “biomarker” OR “signals” OR “mobile health”). Retrieved records were entered into the Covidence review software

(Veritas Health Innovation), and duplicate records were removed.

Inclusion and Exclusion Criteria

As recommended by the Joanna Briggs Institute Reviewers' Manual for scoping reviews, we used the Population, Concept, and Context framework to inform our inclusion and exclusion criteria [41].

The inclusion criteria for study population are as follows: individuals with OCD as the primary diagnosis and of any age and sex.

Individuals who did not have OCD as a primary diagnosis were excluded from the study to avoid confounding population factors. We applied no sex or age restrictions because OCD can affect any sex and can develop at any age, including in childhood.

The inclusion criteria for study concept are as follows: biobehavioral technology or smartphone-based technology involved in the assessment or monitoring of OCD symptoms or treatment of OCD and does not use CBT and ERP.

We excluded CBT and ERP digital implementation studies because recent reviews have already covered these interventions [29,42,43].

The inclusion criteria for study context are as follows: any care setting, including inpatient, outpatient, or natural environment (eg, at home, work, or school), and English-language studies.

OCD symptoms can be potentially tracked and treated in multiple environments, including inpatient, outpatient, or naturalistic settings, depending on the severity of symptoms and the specific needs of the individual. Therefore, we chose not to restrict our search criteria to a specific setting. We included full-text English-language studies and excluded studies with only an English translation of the abstract to ensure alignment with the inclusion and exclusion criteria and appropriately extract all relevant items. In addition, our preliminary search indicated that most studies used various metrics to monitor OCD symptom burden, so not all conceptual results of interest were included in the abstract. Finally, our preliminary search yielded only a few non–English-language studies.

We decided to exclude review articles, meta-analyses, conference abstracts, and thesis defenses because these either miss elements relevant to our extraction or lack the academic rigor of the peer-review process. Given the long history of biobehavioral technology research and implementation in health care, we considered any publication date.

Together, based on the Population, Concept, and Context framework, the following study inclusion criteria were used: OCD is a primary diagnosis, studies conducted in any care setting, participants of any age and sex, study uses biobehavioral technology or smartphone-based technology, study is peer reviewed, study contains original content, and study is in the English language. Exclusion criteria were as follows: non-OCD primary diagnosis or nonclinical population; non–mobile-based technology; CBT and ERP digital implementation study; and studies that are review articles, meta-analyses, conference abstracts, or thesis defenses. No restrictions were placed on the date of publication of the included studies.

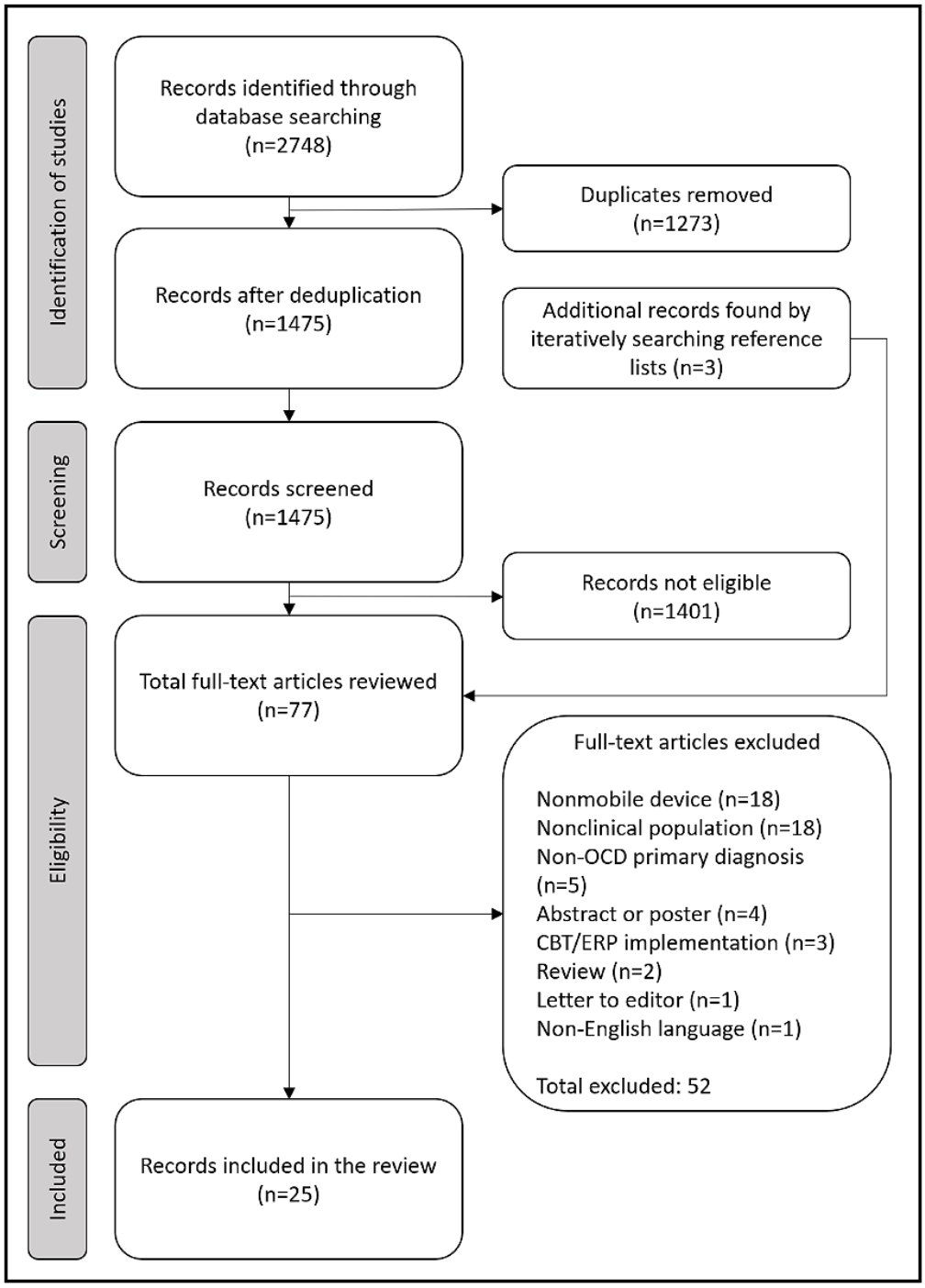
Two authors, AF and RL, screened all studies separately using the blinded screening feature of Covidence; consensus was achieved through discussion between authors for any records with conflicting screening. From this initial screening, full texts from relevant records were obtained. A total of 3 additional studies were identified for full-text review from the reference lists of other reviewed studies. Finally, AF and RL independently extracted the relevant articles and came to consensus on the final extracted items through regular discussion.

Results

Overview

The scoping review was conducted using the Covidence review software, which facilitates the collation of citations with automatic deduplication, allows for blinded screening and review of articles by individual reviewers, tracks articles through the review process, and records reasons for study exclusion. Figure 1 shows the results of the systematic search, study screening, and review process conducted in Covidence. A total of 2748 records were identified across the 4 databases indicated earlier. Following the removal of duplicate records (1273/2748, 46.32%), 1475 (53.68%) of the 2748 studies remained for title and abstract screening. After screening 1475 studies, 1401 (94.98%) studies were found to be ineligible for inclusion, leaving 74 (5.02%) studies for full-text review; an additional 3 studies were identified from the reference list of the reviewed studies, resulting in a total of 77 studies undergoing full-text review. From these 77 studies, 52 (68%) were excluded, mostly studies using nonmobile devices and assessing nonclinical populations or individuals without a diagnosis of OCD. Studies were also excluded if they were conference abstracts or posters, studied the digital implementation of CBT and ERP, were not peer reviewed or were a review, or were not English-language studies.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram. CBT: cognitive behavioral therapy; ERP: exposure-response prevention; OCD: obsessive-compulsive disorder.



Finally, data were extracted from a total of 25 (N=77, 32%) studies. The key findings from these studies, including the type of wearable device or smartphone app, the method of data collection (eg, active vs passive), whether an intervention was implemented, and the overall study results, are discussed in the tables.

Studies Without Digital or Mobile Interventions and With Passive Data Collection

We broadly divided the studies into those that contained digital or mobile interventions and those that did not. We further divided studies that did not include a digital or mobile intervention based on whether they collected mobile or wearable data completely passively or had an active data collection component (ie, requiring participants to enter data or directly engage with an app). Table 1 lists studies lacking a mobile or digital intervention and collecting mobile or wearable data in a passive manner.

Table 1. Studies lacking digital or mobile interventions with passive mobile or wearable data collection, organized by publication year.

Study (author, year, country)	Population (age range [years])	Technology	Mobile or wearable data collected and collection method	Mobile or wearable collection method setting	Nonmobile or wearable data collected and collection method	Study aim	Main results
Millet et al [44], 1998, France	Adult men (26-43)	Actimeter (Gaewilher Electronic)	• Motor activity: passive collection through actigraphy	Inpatient	• Plasma melatonin: active collection • Plasma cortisol: active collection • Axillary temperature: active collection • Symptom burden: active collection through self-report	To assess for differences in circadian variations in plasma melatonin, plasma cortisol, axillary temperature, motor activity, and obsessive-compulsive and depressive symptoms between patients with OCD[a] and controls	• Circadian variation in Hospital Anxiety Depression scale scores • No significant differences in other measures between OCD and control groups
Alfano and Kim [45], 2011, United States	Youth males and females (7-11)	ActiGraphGT1M (ActiGraph)	• Motor activity: passive collection through actigraphy	Natural environment	• Sleep metrics: active collection through parent report • Sleep metrics: active collection through self-report	To assess for differences in sleep patterns in children with OCD compared with controls using objective measures	• Differences in actigraphy measures of sleep in the OCD group, including reduced total sleep time, increased wake after sleep onset, and increased duration of awakenings • Negative correlation between total sleep time and CY-BOCS[b] in the OCD group.
Drummond et al [46], 2012, England	Adult men and women (20-62)	Actiwatch-L (CamNtech Ltd)	• Motor activity: passive collection through actigraphy	Inpatient	• Sleep metrics: active collection through nursing report • Sleep metrics: active collection through self-report	To determine the acceptability, reliability, and validity of using actigraphy to assess sleep patterns in inpatients with severe, refractory OCD	• 59% (36/61) of patients who were admitted agreed to participate in the study. • 81% (29/36) of patients wore the actigraph for 10-20 days, 8% (3/36) wore the actigraph for up to 10 days, and 11% (4/36) were unable to wear the actigraph. • Delayed sleep phase detected by actigraphy showed good agreement with nursing report and self-report.
Pittig et al [47], 2013, United States	Adult men and women (not specified)	LifeShirt system (VivoMetrics)	• Electrocardiogram: passive collection • Respiration: passive collection • Postural data: passive collection	Outpatient	• Symptom burden: active collection through self-report • Task-related subjective units of distress: active collection through self-report	To assess for differences in HR[c] and HRV[d] in patients with anxiety disorders at rest, during stress, and during relaxation and to determine relationships with demographic or clinical variables	• At baseline and during hyperventilation, lower high-frequency HRV in anxiety disorder group • Greater HR during hyperventilation in PD[e] and GAD[f] • Medication status impacted HRV in patients with OCD • Age and sex related to multiple physiological variables

Study (author, year, country)	Population (age range [years])	Technology	Mobile or wearable data collected and collection method	Mobile or wearable collection method setting	Nonmobile or wearable data collected and collection method	Study aim	Main results
Bussing et al [48], 2015, United States	Youth males and females (7-17)	Actical (Mini Mitter)	• Motor activity: passive collection through actigraphy	Natural environment	• Sleep metrics: active collection through self-report • Activation symptoms: active collection through parent report	To determine whether actigraphy can detect SSRI[g] activation syndrome in youths with OCD relative to parent-rated measures	• 35% of daytime and 20% of nighttime actigraphy data were missing. • Female sex associated with lower daytime activity. • Parent report of daytime and nighttime activities was associated with average activity.
Donse et al [49], 2017, Netherlands	Adult men and women (not specified)	Actiwatch Spectrum Plus or Actiwatch 2 (Respironics-Philips)	• Motor activity: passive collection through actigraphy	Natural environment	• Sleep metrics: active collection through self-report	To assess for sleep disturbances in adults with OCD using actigraphy and self-report and to determine whether sleep disturbance can predict responsiveness to rTMS[h] treatment	• Difference in self-report measures of sleep disturbance in patients with OCD • Difference in actigraphic measures of sleep disturbance in patients with OCD • Circadian rhythm sleep disorder model predicted rTMS treatment nonresponse.
Jaspers-Fayer et al [50], 2018, Canada	Youth males and females (8-18)	Fitbit Flex (Fitbit Inc)	• Motor activity: passive collection through actigraphy	Natural environment	• Sleep metrics: active collection through parent report • Sleep metrics: active collection through self-report	To assess for sleep disturbances in children and adolescents with OCD using actigraphy and parent report and self-report	• 72% (18/25) of patients with OCD compared with 15% (4/26) of controls met criteria for sleep disturbance by parent report. • Actigraphy and self-report found longer times between going to bed and falling asleep and longer wake after sleep onset in patients with OCD.
Coles et al [51], 2020, United States	Adult men and women (not specified)	Micro Motionlogger watch (Ambulatory Monitoring Inc)	• Motor activity: passive collection through actigraphy	Natural environment	• Salivary melatonin: active collection • Sleep metrics: active collection through self-report	To assess for differences in sleep parameters between adults with OCD and control adults using self-report, salivary melatonin levels, and wrist actigraphy and to determine whether sleep parameters correlate with symptom burden	• 40% (6/15) of patients with OCD met criteria for DSWPD[i]. • Dim light melatonin onset occurred later in patients with OCD. • Actigraphy data closely mirrored self-report sleep metrics.
Cox and Olatunji [52], 2022, United States	Adult men and women (18-53)	ActiGraph wGT3X-BT (ActiGraph)	• Motor activity: passive collection through actigraphy	Natural environment	• Sleep metrics: active collection through self-report • OCD severity: active collection through self-report		

Study (author, year, country)	Population (age range [years])	Technology	Mobile or wearable data collected and collection method	Mobile or wearable collection method setting	Nonmobile or wearable data collected and collection method	Study aim	Main results
						To examine functional relationship between measures of sleep and delayed circadian rhythms in patients with OCD and its association with OCD symptom severity	• Circadian rhythms are delayed in patients with OCD compared with control patients without psychiatric diagnosis. • Measures of delayed circadian rhythms were associated with OCD symptoms. • Measures of sleep disturbance were not significantly different between individuals with OCD and HC[j], including objective sleep time measured from actigraphy, except for higher insomnia symptoms, which were associated with higher OCD symptoms. • MEQ[k] and DSWPD mediate OCD symptom severity through insomnia in mediation modeling.
Thierfelder et al [53], 2022, Germany	Youth males and females (13-17)		• HR: passive collection through electrocardiogram chest belt • Motor activity: passive collection through wrist-based sensors • Eye-tracking: passive collection through head-mounted device • Data aggregation and processing: passive collection through Surface Pro and custom-built software	Outpatient	• N/A[l]	Pilot study to demonstrate that the collected sensor data capture features of stress reactions, compulsive behavior, and relief from anxiety in an outpatient setting in adolescents with OCD	

Study (author, year, country)	Population (age range [years])	Technology	Mobile or wearable data collected and collection method	Mobile or wearable collection method setting	Nonmobile or wearable data collected and collection method	Study aim	Main results
		Movesense HR2, electrocardiogram chest belt (Suunto); Opal, wrist-based sensor (APDM Inc); *Look!*, custom-built eye tracker with 2 infrared and 1 field camera; Microsoft Surface Pro 7, aggregator device receiving sensor signals and pushing through recording and streaming pipeline (Surface Pro 7 i7/16GB/256GB, Microsoft Corp); and Aggregator Software, custom developed software for data processing and user interface to connect, control, and record the sensors					• $RMSSD^m$ (measure of HRV) decreased for all participants with increasing OCD-related stress. • HR (BPM^n) increased or remained stable with increasing stress. • Relief from stress is generally accompanied by an increase in the RMSSD of HRV and decreased or stable HR. • Increases in physical activity are generally accompanied by a drop in RMSSD and more elevated HR than in a stressful event. • Increase in movement energy can be observed from wrist-based sensors with increasing OCD-related stress. • Repetitive compulsive behavior (checking bag) captured by hand sensors showed a unique frequency distribution compared with other repetitive but noncompulsive behavior. • While refraining from compulsion to wash hands, participants exposed to contamination in public bathroom fixated on the public toilet (59%), floor (27%), and sink (14%).
Gajadien et al [54], 2023, Netherlands	Adult men and women (not specified)	ActTrust (Condor Instruments)	• Motor activity: passive collection through actigraphy	Natural environment	• Sleep metrics: active collection through self-report	To investigate potential differences in sleep parameters between responders and nonresponders to rTMS and to examine the ability of sleep parameters to predict rTMS response	

Study (author, year, country)	Population (age range [years])	Technology	Mobile or wearable data collected and collection method	Mobile or wearable collection method setting	Nonmobile or wearable data collected and collection method	Study aim	Main results
							• Reduction in OCD and depressive symptoms after rTMS treatment • No baseline characteristics significantly different between rTMS responders and nonresponders • Actigraphy parameters did not meet the effect size requirement of Cohen d ≥0.5 for inclusion as a predictor • Discriminant model including subjective sleep quality, sleep latency, daytime dysfunction, and HSDQ[o] insomnia could predict response to rTMS with an AUC[p] of 0.813, a sensitivity of 76%, and a specificity of 50% • Circadian rhythm sleep disorder model from the study by Donse et al [49] was not significant and could not be replicated in this study

[a]OCD: obsessive-compulsive disorder.

[b]CY-BOCS: Children's Yale-Brown Obsessive Compulsive Scale.

[c]HR: heart rate.

[d]HRV: heart rate variability.

[e]PD: panic disorder.

[f]GAD: generalized anxiety disorder.

[g]SSRI: selective serotonin reuptake inhibitor.

[h]rTMS: repetitive transcranial magnetic stimulation.

[i]DSWPD: delayed sleep-wake phase disorder.

[j]HC: healthy control.

[k]MEQ: Morningness-Eveningness Questionnaire.

[l]N/A: not applicable.

[m]RMSSD: root mean square of successive differences.

[n]BPM: beats per minute.

[o]HSDQ: Holland sleep disorder questionnaire.

[p]AUC: area under the curve.

Overall, the studies in this category recruited a mix of youth and adult participants and a mix of male and female participants. A total of 2 (18%) of the 11 studies in this group collected wearable data using advanced body sensors that required monitoring in an outpatient setting. One study used a shirt embedded with sensors to collect physiological data, including heart rate and heart rate variability (HRV; calculated from a continuous electrocardiogram), respiration rate, and postural data in patients with anxiety disorders, including OCD. Briefly, the authors found reduced high-frequency HRV at baseline and during hyperventilation in patients with anxiety compared with control participants. They also found a higher heart rate in patients with panic disorder and generalized anxiety disorder during hyperventilation [47]. The other study piloted the ability

of a variety of sensors to capture OCD symptomology for a larger study. The team used multiple sensors, including an electrocardiogram chest belt, wrist-based sensors, and a custom-built eye-tracking device to measure gaze fixation during OCD-induced stress as well as heart rate, HRV, and hand motor activity during planned OCD-triggering events [53]. HRV decreased with higher stress levels and increased during rest, whereas heart rate either increased or remained stable with higher stress levels and decreased during rest. Additionally, HRV and motor activity data yielded data patterns that distinguished OCD-induced stress from physical activity.

All other studies in this group used actigraphy as an objective measure of activity. Most studies (8/11, 73%) used actigraphy at night as a measure of sleep while also collecting self-report, parent report, or nursing report as additional metrics of sleep quality and quantity. A majority of studies (7/11, 64) collected actigraphy measures in a naturalistic home environment; however, 2 (18%) of the 11 studies were conducted in an inpatient setting, with 1 study focusing on individuals with treatment-refractory OCD [46] and the other study requiring inpatient hospitalization to draw concurrent blood samples [44]. In general, studies found differences in objective and subjective measures of sleep in patients with OCD, such as decreased total sleep time [45,49], increased number of awakenings after sleep onset [45,50], increased duration of awakenings [45], increased time to fall asleep [49,50], later midsleep timing [52], and presence of delayed sleep phase disorder [46,51,52]. However, 1 study provided contrary evidence on sleep disturbance, which was largely nonsignificant between participants with OCD and control participants in subjective and actigraphy measures [52]. Another study analyzed actigraphy and self-reported sleep measures between responders and nonresponders to repetitive transcranial magnetic stimulation (rTMS); the authors found that a circadian rhythm sleep disorder model could discriminate between responders and nonresponders to rTMS treatment with a sensitivity of approximately 84%. An insomnia model could not discriminate between these groups [49]. A more recent study from the same group, however, did not identify the circadian rhythm sleep disorder model as a potential predictor of rTMS response but did find measures of sleep disturbance, as measured by self-report but not actigraphy measures, to be predictive of rTMS response, with an area under the curve of 0.813, sensitivity of 76%, and specificity of 50% [54]. Other investigators have used actigraphy to measure activity in patients with OCD during daytime hours. One study reported no abnormalities in circadian variability compared with controls [44]. Another study that assessed behavioral activation from selective serotonin reuptake inhibitor treatment reported lower daytime activity in girls and an association of parent reports with actigraphy measures of activity [48].

Studies Without Digital or Mobile Interventions and With Active or Mixed Data Collection

All the study participants in this category were adults. Ecological momentary assessments (EMAs) occur in a naturalistic setting and were used in all studies that collected mobile or wearable data in an active or mixed fashion (active and passive concurrently, albeit only Brown et al [55] used a mixed collection method) (Table 2).

Table 2. Studies without digital or mobile interventions and with active or mixed mobile or wearable data collection, organized by publication year.

Study (author, year, country)	Population (age range [years])	Technology	Mobile or wearable data collected and collection method	Mobile or wearable collection method setting	Nonmobile or wearable data collected and collection method	Study aim	Main results
Gloster et al [56], 2008, United States	Adult men and women (20-62)	Palm Zire 21 personal data assistant (Palm Inc)	• Symptom burden: active collection through self-report • Context and social interaction: active collection through self-report	Natural environment	N/A[a]	To determine the accuracy of retrospective estimates of daily OCD[b] symptom burden and symptom covariation relative to prospectively collected EMAs[c] in patients with OCD	• Retrospective recall of OCD symptoms was generally consistent with EMA data, although there was occasional underestimation of the frequency of OCD behaviors • Consistent overestimation of the covariation of symptoms with nonsymptomatic variables
Tilley and Rees [57], 2014, Australia	Adult men and women (28-54)	SMS text message–based prompts; Olympus WS-110 digital voice recorder (Olympus, Tokyo, Japan)	• Symptom burden: active collection through self-report	Natural environment	N/A	To determine whether the use of EMA can provide additional diagnostic information in those with OCD	• Fewer symptoms were endorsed by EMA, although new types of symptoms were reported.
Rupp et al [58], 2019, Germany	Adult men and women (not specified)	movisenseXS (movisens GmbH) implemented on Motorola Moto G2 (Lenovo)	• Symptom burden: active collection through self-report • Emotions related to OCD: active collection through self-report • Behaviors related to OCD: active collection through self-report	Natural environment	Feasibility of EMA: active collection through self-report	To determine the feasibility and effectiveness of using EMA to assess OCD symptoms before and after psychotherapy treatment	• 28.11% (851/3027) of EMA responses removed during data cleansing • Questions regarding acceptability, practicability, representativeness, and reactivity were rated fairly, and responses did not change before and after treatment • Reductions in avoidance and obsessions following treatment

Study (author, year, country)	Population (age range [years])	Technology	Mobile or wearable data collected and collection method	Mobile or wearable collection method setting	Nonmobile or wearable data collected and collection method	Study aim	Main results
Brown et al [55], 2020, United States	Adult men and women (not specified)	Fitbit Alta (Fitbit Inc), Twilio technology (Twilio Inc), and Way to Health Platform [59]	• Behavior and physiology: passive collection through Fitbit • Symptom burden: active collection through self-report • Social interaction and context: active collection through self-report	Natural environment	Acceptability: active collection through qualitative interview	To assess patients' and clinicians' perspectives on the use of a wearable biosensor and EMAs in measurement of OCD symptoms	• High EMA response rate (90.2%) with moderate adherence to physical activity (57.7%) and sleep (52.2%) data collection • Multiple patient qualitative themes from Fitbit and EMA use were generally positive, although there were some concerns about technology use and data accuracy • Clinician themes included concerns about amount of data and integration into clinical care
Rupp et al [60], 2020, Germany	Adult men and women (not specified)	movisenseXS (movisens GmbH) implemented on Motorola Moto G2 (Lenovo)	• Emotions related to OCD: active collection through self-report • Behaviors related to OCD: active collection through self-report	Natural environment	N/A	To use pretreatment and posttreatment EMAs to compare the effects of 2 weeks of CR[d] treatment on OCD with those of DM[e] treatment on OCD	• Some baseline use of therapy techniques by participants before treatment • Increase in the use of psychotherapy strategies and behaviors after treatment • No difference between different therapy modalities in the frequency of use, perceived difficulty, and the experience of relief after treatment

[a]N/A: not applicable.

[b]OCD: obsessive-compulsive disorder.

[c]EMA: ecological momentary assessment.

[d]CR: cognitive restructuring.

[e]DM: detached mindfulness.

A series of studies assessed whether the use of EMA could outperform retrospective symptom recall [56], uncover new OCD symptoms [57], and feasibly monitor symptoms [55]. EMA reported a slightly lower frequency [56] and burden [57] of OCD symptoms than clinician-administered Yale-Brown Obsessive Compulsive Scale (Y-BOCS) or Obsessive Compulsive Inventory-Revised, although EMA captured novel, previously unreported OCD symptoms [57]. Although patients' perspectives on the use of EMA were generally positive, clinicians expressed some concern about the amount of data collected and the integration of EMA into clinical care [55]. Rupp et al [58] assessed the feasibility of using EMA to assess OCD symptom burden before and after participants completed detached mindfulness or cognitive restructuring psychotherapy. Participants generally rated EMA highly in terms of acceptability, practicability, and representativeness. The data were quite noisy; however, approximately 28.11% (851/3027) of the noise was removed during data cleansing [58]. A separate study assessed the results of these psychotherapy interventions and found no significant differences between the detached mindfulness and cognitive restructuring therapies in the frequency of their use, perceived difficulty, or the experience of relief after treatment [60].

Studies With Digital or Mobile Interventions

All studies in this section include some form of mobile or digital intervention; however, we excluded studies that focused on CBT and ERP implementation. All but 1 study (8/9, 89%) also involved the collection of active or passive digital or mobile data (Table 3).

Table 3. Studies with digital or mobile interventions, organized by publication year.

Study (author, year, country)	Population (age [years])	Technology	Mobile or wearable data collected and collection method	Nonmobile or wearable data collected and collection method	Digital or mobile intervention	Digital or mobile intervention setting	Study aim	Main results
Le Boeuf [61], 1974, England	Adult man (49)	Portable shock box carried in jacket pocket and connected via electrodes to the forearm and base of the index finger; shock delivered if circuit completed through the immersion of hands in water	• Presence of water: passive collection through shock device	• Daily handwashing frequency: active collection through self-report	Shock device was turned on for specified periods to provide positive punishment for handwashing	Natural environment	To determine the efficacy of an automated shocking device in the treatment of compulsive handwashing	• Decrease in daily handwashing following 2 weeks of shock box use
Olbrich et al [62], 2016, Germany	Adult man (31)	Geo-Feedback App (developed by S Olbrich)	• Position: passive collection through GPS	• Time to reach treatment clinic: active collection through self-report	Mobile app provides the user a notification if they have not moved a predefined distance in a given length of time	Natural environment	To determine whether smartphone-based feedback can be used to treat OCD[a]	• Use of mobile app decreased the time needed to reach treatment clinic (1 mile distance) from 2 hours to 1 hour. • With the addition of consistent ERP[b] and app use, time to reach the clinic decreased to 20 minutes. • Patient reported the fear of attracting attention from app notifications as a negative reinforcer.
Kashyap et al [63], 2019, India	Adult man (29)	CogTrain App (developed by P Reddy and S Mandadi)	• N/A[c]	• Cognitive and symptom assessments: active collection through clinician-administered	Mobile app for cognitive training (coupled with in-person therapist-guided sessions)	Natural environment	To report on the use of cognitive training as an intervention for OCD using a custom smartphone app, therapist training, and various freely available smartphone apps	• Patient completed therapist-guided cognitive training, mindfulness practices, and ADL[d] training. • Patient completed cognitive training tasks and used the CogTrain App. • Over 12 weeks, patient had improvement in symptom burden and improvement in some cognitive measures.

Study (author, year, country)	Population (age [years])	Technology	Mobile or wearable data collected and collection method	Nonmobile or wearable data collected and collection method	Digital or mobile intervention	Digital or mobile intervention setting	Study aim	Main results
Arevian et al [64], 2020, United States	Adult men and women (18-69)	Chorus platform (Chorus Innovations Inc) and SMS text messages	• Symptom burden and response to treatment: active collection through self-report	• Feasibility and acceptability: active collection through clinician-administered survey	SMS text messages sent to participants to encourage adherence to treatment, remind them to take medication and engage in exposures, and provide information in addition to IOP[e] treatment	Natural environment	To evaluate the usability of a mobile texting app, to evaluate the feasibility of app development with patients and providers, and to describe the types of texting apps developed	• 1787 messages sent and 80 responses received • Various types of messages were created, and overall themes for messages were personalization to individuals and use of humor • Most patients expressed positive feedback about the development and use of messages • Themes from workgroups included treatment engagement, personalization of treatment, motivation, and after-hours care
Olsen et al [65], 2020, United States	Adult man (20s)	Activa PC+S (Medtronic) and smartphone-based EMA	• Intracranial LFP[f]: passive collection through DBS[g] system • Motivation and functionality: active collection through EMA	• MSIT[h]: active collection • Symptom burden: active collection through clinician-administered measures	Open-loop, dual-site DBS to the bilateral VC/VS[i] and SMA[j]	Natural environment and outpatient	To test the feasibility of combining VC/VS DBS with frequency-mismatched stimulation of the SMA in treating refractory OCD	• Small decrease in Y-BOCS[k] with cortical stimulation and small increase in MADRS[l] • PGI-I[m] improved with the addition of cortical stimulation • MSIT reaction time improved with dual-site stimulation • Cortical-striatal synchrony increased with dual-site stimulation • Various changes in power spectra through study • Random forest model predicting PGI-I performed with 92% accuracy, and cortical-striatal gamma and theta synchrony were important features

Study (author, year, country)	Population (age [years])	Technology	Mobile or wearable data collected and collection method	Nonmobile or wearable data collected and collection method	Digital or mobile intervention	Digital or mobile intervention setting	Study aim	Main results
Provenza et al [66], 2021, United States	Adult men and women (31-40)	Summit RC+S (Medtronic); Apple Watch (Apple Inc); StriveStudy mobile app (Rune Labs); Honeycomb task app; actiCAP electroencephalogram cap (Brain Products GmbH); GoPro Hero 6 (GoPro Inc); H4n Pro 4-track Portable Recorder (Zoom Corp); and AFAR[n] computer-vision (Carnegie Mellon University)	• Intracranial electrophysiology: passive collection through DBS device • Heart rate, blood volume pulse, and acceleration: passive collection through Apple Watch • OCD symptom severity: active collection through self-report via StriveStudy app • Performance on cognitive and behavioral tasks: active collection through Honeycomb app • Extracranial electrophysiology: passive collection through actiCAP • Facial movements: passive collection through GoPro and AFAR • Speech: passive collection through H4n recorder	• Symptom burden: active collection through clinician-administered measures	Open-loop DBS to bilateral VC/VS or BNST[o]	Natural environment and outpatient	To identify the neural biomarkers of OCD through (1) the measurement of intracranial and extracranial electrophysiology, (2) self-reported OCD symptom burden, (3) objectively measured affective state, and (4) the evaluation of physiology for the purpose of developing an adaptive DBS for OCD	

Study (author, year, country)	Population (age [years])	Technology	Mobile or wearable data collected and collection method	Nonmobile or wearable data collected and collection method	Digital or mobile intervention	Digital or mobile intervention setting	Study aim	Main results
								• AFAR software estimated positive affect and head velocity, and these data were synchronized with blood volume pulse, electrocardiogram, electroencephalogram, LFP from DBS, and the acceleration of INS^p • Increase in positive affect and subjective positive feelings during DBS programming session in participant 5 • Self-reports were synchronized to at-home, wirelessly streamed DBS LFP, routine at-home tasks, and psychophysiological tasks • In total, across the 3 participants, over 1000 hours of at-home intracranial physiology was recorded • 1 participant completed an at-home LFP recording for 3 continuous days; 41 OCD symptom intensity ratings were collected during this period (range 0-8); LFP frequency band power was examined in the minute before and after self-report; there was a strong negative correlation between power in the delta band and OCD symptom severity in both the left (R=−0.593) and

Study (author, year, country)	Population (age [years])	Technology	Mobile or wearable data collected and collection method	Nonmobile or wearable data collected and collection method	Digital or mobile intervention	Digital or mobile intervention setting	Study aim	Main results
								right (R=−0.557) VC/VS; correlations to planned ERP exposures in this participant were also seen in the delta band
Hawley et al [67], 2021, Canada	Adults of unspecified sex (not specified)	Muse, a consumer-grade electroencephalogram headset device with a mobile app (InteraXon Inc)	• Electroencephalogram: passive collection through Muse device	• Symptom burden: active collection through self-report	Muse technology–guided neurofeedback with daily guided mindfulness	Natural environment and outpatient	To determine whether technology-supported mindfulness can improve OCD symptom burden, increase self-reported mindfulness, and increase electroencephalogram-derived indicators of mind wandering	• Decrease in Y-BOCS-SR[q] in the active treatment group • Increased alpha and beta electroencephalogram power in the treatment group • Alpha and beta power predicted Y-BOCS-SR decrease • Measures of mind wandering predicted Y-BOCS-SR
Hawley et al [68], 2021, Canada	Adults of unspecified sex (not specified)	Muse, a consumer-grade electroencephalogram headset device with a mobile app (InteraXon Inc)	• Electroencephalogram: passive collection through Muse device	• Symptom burden: active collection through self-report	Muse technology–guided neurofeedback with daily guided mindfulness	Natural environment and outpatient	To determine whether technology-supported mindfulness training is associated with decreased cognitive vulnerability, improved attention, reduced OCD symptom burden, and the existence of a relationship between electroencephalogram-derived markers of attention and clinical variables	• Decrease in Y-BOCS-SR in the active treatment group • Increased alpha and beta electroencephalogram power in the treatment group • OBQ[r] perfectionism or certainty and importance or control and Y-BOCS-SR bidirectionally predicted changes in each value • Alpha power and OBQ perfectionism or certainty bidirectionally predictive of each other
Fridgeirsson et al [69], 2023, Netherlands	Adult men and women (30-69)	Medtronic Percept or Activa PC+S with 3389 DBS leads (Medtronic)	• Intracranial LFP: passive collection through DBS system	• Symptom burden: active collection through self-report		Outpatient		

Study (author, year, country)	Population (age [years])	Technology	Mobile or wearable data collected and collection method	Nonmobile or wearable data collected and collection method	Digital or mobile intervention	Digital or mobile intervention setting	Study aim	Main results
					Open-loop DBS to bilateral vALIC[s], although DBS was not active during the study period		To identify an electrophysiologic biomarker of OCD symptoms in adults implanted with DBS in the vALIC through machine learning approaches	• Obsession induction increased VAS[t] scores for anxiety, agitation, obsession, and compulsions • Power in all examined frequency bands increased during compulsions and relief state relative to baseline • Total balanced accuracy of predicting individuals from baseline LFP data was 18.9% for a boosted trees model and 32.6% for a deep learning model compared with a chance level of 9% ($P<.05$) • Patient-specific models showed an average accuracy of 32.5% in predicting the symptom state of individual patients using boosted trees and 38.8% accuracy using deep learning • Deep learning reached an average AUC[u] of 78.2% for compulsions, 62.1% for obsessions, 58.7% for baseline, and 59.7% for relief

[a]OCD: obsessive-compulsive disorder.

[b]ERP: exposure-response prevention therapy.

[c]N/A: not applicable.

[d]ADL: activities of daily living.

[e]IOP: intensive outpatient program.

[f]LFP: local field potential.

[g]DBS: deep brain stimulation.

[h]MSIT: multisource interference task.

[i]VC/VS: ventral capsule/ventral striatum.

[j]SMA: supplementary motor area.

[k]Y-BOCS: Yale-Brown Obsessive Compulsive Scale.

[l]MADRS: Montgomery-Asberg Depression Rating Scale.
[m]PGI-I: patient global impression of improvement.
[n]AFAR: automatic facial affect recognition.
[o]BNST: bed nucleus of the stria terminalis.
[p]INS: implanted neural stimulator.
[q]Y-BOCS-SR: Yale-Brown Obsessive Compulsive Scaleself-report.
[r]OBQ: Obsessive Beliefs Questionnaire.
[s]vALIC: ventral anterior limb of internal capsules.
[t]VAS: visual analog scale.
[u]AUC: area under the curve.

All studies applied interventions in a natural environment, with some providing constant treatment via deep brain stimulation (DBS) [65,66] and others using a 2-pronged at-home and in-clinic interventional approach [67,68]. A total of 2 (22%) of the 9 studies developed novel interventions aimed specifically at OCD symptoms. In a single-participant case report, Le Boeuf [61] created a wearable device that provided a mild electric shock to the user if an electrical circuit was completed when the user's hands were in contact with water, presumably during a washing compulsion. The participant had severely impairing, compulsive handwashing before treatment and had a marked and durable improvement in symptom burden soon after beginning the use of the wearable device [61]. Even more notable is that this study was completed in 1974—well before the advent of smartphones or modern wearable biosensors. In another single-participant case report, Olbrich et al [62] developed a smartphone app to address severe harm-based obsessions and checking compulsions that prevented the participant from attending psychotherapy appointments. The smartphone app tracked the user's location and sent a reminder signal if the participant had not moved a predefined distance. Use of the app reduced the time required for the participant to reach the clinic by 50% (2 hours to 1 hour for a travel distance of 1 mile); once the patient was able to re-engage in ERP (and continue to use the app), he reached the clinic in 20 minutes. He endorsed that the app served as a negative reinforcer in that he feared drawing attention to himself if the app made a signal noise [62].

A total of 2 (22%) of the 9 studies used DBS as a treatment modality, collected longitudinal intracranial physiological measures from the DBS electrodes, and collected passive and active digital and wearable metrics. In a case report of a patient receiving dual-site stimulation in the ventral capsule/ventral striatum and supplementary motor area, Olsen et al [65] found a small improvement in Y-BOCS score and more robust improvement in patient global impression of improvement following dual-site stimulation compared with single ventral capsule/ventral striatum stimulation. Unexpectedly, cortical-striatal synchrony increased with dual-site stimulation, and random forest modeling showed that cortical-striatal gamma and theta synchrony predicted patient global impression of improvement with 92% accuracy [65]. Provenza et al [66] took a multimodal approach to studying OCD by chronically recording intracranial local field potentials (LFPs) while also densely collecting other measures, including heart rate, self-report symptom burden, facial features, and speech samples. The authors presented data on the first few participants in a planned, larger study. They found significant correlations of facial affect with subjective improvement during DBS programming, negative correlations of delta band power with OCD symptom severity, and correlations of delta band power with ERP [66]. Fridgeirsson et al [69] used machine learning models to analyze LFP data collected during rest and symptom provocation in an outpatient setting while DBS was turned off. Machine learning modeling of resting LFP identified individual patients significantly above the chance level. Modeling of LFP collected during symptom provocation and relief of symptoms predicted symptom state with an average accuracy of 32.5% and 38.8% for the boosted trees and deep learning model, respectively, with the latter reaching an average area under the curve of 78.2% for compulsions, 62.1% for obsessions, 58.7% for baseline, and 59.7% for relief [69].

A total of 2 (22%) of the 9 studies from 1 research group used a mobile electroencephalogram device for monitoring during biofeedback treatment. They found that active treatment reduced Y-BOCS self-report scores and increased electroencephalogram alpha and beta power in proportion to the improvement in OCD symptoms [67]. They also found that alpha power and the perfectionism or certainty subscores from the Obsessive Beliefs Questionnaire were reciprocally associated with one another across time [68].

Arevian et al [64] developed and tested the usability and feasibility of a mobile texting app cocreated by therapists and patients for use in an OCD treatment clinic. The app prompted patients with SMS text messages, and some SMS text messages also requested patient engagement and response. The authors found that the types of SMS text messages created generally focused on personalizing treatment for the individual and using humor to aid in treatment. Approximately 80 to 90% of the patients expressed positive sentiments about the development and use of the app. Themes from working groups of therapists regarding app development and use included treatment engagement, personalization of treatment, motivation, and provision of after-hours care [64].

Finally, 1 (11%) of the 9 studies used a custom mobile phone app aimed at improving cognition in individuals with OCD; this was coupled with therapist-led treatment in a single individual with subjective cognitive complaints [63]. The authors reported that this multimodal approach was associated with improvements in subjective and objective measures of cognition.

Discussion

Overview

Wearable sensors and smartphone-based apps are increasingly being used in medicine and health broadly and in psychiatry specifically to monitor symptoms, diagnose diseases, and predict responses to treatment [3-5,7,8,11,12]. Given that OCD is a chronic, fluctuating condition with significant personal and economic costs, there is a unique opportunity to implement this evolving digital data collection framework to improve our understanding of disease mechanism and improve treatment and clinical outcomes [32-35]. This scoping review maps the extant literature on the use of wearable and smartphone-based technologies in tracking, diagnosing, and predicting clinical outcomes in individuals with OCD. The included studies were broadly divided into studies with digital or mobile interventions and those without. Studies without such interventions were further categorized based on whether they solely collected mobile or wearable data passively or involved an active component in data collection. The results of recent reviews of technology use in OCD do not meaningfully overlap with our results: Cooper et al [38] explored the use of technology in facilitating therapist-delivered psychotherapy in person or by webcam, assessment and prediction of OCD symptoms, and interventions in treating OCD, with the results of none of the reviewed studies overlapping ours; Ferreri et al [29] focused broadly on the use of technology in the assessment and prediction of and interventions for OCD, with the results of only 2 reviewed studies overlapping with ours [57,62].

Principal Findings

We found several broad themes through this study. First, except for 2 (8%) of the 25 studies [44,61], the reviewed studies indicate that the use of wearable sensors or mobile apps in evaluating and treating OCD has developed within the past 15 years, with over half (15/25, 60%) of the studies having been conducted in the last 5 years. This speaks not only to the novelty of these methods in psychiatry but also to the increasing pace of adoption of mobile and wearable technologies in health and medicine. Second, regarding the types of technology, most studies using fully passive mobile or wearable data collection used actigraphy to assess sleep or, less frequently, daytime movement patterns. These studies generally reported good agreement between objective actigraphy data and patient-, parent-, or nurse-reported subjective metrics. However, no study used actigraphy as the sole measure of sleep, which may be indicative of the current limitations of actigraphy. Several studies used extensive, nonconsumer sensors and modalities, including custom-built hardware and software, to passively track OCD symptoms. These efforts highlight the potential therapeutic benefit of tracking OCD symptoms passively and the desire for higher-performance sensors that are not available in off-the-shelf solutions. Many studies that actively collect mobile or wearable data use EMA to assess OCD symptom severity and burden in a naturalistic manner. In general, EMA is well tolerated by participants and appears to uncover new OCD symptoms not reported on retrospective questionnaires, although it may underestimate the overall OCD symptom burden.

We found that mobile or digital interventions are varied and diverse. They include apps and devices that provide negative reinforcement, apps that provide cognitive training, apps facilitating bidirectional texting and SMS text messaging between providers and patients, electroencephalogram-based biofeedback devices, and open-loop DBS with concurrently recorded intracranial LFP. Studies that leveraged mobile or digital interventions were often case reports with a single male participant. The dearth of studies involving a larger and more diverse participant pool highlights the novelty of such interventions. Nevertheless, given the rapidity of technology development and adoption, we anticipate the depth and breadth of mobile and digital interventions to continue to expand with increasing speed. Finally, approaches to data privacy and security are often underreported. This is a critical issue to address given the user concerns about these technologies [70-72], the ongoing integration of technology into health care, and the potential for malicious use of data [73].

The findings of this review highlight several important considerations for future studies and the implementation of digital health technologies in clinical practice. First, the consistency and standardization of data collection and analysis are important and likely to improve both study quality and public perception of digital or wearable technology research. To facilitate this effort, future studies will benefit from the use of a conceptual framework that allows one to identify important metrics to assess, determine on what timescale to collect these measures, and decide how to implement appropriate and statistically sound analytic methods. Two commonly used and conceptually overlapping frameworks—behavioral signal processing [74] and digital phenotyping [75,76]—share important features and aims: acquisition of multimodal and ecologically valid data, selection of analytic methods suited to the acquired data, and development of models to predict clinical course and treatment response. Both approaches have been usefully implemented in psychiatric conditions as diverse as schizophrenia [77], depression [78,79], anxiety [80,81], and autism spectrum disorder [82,83].

Second, wearable- and smartphone-based studies have the potential to improve treatment outcomes through the development of intervention decision models, which are collections of strategies and policies for the evaluation and treatment of patients and are commonly used in diverse fields of medicine [84]. Decision models operate most effectively when the illness phenomenology (ie, signs and symptoms) maps onto an understanding of the pathophysiology. Wearable devices and smartphones will allow for the ongoing collection of diverse, dense, longitudinal data sets that can improve our understanding of the signs and symptoms of psychiatric disorders; in conjunction with research on the pathophysiology of mental illness, these complementary approaches will lead to the development of much-needed decision models in psychiatry [84].

Third, the included studies were conducted across the globe in countries, including India, Australia, Canada, the United States, and several European nations. Globally, smartphone use ranges from 70% to 85% of the population and is steadily increasing [85,86]. Furthermore, mental illness is prevalent throughout the

world [87,88], and even within the United States, there are disparities in access to care based on race and ethnicity [89,90]. We also found that studies were conducted across age groups, from children and adolescents to those in their 60s, and in both males and females (although not all studies reported age). Taking these themes together, wearable- and smartphone-based studies can, and should, be conducted in diverse settings and populations around the world. This naturally lends itself to large, concurrent studies that are scaled up to include many more participants so that variability in measures can effectively be captured and analyzed.

Fourth, the declining costs of technology, ubiquitous use of smartphones and their associated functionalities (eg, user interface, cloud connection, and data storage and sharing), and integration of artificial intelligence for high-dimensional data processing have enabled real-time monitoring of various health-related biomarkers via wearable biosensors [6]. For instance, a recently proposed study uses an armband biosensor to passively monitor diverse physiological parameters in patients with COVID-19. An associated smartphone app receives and stores real-time data from the sensor and subsequently uploads them to a cloud-based server, where further processing occurs via machine learning. The results can then be displayed to a clinician via a web-based dashboard with an overall goal of early detection of disease progression [91]. From our review, Provenza et al [66] took a similar approach in capturing diverse streams of wearable and mobile data concurrently. They demonstrated an approach to combining these data into a broader scientific and clinical picture [66]. We anticipate that future studies in psychiatry will further integrate actively and passively collected wearable and mobile data, on-device and cloud-based storage, and real-time data extraction and analysis to produce actionable information that patients and clinicians can use to guide care.

Finally, the ethical and legal frameworks surrounding mobile and wearable data collection and use continue to evolve, particularly as the definitions of devices and apps change [92]. Currently, most devices are not regulated by the Food and Drug Administration, although their features or marketing suggest medical diagnostic capabilities; this leaves manufacturers and, potentially, physicians open to state and federal liabilities should these devices malfunction or fail to perform as advertised [93]. Simon et al [93] suggested changes to state and federal regulations to mitigate this liability, although they also note that best practices developed by physician organizations that specifically address mobile or wearable devices may reduce some legal risk.

Strengths and Limitations

We present an overview of the use of mobile and wearable technologies in the monitoring and treatment of OCD. Our systematic approach to the literature ensured that all indexed studies were included, supported by our identification of an older study not previously captured in reviews [61]. It is possible that relevant non–English-language studies were overlooked, as we focused our review on manuscripts published in English. We divided the included studies into studies passively collecting data, studies actively collecting data, and studies implementing treatment; this decision was based on the structure and findings of these studies and was intended to highlight the current landscape of the field, although other organizational approaches could be validly implemented. We chose not to focus on digital or mobile implementation of CBT and ERP, given the existing recent reviews highlighting the literature covering CBT and ERP [29,42,43], and to identify studies reporting novel digital or mobile treatment approaches. Finally, we briefly suggest areas for ongoing consideration when designing studies and considering the clinical implementation of mobile and wearable technologies in OCD. This field is evolving rapidly, and continued publication of high-quality research is paramount for a safe and trusted uptake of technology by patients and providers.

Acknowledgments

The authors thank Lynn Kysh, Master of Library and Information Science, Master of Public Policy, for her expertise as a clinical and research librarian in discussions on scoping review protocols, topic development, search strategies, and data management. This work was supported by the National Center for Advancing Translational Sciences award UL1TR001855-06A1 to ACF.

Authors' Contributions

ACF, RL, BSP, and SSN conceived and designed the study. ACF completed the database searches and entered the studies into Covidence. ACF and RL screened the titles and abstracts, reviewed the full texts, and extracted the data. ACF created the figure and tables and wrote the manuscript. ACF, RL, BSP, and SSN edited and revised the manuscript. BSP and SSN provided guidance and supervision throughout the review process. All authors discussed the results, contributed to the final manuscript, and read and approved the published version of the manuscript.

Conflicts of Interest

SSN is the chief scientist and a cofounder with equity stake of Behavioral Signals, a technology company focused on creating technologies for emotional and behavioral machine intelligence in consumer services. He is also the chief engineering science officer and a cofounder with equity stake of Lyssn, a technology company focused on tools for supporting training on and the supervision and quality assurance of psychotherapy and counseling.

References

1. Faverio M. Share of those 65 and older who are tech users has grown in the past decade. Pew Research Center. 2022 Jan 13. URL: https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ [accessed 2023-01-07]
2. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. 2020 Jan 9. URL: https://www.pewresearch.org/short-reads/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2023-01-07]
3. Majumder S, Deen MJ. Smartphone sensors for health monitoring and diagnosis. Sensors (Basel) 2019 May 09;19(9):2164 [FREE Full text] [doi: 10.3390/s19092164] [Medline: 31075985]
4. Chan M, Estève D, Fourniols JY, Escriba C, Campo E. Smart wearable systems: current status and future challenges. Artif Intell Med 2012 Nov;56(3):137-156. [doi: 10.1016/j.artmed.2012.09.003] [Medline: 23122689]
5. Kim J, Campbell AS, de Ávila BE, Wang J. Wearable biosensors for healthcare monitoring. Nat Biotechnol 2019 Apr;37(4):389-406 [FREE Full text] [doi: 10.1038/s41587-019-0045-y] [Medline: 30804534]
6. Zhang Y, Hu Y, Jiang N, Yetisen AK. Wearable artificial intelligence biosensor networks. Biosens Bioelectron 2022 Oct 21;219:114825 [FREE Full text] [doi: 10.1016/j.bios.2022.114825] [Medline: 36306563]
7. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
8. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132 [FREE Full text] [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
9. Aranki D, Kurillo G, Yan P, Liebovitz DM, Bajcsy R. Real-time tele-monitoring of patients with chronic heart-failure using a smartphone: lessons learned. IEEE Trans Affect Comput 2016 Jul 1;7(3):206-219. [doi: 10.1109/TAFFC.2016.2554118]
10. Thap T, Chung H, Jeong C, Hwang KE, Kim HR, Yoon KH, et al. High-resolution time-frequency spectrum-based lung function test from a smartphone microphone. Sensors (Basel) 2016 Aug 17;16(8):1305 [FREE Full text] [doi: 10.3390/s16081305] [Medline: 27548164]
11. Hickey BA, Chalmers T, Newton P, Lin CT, Sibbritt D, McLachlan CS, et al. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: a systematic review. Sensors (Basel) 2021 May 16;21(10):3461 [FREE Full text] [doi: 10.3390/s21103461] [Medline: 34065620]
12. Goldberg SB, Lam SU, Simonsson O, Torous J, Sun S. Mobile phone-based interventions for mental health: a systematic meta-review of 14 meta-analyses of randomized controlled trials. PLOS Digit Health 2022;1(1):e0000002 [FREE Full text] [doi: 10.1371/journal.pdig.0000002] [Medline: 35224559]
13. Palmius N, Tsanas A, Saunders KE, Bilderbeck AC, Geddes JR, Goodwin GM, et al. Detecting bipolar depression from geographic location data. IEEE Trans Biomed Eng 2017 Aug;64(8):1761-1771 [FREE Full text] [doi: 10.1109/TBME.2016.2611862] [Medline: 28113247]
14. Faurholt-Jepsen M, Busk J, Þórarinsdóttir H, Frost M, Bardram JE, Vinberg M, et al. Objective smartphone data as a potential diagnostic marker of bipolar disorder. Aust N Z J Psychiatry 2019 Feb;53(2):119-128. [doi: 10.1177/0004867418808900] [Medline: 30387368]
15. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
16. Arevian AC, Bone D, Malandrakis N, Martinez VR, Wells KB, Miklowitz DJ, et al. Clinical state tracking in serious mental illness through computational analysis of speech. PLoS One 2020 Jan 15;15(1):e0225695 [FREE Full text] [doi: 10.1371/journal.pone.0225695] [Medline: 31940347]
17. Karam ZN, Provost EM, Singh S, Montgomery J, Archer C, Harrington G, et al. Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. Proc IEEE Int Conf Acoust Speech Signal Process 2014 May;2014:4858-4862 [FREE Full text] [doi: 10.1109/ICASSP.2014.6854525] [Medline: 27630535]
18. Zhu Y, Jayagopal JK, Mehta RK, Erraguntla M, Nuamah J, McDonald AD, et al. Classifying major depressive disorder using fNIRS during motor rehabilitation. IEEE Trans Neural Syst Rehabil Eng 2020 Apr;28(4):961-969. [doi: 10.1109/TNSRE.2020.2972270] [Medline: 32054581]
19. Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. J Med Internet Res 2020 May 29;22(5):e16875 [FREE Full text] [doi: 10.2196/16875] [Medline: 32348284]
20. Narziev N, Goh H, Toshnazarov K, Lee SA, Chung KM, Noh Y. STDD: short-term depression detection with passive sensing. Sensors (Basel) 2020 Mar 04;20(5):1396 [FREE Full text] [doi: 10.3390/s20051396] [Medline: 32143358]
21. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. Biol Psychiatry 2012 Oct 01;72(7):580-587 [FREE Full text] [doi: 10.1016/j.biopsych.2012.03.015] [Medline: 22541039]
22. Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J Neurolinguistics 2007 Jan;20(1):50-64 [FREE Full text] [doi: 10.1016/j.jneuroling.2006.04.001] [Medline: 21253440]

23. Scott J, Hidalgo-Mazzei D, Strawbridge R, Young A, Resche-Rigon M, Etain B, et al. Prospective cohort study of early biosignatures of response to lithium in bipolar-I-disorders: overview of the H2020-funded R-LiNK initiative. Int J Bipolar Disord 2019 Sep 25;7(1):20 [FREE Full text] [doi: 10.1186/s40345-019-0156-x] [Medline: 31552554]
24. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
25. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 2;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
26. Taylor S, Jaques N, Nosakhare E, Sano A, Picard R. Personalized multitask learning for predicting tomorrow's mood, stress, and health. IEEE Trans Affect Comput 2020;11(2):200-213 [FREE Full text] [doi: 10.1109/TAFFC.2017.2784832] [Medline: 32489521]
27. Yan S, Hosseinmardi H, Kao HT, Narayanan S, Lerman K, Ferrara E. Affect estimation with wearable sensors. J Healthc Inform Res 2020 Sep;4(3):261-294 [FREE Full text] [doi: 10.1007/s41666-019-00066-z] [Medline: 35415445]
28. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
29. Ferreri F, Bourla A, Peretti CS, Segawa T, Jaafari N, Mouchabac S. How new technologies can improve prediction, assessment, and intervention in obsessive-compulsive disorder (e-OCD): review. JMIR Ment Health 2019 Dec 10;6(12):e11643 [FREE Full text] [doi: 10.2196/11643] [Medline: 31821153]
30. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition DSM-5. Virginia, United States: American Psychiatric Association; 2013.
31. Calvocoressi L, Libman D, Vegso SJ, McDougle CJ, Price LH. Global functioning of inpatients with obsessive-compulsive disorder, schizophrenia, and major depression. Psychiatr Serv 1998 Mar;49(3):379-381. [doi: 10.1176/ps.49.3.379] [Medline: 9525801]
32. Angelakis I, Gooding P, Tarrier N, Panagioti M. Suicidality in obsessive compulsive disorder (OCD): a systematic review and meta-analysis. Clin Psychol Rev 2015 Jul;39:1-15. [doi: 10.1016/j.cpr.2015.03.002] [Medline: 25875222]
33. Hollander E, Doernberg E, Shavitt R, Waterman RJ, Soreni N, Veltman DJ, et al. The cost and impact of compulsivity: a research perspective. Eur Neuropsychopharmacol 2016 May;26(5):800-809. [doi: 10.1016/j.euroneuro.2016.02.006] [Medline: 27235690]
34. Catapano F, Perris F, Masella M, Rossano F, Cigliano M, Magliano L, et al. Obsessive-compulsive disorder: a 3-year prospective follow-up study of patients treated with serotonin reuptake inhibitors OCD follow-up study. J Psychiatr Res 2006 Sep;40(6):502-510. [doi: 10.1016/j.jpsychires.2005.04.010] [Medline: 16904424]
35. Eisen JL, Sibrava NJ, Boisseau CL, Mancebo MC, Stout RL, Pinto A, et al. Five-year course of obsessive-compulsive disorder. J Clin Psychiatry 2013 Mar 15;74(03):233-239. [doi: 10.4088/jcp.12m07657]
36. Thorsen AL, Kvale G, Hansen B, van den Heuvel OA. Symptom dimensions in obsessive-compulsive disorder as predictors of neurobiology and treatment response. Curr Treat Options Psychiatry 2018 Mar;5(1):182-194 [FREE Full text] [doi: 10.1007/s40501-018-0142-4] [Medline: 30237966]
37. Bergfeld IO, Dijkstra E, Graat I, de Koning P, van den Boom BJ, Arbab T, et al. Invasive and non-invasive neurostimulation for OCD. Curr Top Behav Neurosci 2021;49:399-436. [doi: 10.1007/7854_2020_206] [Medline: 33550567]
38. Cooper D, Champion SM, Stavropoulos L, Grisham JR. How technology can enhance treatment: a scoping review of clinical interventions for anxiety and obsessive-compulsive spectrum disorders. Br J Clin Psychol 2022 Jan;61 Suppl 1:8-30. [doi: 10.1111/bjc.12279] [Medline: 33570762]
39. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
40. Frank AC, Li R, Peterson B, Narayanan S. Scoping review of wearable and mobile devices used in the assessment and treatment of obsessive-compulsive disorder. OSF Registries. 2022 Aug 9. URL: https://osf.io/2bpwt [accessed 2023-06-30]
41. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
42. Aboujaoude E. Three decades of telemedicine in obsessive-compulsive disorder: a review across platforms. J Obsessive Compuls Relat Disord 2017 Jul;14:65-70. [doi: 10.1016/j.jocrd.2017.06.003]
43. van Loenen I, Scholten W, Muntingh A, Smit J, Batelaan N. The effectiveness of virtual reality exposure-based cognitive behavioral therapy for severe anxiety disorders, obsessive-compulsive disorder, and posttraumatic stress disorder: meta-analysis. J Med Internet Res 2022 Feb 10;24(2):e26736 [FREE Full text] [doi: 10.2196/26736] [Medline: 35142632]
44. Millet B, Touitou Y, Poirier MF, Bourdel MC, Hantouche E, Bogdan A, et al. Plasma melatonin and cortisol in patients with obsessive-compulsive disorder: relationship with axillary temperature, physical activity, and clinical symptoms. Biol Psychiatry 1998 Nov 01;44(9):874-881. [doi: 10.1016/s0006-3223(97)00512-x] [Medline: 9807642]

45. Alfano CA, Kim KL. Objective sleep patterns and severity of symptoms in pediatric obsessive compulsive disorder: a pilot investigation. J Anxiety Disord 2011 Aug;25(6):835-839 [FREE Full text] [doi: 10.1016/j.janxdis.2011.04.004] [Medline: 21570250]
46. Drummond LM, Wulff K, Rani RS, White S, Mbanga-Sibanda J, Ghodse H, et al. How should we measure delayed sleep phase shift in severe, refractory obsessive-compulsive disorder? Int J Psychiatry Clin Pract 2012 Oct;16(4):268-276. [doi: 10.3109/13651501.2012.709866] [Medline: 22809128]
47. Pittig A, Arch JJ, Lam CW, Craske MG. Heart rate and heart rate variability in panic, social anxiety, obsessive-compulsive, and generalized anxiety disorders at baseline and in response to relaxation and hyperventilation. Int J Psychophysiol 2013 Jan;87(1):19-27. [doi: 10.1016/j.ijpsycho.2012.10.012] [Medline: 23107994]
48. Bussing R, Reid AM, McNamara JP, Meyer JM, Guzick AG, Mason DM, et al. A pilot study of actigraphy as an objective measure of SSRI activation symptoms: results from a randomized placebo controlled psychopharmacological treatment study. Psychiatry Res 2015 Feb 28;225(3):440-445 [FREE Full text] [doi: 10.1016/j.psychres.2014.11.070] [Medline: 25535011]
49. Donse L, Sack AT, Fitzgerald PB, Arns M. Sleep disturbances in obsessive-compulsive disorder: association with non-response to repetitive transcranial magnetic stimulation (rTMS). J Anxiety Disord 2017 Jun;49:31-39. [doi: 10.1016/j.janxdis.2017.03.006] [Medline: 28388457]
50. Jaspers-Fayer F, Lin SY, Belschner L, Mah J, Chan E, Bleakley C, et al. A case-control study of sleep disturbances in pediatric obsessive-compulsive disorder. J Anxiety Disord 2018 Apr;55:1-7. [doi: 10.1016/j.janxdis.2018.02.001] [Medline: 29529448]
51. Coles ME, Schubert J, Stewart E, Sharkey KM, Deak M. Sleep duration and timing in obsessive-compulsive disorder (OCD): evidence for circadian phase delay. Sleep Med 2020 Aug;72:111-117. [doi: 10.1016/j.sleep.2020.03.021] [Medline: 32575000]
52. Cox RC, Olatunji BO. Delayed circadian rhythms and insomnia symptoms in obsessive-compulsive disorder. J Affect Disord 2022 Dec 01;318:94-102. [doi: 10.1016/j.jad.2022.08.118] [Medline: 36057288]
53. Thierfelder A, Primbs J, Severitt B, Hohnecker CS, Kuhnhausen J, Alt AK, et al. Multimodal sensor-based identification of stress and compulsive actions in children with obsessive-compulsive disorder for telemedical treatment. Annu Int Conf IEEE Eng Med Biol Soc 2022 Jul;2022:2976-2982. [doi: 10.1109/EMBC48229.2022.9871899] [Medline: 36085677]
54. Gajadien PT, Postma TS, van Oostrom I, Scheepstra KW, van Dijk H, Sack AT, et al. Sleep predicts the response to rTMS and CBT in patients with OCD: an open label effectiveness study. Int J Clin Health Psychol 2023;23(2):100353 [FREE Full text] [doi: 10.1016/j.ijchp.2022.100353] [Medline: 36415607]
55. Brown LA, Narine K, Asnanni A, Simon S, Majeed I, Cohen D. Implementation of technology-driven comprehensive physical health assessment in an anxiety specialty clinic: Preliminary pilot study findings. The Behavior Therapist 2020;43:19-26 [FREE Full text]
56. Gloster AT, Richard DC, Himle J, Koch E, Anson H, Lokers L, et al. Accuracy of retrospective memory and covariation estimation in patients with obsessive-compulsive disorder. Behav Res Ther 2008 May;46(5):642-655. [doi: 10.1016/j.brat.2008.02.010] [Medline: 18417100]
57. Tilley PJ, Rees CS. A clinical case study of the use of ecological momentary assessment in obsessive compulsive disorder. Front Psychol 2014 Apr 17;5:339. [doi: 10.3389/fpsyg.2014.00339] [Medline: 24860521]
58. Rupp C, Falke C, Gühne D, Doebler P, Andor F, Buhlmann U. A study on treatment sensitivity of ecological momentary assessment in obsessive-compulsive disorder. Clin Psychol Psychother 2019 Nov;26(6):695-706. [doi: 10.1002/cpp.2392] [Medline: 31365952]
59. Way to health. Penn Medicine, Center for Health Care Innovation. URL: https://waytohealth.org [accessed 2023-07-03]
60. Rupp C, Gühne D, Falke C, Doebler P, Andor F, Buhlmann U. Comparing effects of detached mindfulness and cognitive restructuring in obsessive-compulsive disorder using ecological momentary assessment. Clin Psychol Psychother 2020 Mar;27(2):193-202. [doi: 10.1002/cpp.2418] [Medline: 31881096]
61. Le Boeuf A. An automated aversion device in the treatment of a compulsive handwashing ritual. J Behav Ther Exp Psychiatry 1974 Dec;5(3-4):267-270. [doi: 10.1016/0005-7916(74)90076-7]
62. Olbrich H, Stengler K, Olbrich S. Smartphone based Geo-Feedback in obsessive compulsive disorder as facilitatory intervention: a case report. J Obs Compuls Relat Disord 2016 Jan;8:75-78. [doi: 10.1016/j.jocrd.2016.01.001]
63. Kashyap H, Reddy P, Mandadi S, Narayanaswamy JC, Sudhir PM, Reddy YC. Cognitive training for neurocognitive and functional impairments in obsessive compulsive disorder: a case report. J Obs Compuls Relat Disord 2019 Oct;23:100480. [doi: 10.1016/j.jocrd.2019.100480]
64. Arevian AC, O'Hora J, Rosser J, Mango JD, Miklowitz DJ, Wells KB. Patient and provider cocreation of mobile texting apps to support behavioral health: usability study. JMIR Mhealth Uhealth 2020 Jul 29;8(7):e12655 [FREE Full text] [doi: 10.2196/12655] [Medline: 32723714]
65. Olsen ST, Basu I, Bilge MT, Kanabar A, Boggess MJ, Rockhill AP, et al. Case report of dual-site neurostimulation and chronic recording of cortico-striatal circuitry in a patient with treatment refractory obsessive compulsive disorder. Front Hum Neurosci 2020;14:569973 [FREE Full text] [doi: 10.3389/fnhum.2020.569973] [Medline: 33192400]

66. Provenza NR, Sheth SA, Dastin-van Rijn EM, Mathura RK, Ding Y, Vogt GS, et al. Long-term ecological assessment of intracranial electrophysiology synchronized to behavioral markers in obsessive-compulsive disorder. Nat Med 2021 Dec;27(12):2154-2164 [FREE Full text] [doi: 10.1038/s41591-021-01550-z] [Medline: 34887577]
67. Hawley LL, Rector NA, DaSilva A, Laposa JM, Richter MA. Technology supported mindfulness for obsessive compulsive disorder: self-reported mindfulness and EEG correlates of mind wandering. Behav Res Ther 2021 Jan;136:103757. [doi: 10.1016/j.brat.2020.103757] [Medline: 33310604]
68. Hawley LL, Rector NA, Richter MA. Technology supported mindfulness for obsessive compulsive disorder: the role of obsessive beliefs. J Anxiety Disord 2021 Jun;81:102405. [doi: 10.1016/j.janxdis.2021.102405] [Medline: 33910091]
69. Fridgeirsson EA, Bais MN, Eijsker N, Thomas RM, Smit DJ, Bergfeld IO, et al. Patient specific intracranial neural signatures of obsessions and compulsions in the ventral striatum. J Neural Eng 2023 Mar 10;20(2). [doi: 10.1088/1741-2552/acbee1] [Medline: 36827705]
70. Motti VG, Caine K. Users' privacy concerns about wearables. In: Brenner M, Christin N, Johnson B, Rohloff K, editors. Financial Cryptography and Data Security. Berlin, Heidelberg: Springer; 2015.
71. Bauer M, Glenn T, Geddes J, Gitlin M, Grof P, Kessing LV, et al. Smartphones in mental health: a critical review of background issues, current status and future concerns. Int J Bipolar Disord 2020 Jan 10;8(1):2 [FREE Full text] [doi: 10.1186/s40345-019-0164-x] [Medline: 31919635]
72. Torous J, Wisniewski H, Liu G, Keshavan M. Mental health mobile phone app usage, concerns, and benefits among psychiatric outpatients: comparative survey study. JMIR Ment Health 2018 Nov 16;5(4):e11715 [FREE Full text] [doi: 10.2196/11715] [Medline: 30446484]
73. Hathaliya JJ, Tanwar S. An exhaustive survey on security and privacy issues in Healthcare 4.0. Comput Commun 2020 Mar;153:311-335. [doi: 10.1016/j.comcom.2020.02.018]
74. Narayanan S, Georgiou PG. Behavioral signal processing: deriving human behavioral informatics from speech and language: computational techniques are presented to analyze and model expressed and perceived human behavior-variedly characterized as typical, atypical, distressed, and disordered-from speech and language cues and their applications in health, commerce, education, and beyond. Proc IEEE Inst Electr Electron Eng 2013 Feb 07;101(5):1203-1233 [FREE Full text] [doi: 10.1109/JPROC.2012.2236291] [Medline: 24039277]
75. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
76. Torous J, Onnela JP, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry 2017 Mar 07;7(3):e1053 [FREE Full text] [doi: 10.1038/tp.2017.25] [Medline: 28267146]
77. Adler DA, Ben-Zeev D, Tseng VW, Kane JM, Brian R, Campbell AT, et al. Predicting early warning signs of psychotic relapse from passive sensing data: an approach using encoder-decoder neural networks. JMIR Mhealth Uhealth 2020 Aug 31;8(8):e19962 [FREE Full text] [doi: 10.2196/19962] [Medline: 32865506]
78. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
79. Kathan A, Harrer M, Küster L, Triantafyllopoulos A, He X, Milling M, et al. Personalised depression forecasting using mobile sensor data and ecological momentary assessment. Front Digit Health 2022;4:964582 [FREE Full text] [doi: 10.3389/fdgth.2022.964582] [Medline: 36465087]
80. Jacobson NC, Feng B. Digital phenotyping of generalized anxiety disorder: using artificial intelligence to accurately predict symptom severity using wearable sensors in daily life. Transl Psychiatry 2022 Aug 17;12(1):336 [FREE Full text] [doi: 10.1038/s41398-022-02038-1] [Medline: 35977932]
81. Demiris G, Corey Magan KL, Parker Oliver D, Washington KT, Chadwick C, Voigt JD, et al. Spoken words as biomarkers: using machine learning to gain insight into communication as a predictor of anxiety. J Am Med Inform Assoc 2020 Jun 01;27(6):929-933 [FREE Full text] [doi: 10.1093/jamia/ocaa049] [Medline: 32374378]
82. McKernan EP, Kumar M, Di Martino A, Shulman L, Kolevzon A, Lord C, et al. Intra-topic latency as an automated behavioral marker of treatment response in autism spectrum disorder. Sci Rep 2022 Feb 28;12(1):3255 [FREE Full text] [doi: 10.1038/s41598-022-07299-w] [Medline: 35228613]
83. Lahiri R, Nasir M, Kumar M, Kim SH, Bishop S, Lord C, et al. Interpersonal synchrony across vocal and lexical modalities in interactions involving children with autism spectrum disorder. JASA Express Lett 2022 Sep;2(9):095202 [FREE Full text] [doi: 10.1121/10.0013421] [Medline: 36097603]
84. Barron DS, Baker JT, Budde KS, Bzdok D, Eickhoff SB, Friston KJ, et al. Decision models and technology can help psychiatry develop biomarkers. Front Psychiatry 2021;12:706655 [FREE Full text] [doi: 10.3389/fpsyt.2021.706655] [Medline: 34566711]
85. Schumacher S, Kent N. 8 charts on internet use around the world as countries grapple with COVID-19. Pew Research Center. 2020 Apr 2. URL: https://www.pewresearch.org/short-reads/2020/04/02/8-charts-on-internet-use-around-the-world-as-countries-grapple-with-covid-19/ [accessed 2022-10-12]

86. Wike R, Silver L, Fetterolf J, Huang C, Austin S, Clancy L, et al. Internet, smartphone and social media use. Pew Research Center. 2022. URL: https://www.pewresearch.org/global/2022/12/06/internet-smartphone-and-social-media-use-in-advanced-economies-2022/ [accessed 2022-10-12]
87. Naveed S, Waqas A, Chaudhary AM, Kumar S, Abbas N, Amin R, et al. Prevalence of common mental disorders in South Asia: a systematic review and meta-regression analysis. Front Psychiatry 2020 Sep 2;11:573150 [FREE Full text] [doi: 10.3389/fpsyt.2020.573150] [Medline: 32982812]
88. Cai H, Jin Y, Liu R, Zhang Q, Su Z, Ungvari GS, et al. Global prevalence of depression in older adults: a systematic review and meta-analysis of epidemiological surveys. Asian J Psychiatr 2023 Feb;80:103417. [doi: 10.1016/j.ajp.2022.103417] [Medline: 36587492]
89. Williams M, Powers M, Yun YG, Foa E. Minority participation in randomized controlled trials for obsessive-compulsive disorder. J Anxiety Disord 2010 Mar;24(2):171-177 [FREE Full text] [doi: 10.1016/j.janxdis.2009.11.004] [Medline: 20143498]
90. Wetterneck CT, Little TE, Rinehart KL, Cervantes ME, Hyde E, Williams M. Latinos with obsessive-compulsive disorder: mental healthcare utilization and inclusion in clinical trials. J Obsessive Compuls Relat Disord 2012 Apr;1(2):85-97 [FREE Full text] [doi: 10.1016/j.jocrd.2011.12.001] [Medline: 29057210]
91. Wong CK, Ho DT, Tam AR, Zhou M, Lau YM, Tang MO, et al. Artificial intelligence mobile health platform for early detection of COVID-19 in quarantine subjects using a wearable biosensor: protocol for a randomised controlled trial. BMJ Open 2020 Jul 22;10(7):e038555 [FREE Full text] [doi: 10.1136/bmjopen-2020-038555] [Medline: 32699167]
92. Maaß L, Freye M, Pan CC, Dassow HH, Niess J, Jahnel T. The definitions of health apps and medical apps from the perspective of public health and law: qualitative analysis of an interdisciplinary literature overview. JMIR Mhealth Uhealth 2022 Oct 31;10(10):e37980 [FREE Full text] [doi: 10.2196/37980] [Medline: 36315221]
93. Simon DA, Shachar C, Cohen IG. Unsettled liability issues for "prediagnostic" wearables and health-related products. JAMA 2022 Oct 11;328(14):1391-1392. [doi: 10.1001/jama.2022.16317] [Medline: 36136366]

Abbreviations

CBT: cognitive behavioral therapy
DBS: deep brain stimulation
EMA: ecological momentary assessment
ERP: exposure-response prevention
HRV: heart rate variability
LFP: local field potential
OCD: obsessive-compulsive disorder
rTMS: repetitive transcranial magnetic stimulation
Y-BOCS: Yale-Brown Obsessive Compulsive Scale

Edited by J Torous; submitted 15.01.23; peer-reviewed by D Valdes, A Angelucci, M Khazaneha, XZ Hu; comments to author 14.04.23; revised version received 27.05.23; accepted 13.06.23; published 18.07.23.

Please cite as:
Frank AC, Li R, Peterson BS, Narayanan SS
Wearable and Mobile Technologies for the Evaluation and Treatment of Obsessive-Compulsive Disorder: Scoping Review
JMIR Ment Health 2023;10:e45572
URL: https://mental.jmir.org/2023/1/e45572
doi:10.2196/45572
PMID:37463010

Review

Use of the Maslach Burnout Inventory Among Public Health Care Professionals: Scoping Review

Juliana Pontes Soares[1], PhD; Rayssa Horácio Lopes[1], PhD; Paula Beatriz de Souza Mendonça[2], MSc; Cícera Renata Diniz Vieira Silva[3], PhD; Cláudia Cristiane Filgueira Martins Rodrigues[4], PhD; Janete Lima de Castro[1], PhD

[1]Postgraduate Program in Collective Health, Federal University of Rio Grande do Norte, Natal, Brazil

[2]Postgraduate Program in Collective Health, Federal University of Espírito Santo, Vitória, Brazil

[3]Technical School of Health of Cajazeiras, Federal University of Campina Grande, Cajazeiras, Brazil

[4]School of Health, Federal University of Rio Grande do Norte, Natal, Brazil

Corresponding Author:
Juliana Pontes Soares, PhD
Postgraduate Program in Collective Health
Federal University of Rio Grande do Norte
Av Senador Salgado Filho, 3000 – Bairro: Lagoa Nova
Natal, 59078–970
Brazil
Phone: 55 83991627740
Email: juliana.pontes@ufrr.br

Abstract

Background: Work can be considered a source of living, well-being, and socioeconomic development. When the work environment negatively influences individuals, it may trigger emotional disturbances, behavioral problems, chronic stress conditions, and illnesses such as burnout syndrome (BS). Recently, studies on BS have increased and placed a special focus on health care professionals. The prevalence of BS among health professionals is associated with their chronic exposure to human hardship and long working hours without proper rest. These factors have contributed to greater stress and high physical and emotional exhaustion levels.

Objective: This study aims to identify and map studies using the Maslach Burnout Inventory (MBI) scale to identify burnout syndrome in health professionals working in public health services.

Methods: This scoping review was developed based on the Joanna Briggs Institute (JBI) Reviewers Manual and reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews). A total of 6 databases were searched to identify relevant studies: Embase, LILACS, MEDLINE or PubMed, PsycInfo, Scopus, and Web of Science. Gray literature was consulted on ProQuest Dissertations and Theses Global, Google Scholar, Brazilian Digital Library of Theses and Dissertations, and Open Access Theses and Dissertations. Additionally, the reference lists were searched to retrieve studies not previously selected. The steps followed in this study were proposed by Arskey and O'Malley and Levac, Colquhoun, and O'Brien: identification of research questions, identification of potential studies, study selection, data extraction and imputation, data analyses and interpretation, and consultation with stakeholders. The detailed methodology was published in a protocol.

Results: A total of 55 articles were identified after screening for eligibility criteria, published between 1999 and 2021 in 32 countries. Most reports were published in Brazil, Spain, and China. A total of 22 versions of the MBI were identified, presenting different items, scores, and cutoff points. The included studies had recommendations and implications for clinical practice. The consultation with stakeholders allowed knowledge translation for those interested in BS.

Conclusions: Studies mostly included physicians (34/55, 61.8%) and nurses (24/55, 43.6%), and the original version of MBI was predominantly used. Divergences in BS classification were highlighted, which may be related to MBI cross-cultural adaptations and applications in other countries. This study contributes to the advancement of research regarding burnout syndrome as an occupational illness since it has harmful consequences for workers, health care services, and the quality of care provided to the population.

(JMIR Ment Health 2023;10:e44195) doi:10.2196/44195

KEYWORDS

burnout professional; burnout; health care professional; health personnel; health professionals; Maslach Burnout Inventory; mental health; occupational health; public health services; public health; workplace stress

Introduction

Work can be considered a source of living, well-being, and socioeconomic development. When the work environment negatively influences individuals, it may trigger emotional disturbances, behavioral problems, chronic stress conditions, and illnesses such as burnout syndrome (BS) [1-3]. This term was coined by the psychoanalyst Freudenberger [4] after studying unwanted emotional and interpersonal exhaustion factors that affected people working in health care services [4]. Later, the social psychologist Maslach studied emotions in the workplace and their relationships with illness. The initial study indicated that clinical symptoms of BS were associated with the mental health and social status of the care provider and recipient in the occupational context [1].

The World Health Organization incorporated BS in the latest version of the International Classification of Diseases (ICD-11) as an "occupational phenomenon" resulting from poor management of a demanding work environment [5,6]. Recently, studies on BS have increased and placed special focus on health care professionals [3,7-9]. The prevalence of BS among health care professionals is associated with their chronic exposure to human hardship and long working hours without proper rest. These factors have contributed to greater stress and high physical and emotional exhaustion levels [3,6,10,11].

The public health care workforce is mentioned as the most exposed to BS [11]. Prolonged psychophysical distress, emotional exhaustion, and a lack of personal accomplishment negatively influence occupational performance and mental health. In addition, occupational illness is globally recognized as a public health problem because it encompasses dimensions that go beyond health. High rates of absenteeism, job turnover, and social security expenses also impact the socioeconomic dimension [10,11].

The assessment of BS encompasses a systematic approach based on psychometric variables of a scale conceived by Maslach and Jackson [12], called the Maslach Burnout Inventory (MBI). The MBI is a self-administered scale composed of affirmative sentences about feelings and attitudes toward work. It has 3 dimensions: emotional exhaustion, depersonalization, and a lack of personal accomplishment. This multidimensional model adopted by Maslach and Jackson is the most cited in the literature about BS [1-3,13].

The version published in 1981 was progressively updated, translated, and adapted to meet the needs of different professional groups worldwide [1-3]. The study conducted by De Hert [3] presented 5 versions of the MBI scale: Human Services Survey (MBI-HSS), Human Services Survey for Medical Personnel (MBI-HSS-[MP]), Educators Survey (MBI-ES), General Survey (MBI-GS), and General Survey for Students (MBI-GS [S]). The high number of adapted scales is due to the diverse occupational sectors and the attempt to find meaningful solutions to prevent occupational illness [3,7,8]. Studies using the MBI provided evidence to guide managers and other health authorities in planning, implementing, and evaluating interventions [3,5,7,9]. Moreover, these studies helped develop guidelines to stimulate ideas on workplace organization and improve the physical and mental well-being of health professionals [3,5,7,9].

Currently, a gap concerning mapping the use of MBI among health professionals working in public services has been identified. The MBI is mentioned in the literature as the gold standard for evaluating BS [10], and despite its importance, its use is not standardized. This makes it difficult to compare studies, even when they are developed from the same perspective. Thus, the results of this scoping review may direct initiatives to think about a possible standardization in the use of the MBI, contributing to the advancement of research aimed at understanding the relationship between work and mental illness among professionals. Considering the importance of MBI and the exposure to occupational factors that may lead to mental suffering, this scoping review aimed to identify and map studies using the MBI scale to identify BS in health professionals working in public health services.

Methods

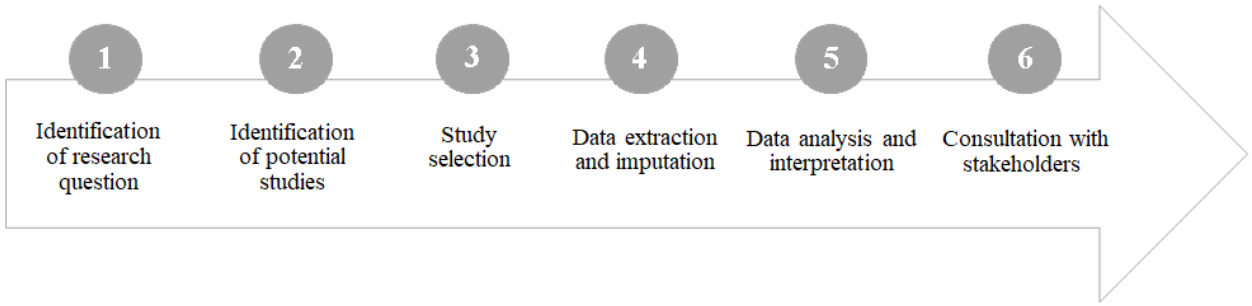
Study Methodology

This scoping review was developed based on the Joanna Briggs Institute Reviewers Manual (JBI) [14] and reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [15]. The steps followed in this study were proposed by Arskey and O'Malley [16] and Levac et al [17] (Figure 1). The detailed methodology was published in a protocol [18]. The main changes in this review were related to the analysis of the latent profile, the form of application of the scales, and the challenges and limitations that were not identified in the included studies.

The research question was developed based on the acronym PCC (Population, Concept, Context) [14]: P–Public health care professionals; C–Maslach Burnout Inventory; and C–BS in health care professionals. The following research questions were identified: (1) How is MBI used to identify BS in health care professionals working in public health services? (2) What is the most studied professional category using MBI? (3) What are the main results of MBI in health care professionals working for public health services? And (4), what are the recommendations for clinical practice arising from the use of MBI?

A total of 6 databases were searched to identify relevant studies: EMBASE, LILACS, MEDLINE or PubMed, PsycInfo, Scopus, and the Web of Science. Gray literature was consulted on ProQuest Dissertations and Theses Global, Google Scholar, Brazilian Digital Library of Theses and Dissertations, and Open Access Theses and Dissertations. Additionally, reference lists were searched to retrieve studies not previously selected.

Figure 1. Steps of the scoping review.



Inclusion criteria encompassed quantitative and qualitative studies that used the MBI scale to identify BS in health care professionals working in public health services, full-text availability, and no language or date restriction. Exclusion criteria encompassed duplicate publications, literature reviews, letters, editorials, theoretical essays, and opinion articles; studies that analyzed BS in other professionals, professionals working in private services, or undergraduate students; studies that used different scales, did not isolate the MBI in the results section, or did not use the MBI to investigate BS; and studies with an unavailable full-text version.

Searches and screenings occurred from February 2 to April 6, 2022. The search strategy was developed using 4 controlled vocabularies (DeCS, MeSH, EMTREE, and APA THESAURUS) [18]. The extraction, conversion, combination, construction, and use model [19] was used and refined with the help of a librarian. Natural and controlled languages were chosen for greater sensitivity and expansion of search results [19].

Selected studies were exported to a Microsoft Excel spreadsheet, and duplicates were manually excluded. A total of 2 independent researchers screened the titles and abstracts, and a third reviewer was consulted to resolve possible disagreements. As recommended by the JBI Manual [14], a pilot study was conducted to evaluate the consistency of the protocol [18] and select studies according to titles and abstracts. Subsequently, full texts were retrieved, and reference lists of the included studies were analyzed.

Data extraction and imputation ensured the consistency and reliability of the results. A total of 2 independent researchers used a data extraction sheet, adapted based on the JBI model [14], containing the following: study, type of study, year of publication, the context of publication, journal, study aims, study design, population, sample, MBI version, MBI domains (emotional exhaustion, depersonalization, and lack of personal accomplishment), number of items, type of Likert scale, cutoff point, results, challenges and limitations in using the scale, and implications for practice.

Quantitative variables (type of publication, year, country, study design, studied professional category, scale versions, domains, number of items, and Likert and cut-off points) were analyzed descriptively and presented as absolute and relative frequencies. Qualitative variables related to recommendations and implications for practice were processed using the IRAMUTEQ software [20]. Textual fragments extracted from the results and conclusion sections produced a text corpus that was further analyzed using similarity analysis.

The results of this review were presented to 7 stakeholders (researchers with experience in the use of MBI and BS), who answered questions regarding the findings in compliance with Levac et al [17]. This procedure intends to exchange knowledge and develop strategies for disseminating results and ideas for future studies. The following questions were asked:

1. In the scientific field, disseminating research findings is very important for academia, health services, management, health professionals, and the entire community. Therefore, we would like your suggestions on ways to disseminate the results of this study.
2. How do you think this study could contribute to standardizing the use of MBI?
3. Based on this scoping review, what ideas would you have for future studies?

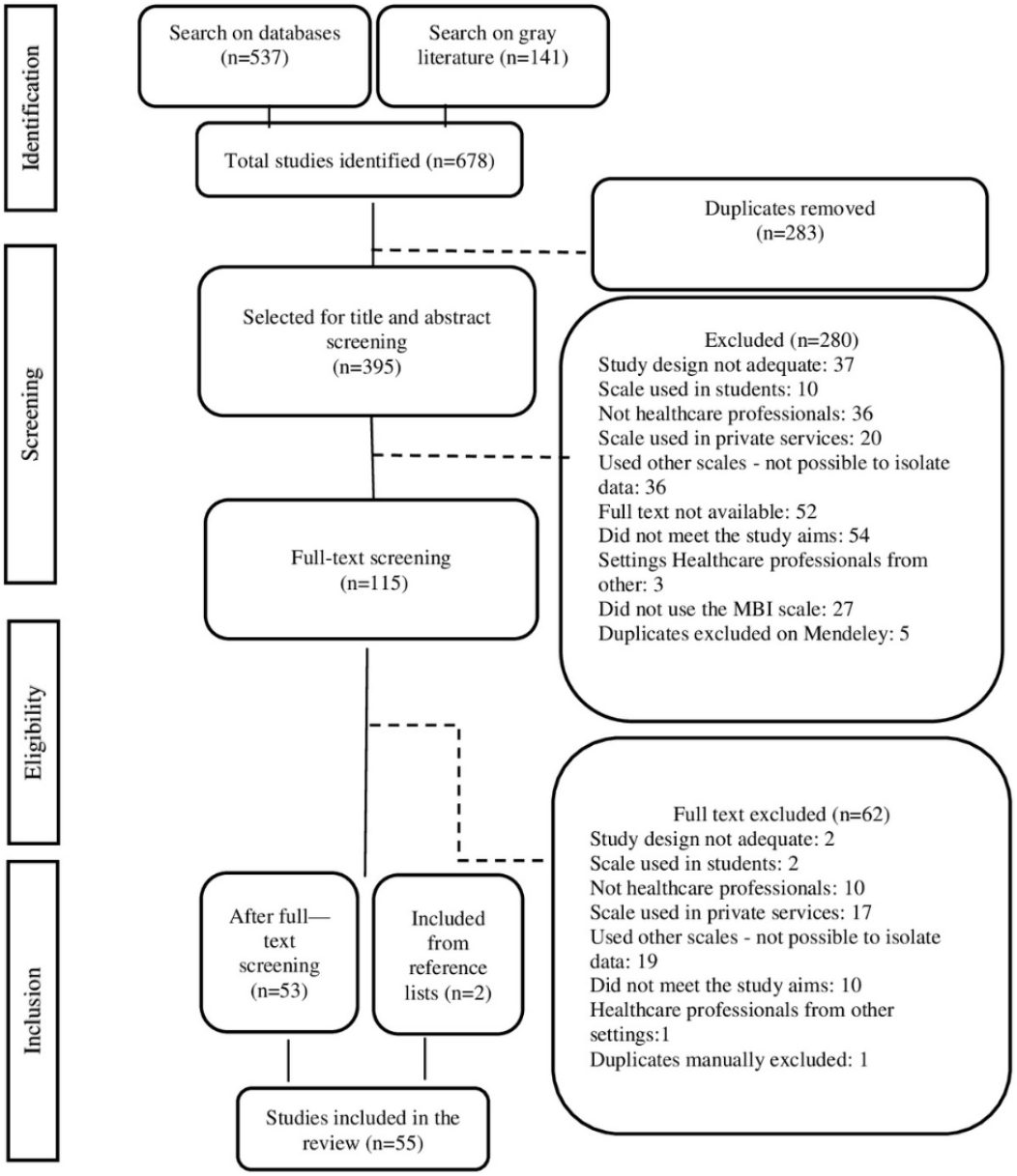
Ethics Approval

Given the participation of stakeholders, the study was submitted and approved by the research ethics committee of the Onofre Lopes University Hospital of the Federal University of Rio Grande do Norte (No. 4,952,319 and CAAE No. 46284921.4.0000.5292) on September 3, 2021.

Results

A total of 678 publications were identified, of which 537 were peer-reviewed studies and 141 were retrieved from gray literature. After screening and eligibility criteria, 53 studies were eligible, and 2 studies were identified after consulting the reference lists; thus, 55 studies were included in this scoping review (Figure 2).

Figure 2. Flowchart of study selection adapted from the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses).



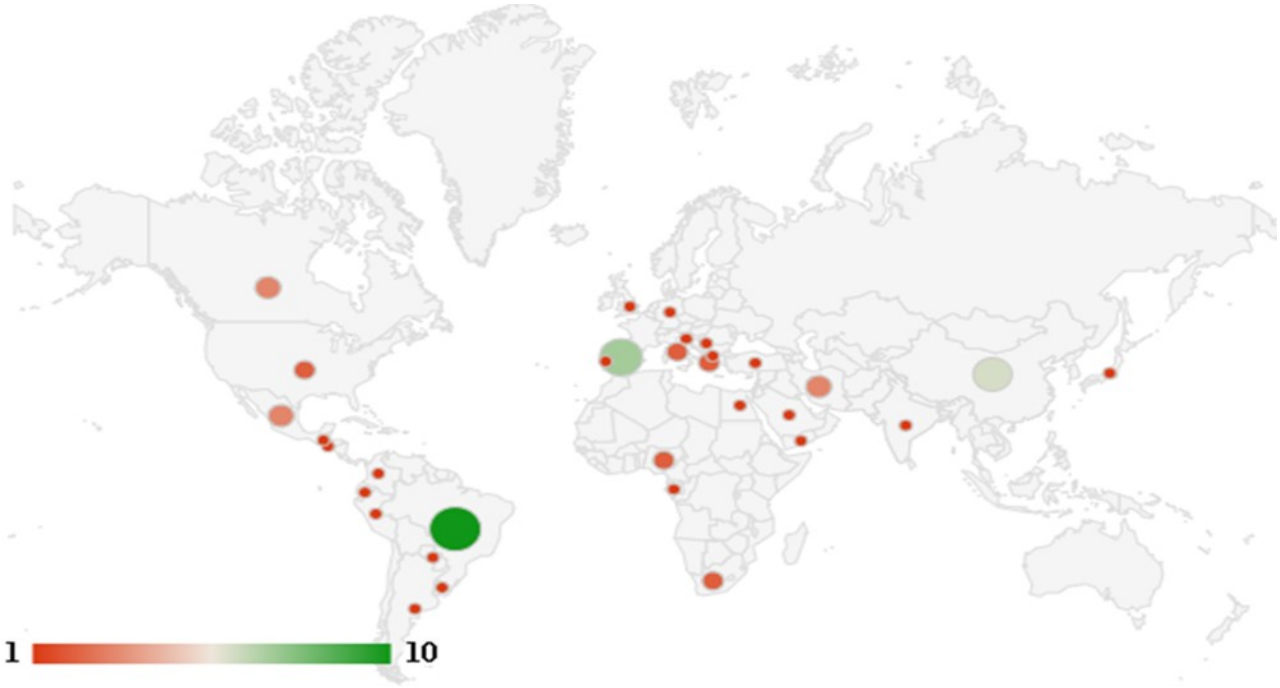
Characteristics of Included Studies

A total of 55 scientific articles published between 1999 and 2021 were included; most were published between 2016 and 2021 (n=32, 58.1%), mainly in 2018 (n=8, 14.5%) (Multimedia Appendix 1).

Studies were developed in several countries, including Brazil (n=10, 15.8%), Spain (n=7, 11.1%), and China (n=6, 9.5%) (Figure 3). For mapping purposes, we adopted the total number of countries (n=63) since 1 study was developed in more than one country.

The study design was reported in 44 (80%) studies; most were cross-sectional (n=37, 67.2%). Studies assessed a single health profession (n=40, 72.7%) or more than one category (n=15, 27.2%). The MBI scale was predominantly used to assess physicians (n=34, 61.8%) and nurses (n=24, 43.6%).

Figure 3. Distribution of studies according to country of origin.



Characteristics of the Maslach Burnout Inventory

A total of 21 versions of the MBI were identified. The original version was the most used (20/55, 36.3%) [20-39] as shown in Table 1.

The MBI scale comprises 3 domains: emotional exhaustion, depersonalization, and lack of personal accomplishment. Most studies (53/55, 96.3%) analyzed all 3 domains [21-73].

A total of 42 (76.3%) studies used an MBI scale composed of 22 items [21-32,34-39,41,42,45,46,48-58,60,62-64,66,68,69,71,72]. Additionally, 11 (20%) studies did not describe which MBI items were used [33,43,47,59,61,65,68,70,73-75].

Answers to the MBI were distributed on a Likert-type scale, and most studies used a 0 to 6 score (26/55, 47.2%) [21,23-25,28,30-32,36,37,41-44,48,49,51,54,57-59,62,64,66,67,69,74]. A total of 16 (29%) studies did not describe how they rated the scores [22,27,33,38,46,47,50,53,55,56,61,65,68,71,75].

Studies used different ways to classify the scores of each MBI domain. The most used cutoff point (10/55 studies, 18.1%) established high (≥27 points), medium (from 19 to 26 points), and low scores (≤19 points) for emotional exhaustion. Depersonalization was considered high (≥10), medium (from 6 to 9), and low (≤6). Personal accomplishment was also represented as high (≤33), medium (from 34 to 39), and low (≥40) [27,31,34,38,58,59,61,65,71,75].

Table 1. Distribution of Maslach Burnout Inventory (MBI) versions used in the included studies.

Version of the MBI	Studies, n (%)
MBI original	20 (36.3)
MBI-HSS[a]	9 (16.3)
Spanish version of MBI	3 (5.4)
Spanish version of MBI-HSS	3 (5.4)
Chinese version of MBI-GS[b]	2 (3.6)
Chinese version of MBI-HSS	2 (3.6)
Short version of MBI	2 (3.6)
French version	1 (1.8)
Persian version	1 (1.8)
Brazilian version	1 (1.8)
Greek version	1 (1.8)
Portuguese version	1 (1.8)
Japanese version	1 (1.8)
English version translated to Arabic	1 (1.8)
German version	1 (1.8)
South African version	1 (1.8)
Turkish version	1 (1.8)
French version for physicians	1 (1.8)
Version validated by Robayo Tamayo	1 (1.8)
Brazilian version validated by Lautert	1 (1.8)
Version validated and adapted by Benevides-Pereira	1 (1.8)

[a]MBI-HSS: MBI Human Services Survey.

[b]MBI-GS: MBI General Survey.

Recommendations and Implications for Practice: Similarity Analysis

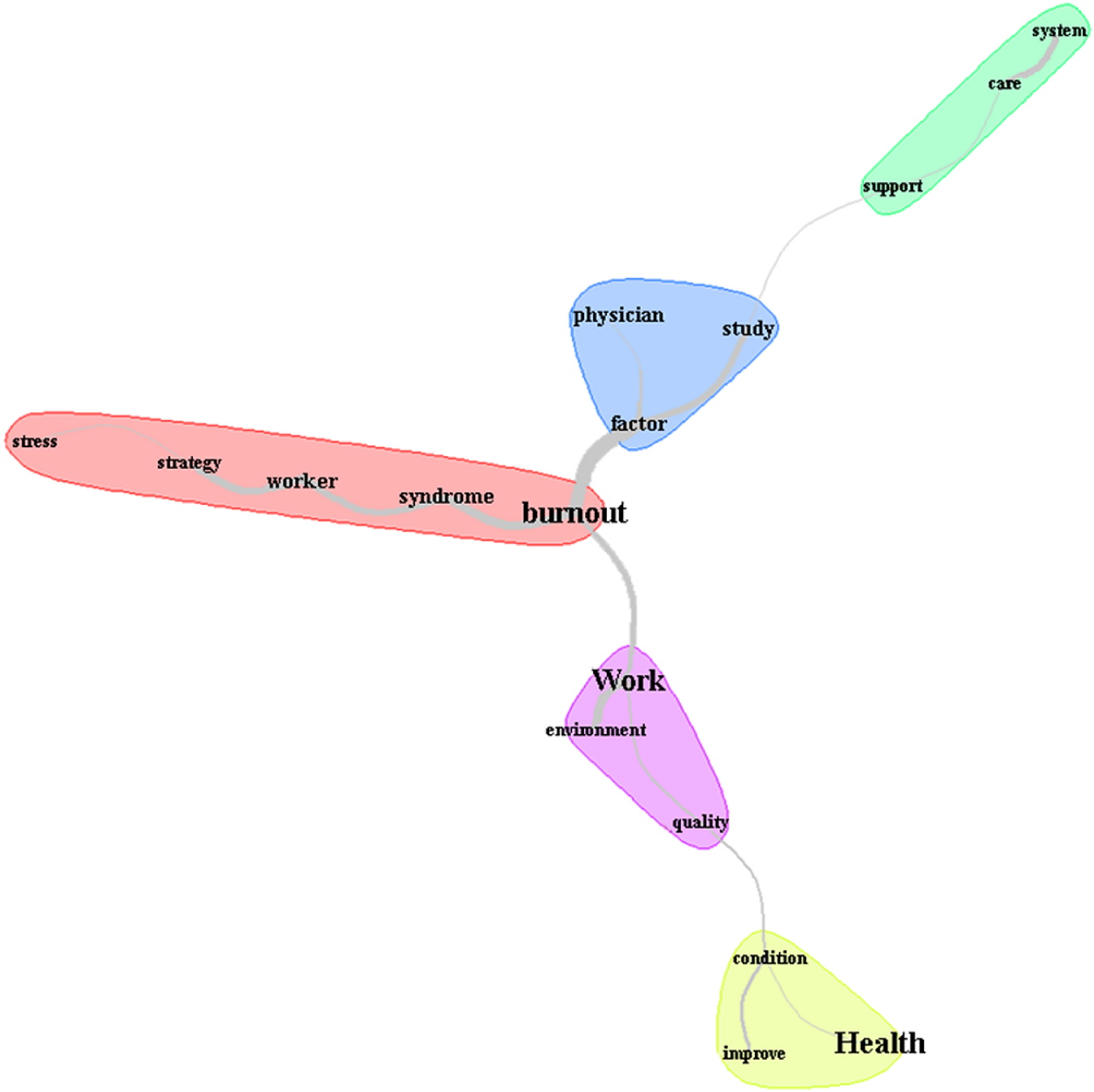
A text corpus was developed based on information from the included studies to analyze the recommendations and implications for practice. The text corpus underwent a lexicographic analysis using the IRAMUTEQ software 19 to generate the similarity analysis. The analyzed text corpus consisted of 49 texts (87 text segments) and a 64% usage rate (2933 occurrences and 875 forms). In the similarity analysis, the cutoff point was adopted as the triple of the quotient between occurrences and forms; thus, words with a frequency of ≥10 showing theoretical coherence with the research topic were inserted.

The image created was composed of 5 community halos (Figure 4) that highlighted the words “health,” “work,” “burnout,” and “physician.” The central community was formed with the word “burnout,” branching into “syndrome,” “worker,” “strategy,” and “stress.” The central community branches into 2 communities in the upper portion. The first includes the words “physician,” “factor,” and “study,” and joins with the words “system,” “support,” and “care.” In the lower portion, the central community also branches into 2 communities, revealing the words “work,” “environment,” and “quality,” joining with “health,” “condition,” and “improve.”

According to the included studies, recommendations and implications for practice addressing BS require strategies focused on stress management and strengthening the health systems. Health systems should develop studies regarding factors in the work environment impacting the quality of care. This approach could improve the health conditions of professionals, especially physicians.

The included studies did not report limitations or challenges in using the MBI. Also, a causal relationship between events was not established because most study designs were cross-sectional.

Figure 4. Image created from the similarity analysis with recommendations and implications for practice based on using the Maslach Burnout Inventory (MBI) scale. Fruchterman Reingold presentation, chi-square score with community and halo.



Discussion

Main Findings

Mapping the literature on the Maslach Burnout Inventory allowed for the observation of ways to identify the presence of BS in professionals working in public health services. Several studies using the MBI were developed in Brazil, a country of continental dimensions and cultural diversity that challenges the health status of professionals [76]. The structural challenges in the public Brazilian Unified Health System (SUS) have been marginally addressed, often in a fragmented and superficial way [77].

Most professionals within the SUS face conditions including precarious and fragile employment contracts, low wages, a high workload due to a lack of complete teams, poor working conditions and infrastructure, and a lack of supplies and equipment. The sum of these conditions compromises the health status and may result in physical and mental illness [77-79]. Work-related mental disorders, including BS, may be associated with work organization in the health field. This relationship could explain the high number of studies conducted in Brazil and identified in this review [29,31,35,38,52,55,63,71-73].

Spain was the second country in the number of studies included in this review [21,34,45,46,48-50]. Research shows that health systems in Brazil and Spain have similarities linked to the implementation of national health systems, organizational principles, universal coverage, and reorganization involving primary health care. Also, both countries face operational challenges in ensuring universality, dealing with budget cuts, and managing the expansion of private health insurance [80,81].

Studies using the MBI for assessing BS focused most on physicians. Literature reports a prevalence of BS close to or higher than 50% among these professionals, especially for those involved with direct patient care [82-86]. BS negatively impacts patients, health organizations, and systems, increasing the risks of errors during medical procedures [82,84,87]. The predictors of BS among physicians include high workload (eg, long and stressful night shifts), work conflicts, increased digitization and work bureaucracies, and a lack of continued education, support from colleagues, and autonomy at work [82-84,88]. The second professional category most present in the studies was nursing. These professionals are also exposed to a greater workload as well as emotional stressors because they are closer to patients and because they often work in different work environments despite the low wages offered to the category. This sum of factors results in negative impacts on the mental health of this workforce, contributing to the development of BS [89].

Among the selected studies, physicians and nurses worked mostly in primary health care (PHC) services [22,29,35,36,42,63,65]. It is known that this level of care is considered the gateway to health systems and therefore receives a greater demand from users who seek the service because it is closer to their homes. This greater demand overloads professionals, leaving them susceptible to the development of BS [62].

Regarding the type of study, a cross-sectional design was predominantly used. This design favors the description of specific population characteristics at a single point in time, contributing to a representative sample of the population. Advantages of the cross-sectional design include low cost, low risk of data loss, and identification of the prevalence of a certain phenomenon. However, this design does not establish a causal relationship between events [90,91].

Although the MBI neither identifies antecedent nor consequent factors linked to BS, it allows the identification of suggestive or onset signs of the syndrome [92]. Despite its general use, no consensus on how to interpret the MBI is currently available, justifying the differences in the scoring and description of results among studies. Variations in the frequency and cutoff points of the scale require accurate description in each study to avoid divergences in the results [6,93]. Most studies included in this review used the score (Likert-type scale and number of items) corresponding to the original version of the MBI. However, several studies did not follow the cutoff points of the MBI guidelines, which include high scores in the subscales of emotional exhaustion and depersonalization and low scores in personal performance. Studies mostly considered high scores in all 3 subscales as indicative of BS [27,31,34,38, 58,59,61,65,71,75].

In addition to adapted, translated, and validated versions, an Indian study translated the MBI into 3 indigenous languages to better fit the studied population [74]. The various ways of interpreting the MBI show that some authors consider the tridimensionality (ie, high scores for emotional exhaustion and depersonalization and low scores for personal performance) for the BS outcome [12]. Other authors consider only 1 dimension, regardless of which shows the higher scores [94], or 2 dimensions (high scores for emotional exhaustion and depersonalization) [95]. In this review, most studies used tridimensionality as the BS outcome, as recommended by the original version.

The cross-cultural adaptation of the MBI is crucial, but differences in interpreting the scale hamper the comparison among studies [6,93]. One study used previously described cutoff points in a population, thus, allowing the comparison of results with other publications [63]. Another study used the cutoff points recommended by the MBI guideline to enable international comparison [21].

BS challenges health care systems worldwide, affecting between 25% and 75% of health care professionals [96]. The prevalence also varies between countries, professional specialties, and work sectors [96]. Regarding recommendations and implications for practice arising from the use of MBI, our results highlight the importance of a holistic view of health for professionals working in public health services. Furthermore, the findings support the implementation of prevention programs that should be based on a set of individual and organizational strategies.

Implementing strategies that minimize occupational stress may contribute to reducing BS in health services. Strategies must be aimed at the worker individually and collectively in an organizational way. The reorganization of the work process, staff sizing according to the needs of the demands, distribution of activities according to the response capacity, recognition of the first signs that characterize BS, modification of stressors, elaboration and implementation of organizational policies that improve the quality of life and offer emotional support, as well as the inclusion of integrative and complementary practices [97-99], are strategies that may positively impact the health and well-being of health professionals and improve the quality of care for patients seeking public health services worldwide.

Stakeholder Consultation

When asked about ways of disseminating our research findings, stakeholders reported the dissemination of the results could be performed in the following ways: scientific articles, including graphs and tables; in health centers, support centers for health workers, and with health care managers; social media networks (ie, connecting researchers in teaching and management institutions); indexed journals in international, open, and peer-reviewed databases, printed or web-based; and scientific events.

In response to the contributions of this review to standardizing the use of MBI, participants answered that the scale was broad and reliable, and its standardization would be important for allowing comparisons among studies. They also reported that standardizing the use of MBI was needed to critically evaluate and compare the different versions and indicate their advantages and disadvantages.

Regarding suggestions for future studies, participants emphasized the importance of developing studies with primary health care professionals and professors since they are exposed to several psychological risks but are little studied. Participants highlighted that the scale was used exclusively to assess BS and did not consider the antecedents or consequences of the

syndrome. Therefore, studies on different ways of measuring the predisposition to BS are needed. Future studies should also include a systematic review with meta-analysis and provide robust data on the profiles of participants included in studies using the MBI.

Strengths and Limitations of the Study

This scoping review identified and mapped the use of the MBI scale among professionals in public health services. We expect an increase in the number of publications on the use of MBI in professionals working in these settings, especially cross-sectional studies, as they are important to explore local contexts.

This research followed the methodological recommendations proposed by the JBI and was conducted without date or language limits to allow a broad inclusion of the literature. Furthermore, we shared the main results with 7 specialists in the area, favoring knowledge transfer with interested parties [17].

A limitation of this study is that some relevant databases may have been missed despite extensive searches. Moreover, although we double-checked the data analysis, loss of information may have occurred during the translation of the studies into Portuguese (ie, the native language of the authors). Another possible limitation concerns the inclusion of public health professionals in the selection of studies. We highlight the difficulty of identifying this workforce in studies where it was not clearly stated that it was a public health service, and therefore some important information may not have been included due to the exclusion of studies that did not show the scenario in which the study was conducted.

Conclusions

We identified and mapped the use of the MBI scale among professionals from public health services in studies from different countries, mainly performed with physicians and nurses. Among different versions of the Maslach Burnout Inventory, the original instrument is the most widely applied to study BS in health professionals from public health systems.

The findings identified divergences regarding how studies assessed BS, probably due to cross-cultural adaptations of the scale in different countries. The study contributes to the advancement of research on BS as an occupational illness since, when answering the MBI, the participants refer to situations experienced in their work environment, reinforcing the relationship between the stressors present in the work environment and the development of BS, which characterizes it as an occupational disease that can have consequences for workers, for health services, and for the quality of the assistance provided to the population.

Acknowledgments

The authors would like to thank Wánderson Cássio Oliveira Araújo for their support in developing the search strategy and Probatus Academic Services for providing scientific language translation, revision, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of included studies.
[DOCX File , 47 KB - mental_v10i1e44195_app1.docx]

References

1. Maslach C, Schaufeli WB, Leiter MP. Job burnout. Annu Rev Psychol 2001;52:397-422. [doi: 10.1146/annurev.psych.52.1.397] [Medline: 11148311]
2. Shaikh AA, Shaikh A, Kumar R, Tahir A. Assessment of burnout and its factors among doctors using the abbreviated maslach burnout inventory. Cureus 2019;11(2):e4101 [FREE Full text] [doi: 10.7759/cureus.4101] [Medline: 31057995]
3. De Hert S. Burnout in healthcare workers: prevalence, impact and preventative strategies. Local Reg Anesth 2020;13:171-183 [FREE Full text] [doi: 10.2147/LRA.S240564] [Medline: 33149664]
4. Freudenberger HJ. Staff burn-out. J Social Issues 1974;30:159-165. [doi: 10.1111/j.1540-4560.1974.tb00706.x]
5. Burn-out an "occupational phenomenon": international classification of diseases. World Health Organization. 2019. URL: https://www.who.int/news/item/28-05-2019-burn-out-an-occupational-phenomenon-international-classification-of-diseases [accessed 2022-06-27]
6. Perniciotti P, Serrano CV, Guarita RV, Morales RJ, Romano BW. Síndrome de burnout nos profissionais de saúde: atualização sobre definições, fatores de risco e estratégias de prevenção. Rev SBPH 2020;23(1):35-52 [FREE Full text]
7. de Beer LT, Schaufeli WB, De Witte H, Hakanen JJ, Shimazu A, Glaser J, et al. Measurement invariance of the burnout assessment tool (BAT) across seven cross-national representative samples. Int J Environ Res Public Health 2020;17(15):5604 [FREE Full text] [doi: 10.3390/ijerph17155604] [Medline: 32756483]
8. Baugh JJ, Takayesu JK, White BA, Raja AS. Beyond the Maslach burnout inventory: addressing emergency medicine burnout with Maslach's full theory. J Am Coll Emerg Physicians Open 2020;1(5):1044-1049 [FREE Full text] [doi: 10.1002/emp2.12101] [Medline: 33145555]

9. Bui T. Relação esforço-recompensa, comprometimento excessivo e esgotamento: um estudo transversal entre profissionais de saúde Vietnamitas. Psicol Conv 2022;9:1.
10. Vieira I, Russo JA. Burnout and stress: between medicalization and psychologization. Physis 2019;29(2):e290206. [doi: 10.1590/S0103-73312019290206]
11. Petrelli F, Scuri S, Tanzi E, Nguyen C, Grappasonni I. Public health and burnout: a survey on lifestyle changes among workers in the healthcare sector. Acta Biomed 2019;90(1):24-30 [FREE Full text] [doi: 10.23750/abm.v90i1.7626] [Medline: 30889151]
12. Maslach C, Jackson SE. The measurement of experienced burnout. J. Organiz. Behav 1981;2(2):99-113. [doi: 10.1002/job.4030020205]
13. Campos ICM, Pereira SS, Schiavon ICA, Alves M. Maslach burnout inventory-human services survey (MBI-HSS): revisão integrativa de sua utilização em pesquisas Brasileiras. Arq Ciências Saúde UNIPAR 2020;24(3):187-195. [doi: 10.25110/arqsaude.v24i3.2020.7875]
14. Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11. Scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, South Australia: Joanna Briggs Institute; 2020:407-452.
15. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
16. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
17. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
18. Soares JP, Lopes RH, de Souza Mendonça PB, Silva CRDV, Rodrigues CCFM, de Castro JL. Use of the Maslach burnout inventory among public health care professionals: protocol for a scoping review. JMIR Res Protoc 2022;11(11):e42338 [FREE Full text] [doi: 10.2196/42338] [Medline: 36318252]
19. Araújo WCO. Recuperação da informação em saúde: construção, modelos e estratégias. ConCI 2020;3(2):100-134. [doi: 10.33467/conci.v3i2.13447]
20. Camargo BV, Justo AM. IRAMUTEQ: um software gratuito para análise de dados textuais. Temas Psicol 2013;21(2):513-518. [doi: 10.9788/tp2013.2-16]
21. Grau A, Flichtentrei D, Suñer R, Prats M, Braga F. Influencia de factores personales, profesionales y transnacionales en el síndrome de burnout en personal sanitario hispanoamericano y español (2007). Rev Esp Salud Publica 2009;83(2):215-230 [FREE Full text] [doi: 10.1590/s1135-57272009000200006] [Medline: 19626249]
22. Beltrán CA, Baltazar RG, Moreno MP, Santacruz GH. Factores de riesgo psicosocial laborales, apoyo social y síndrome de burnout en médicos de familia de tres instituciones públicas de salud de Guadalajara (México). Rev Salud Uninorte 2013;29(3):487-500 [FREE Full text]
23. Bijari B, Abassi A. Prevalence of burnout syndrome and associated factors among rural health workers (behvarzes) in South Khorasan. Iran Red Crescent Med J 2016;18(10):e25390 [FREE Full text] [doi: 10.5812/ircmj.25390] [Medline: 28180014]
24. Bressi C, Porcellana M, Gambini O, Madia L, Muffatti R, Peirone A, et al. Burnout among psychiatrists in Milan: a multicenter survey. Psychiatr Serv 2009;60(7):985-988 [FREE Full text] [doi: 10.1176/ps.2009.60.7.985] [Medline: 19564233]
25. Hayter M. Burnout and AIDS care-related factors in HIV community clinical nurse specialists in the North of England. J Adv Nurs 1999;29(4):984-993. [doi: 10.1046/j.1365-2648.1999.00973.x] [Medline: 10215992]
26. Lasebikan VO, Oyetunde MO. Burnout among nurses in a Nigerian general hospital: prevalence and associated factors. ISRN Nurs 2012;2012:402157 [FREE Full text] [doi: 10.5402/2012/402157] [Medline: 22619733]
27. Lee FJ, Stewart M, Brown JB. Stress, burnout, and strategies for reducing them: what's the situation among Canadian family physicians? Can Fam Physician 2008;54(2):234-235 [FREE Full text] [Medline: 18272641]
28. Lesić ARA, Stefanovic NP, Perunicić I, Milenković P, Tosevski DL, Bumbasirević MZ. Burnout in Belgrade orthopaedic surgeons and general practitioners: a preliminary report. Acta Chir Iugosl 2009;56(2):53-59. [doi: 10.2298/aci0902053l] [Medline: 19780331]
29. Lorenz VR, Benatti MCC, Sabino MO. Burnout e estresse em enfermeiros de um hospital universitário de alta complexidade. Rev Lat-Am Enferm 2010;18(6):1084-1091. [doi: 10.1590/S0104-11692010000600007]
30. Okwaraji FE, Aguwa EN. Burnout and psychological distress among nurses in a Nigerian tertiary health institution. Afr Health Sci 2014;14(1):237-245 [FREE Full text] [doi: 10.4314/ahs.v14i1.37] [Medline: 26060486]
31. Paiva CE, Martins BP, Paiva BSR. Doctor, are you healthy? A cross-sectional investigation of oncologist burnout, depression, and anxiety and an investigation of their associated factors. BMC Cancer 2018;18(1):1044 [FREE Full text] [doi: 10.1186/s12885-018-4964-7] [Medline: 30367614]
32. Rachiotis G, Kourousis C, Kamilaraki M, Symvoulakis EK, Dounias G, Hadjichristodoulou C. Medical supplies shortages and burnout among Greek health care workers during economic crisis: a pilot study. Int J Med Sci 2014;11(5):442-447 [FREE Full text] [doi: 10.7150/ijms.7933] [Medline: 24688306]

33. Renzi C, Tabolli S, Ianni A, Di Pietro C, Puddu P. Burnout and job satisfaction comparing healthcare staff of a dermatological hospital and a general hospital. J Eur Acad Dermatol Venereol 2005;19(2):153-157. [doi: 10.1111/j.1468-3083.2005.01029.x] [Medline: 15752281]
34. Rionda IS, Cortés-García L, de la Villa Moral Jiménez M. The role of burnout in the association between work-related factors and perceived errors in clinical practice among Spanish residents. Int J Environ Res Public Health 2021;18(9):4931 [FREE Full text] [doi: 10.3390/ijerph18094931] [Medline: 34066327]
35. de Lima Trindade L, Lautert L. [Syndrome of burnout among the workers of the strategy of health of the family]. Rev Esc Enferm USP 2010;44(2):274-279 [FREE Full text] [doi: 10.1590/s0080-62342010000200005] [Medline: 20642035]
36. Zarei E, Ahmadi F, Sial MS, Hwang J, Thu PA, Usman SM. Prevalence of Burnout among Primary Health Care Staff and Its Predictors: A Study in Iran. Int J Environ Res Public Health 2019 Jun 25;16(12):2249 [FREE Full text] [doi: 10.3390/ijerph16122249] [Medline: 31242691]
37. Zhou Y, Gu W, Tao R, Chen C. The impact of gender and working hours on pulmonary physician burnout. Ann Transl Med 2020;8(18):1166 [FREE Full text] [doi: 10.21037/atm-20-5827] [Medline: 33241015]
38. Mota CM, Dosea GS, Nunes PS. [Assessment of the prevalence of burnout syndrome in community health agents of the city of Aracaju in the state of Sergipe, Brazil]. Cien Saude Colet 2014;19(12):4719-4726 [FREE Full text] [doi: 10.1590/1413-812320141912.02512013] [Medline: 25388180]
39. Zha N, Patlas MN, Neuheimer N, Duszak R. Prevalence of burnout among Canadian radiologists and radiology trainees. Can Assoc Radiol J 2018;69(4):367-372 [FREE Full text] [doi: 10.1016/j.carj.2018.05.005] [Medline: 30270152]
40. Khalafallah AM, Lam S, Gami A, Dornbos DL, Sivakumar W, Johnson JN, et al. Burnout and career satisfaction among attending neurosurgeons during the COVID-19 pandemic. Clin Neurol Neurosurg 2020;198:106193 [FREE Full text] [doi: 10.1016/j.clineuro.2020.106193] [Medline: 32942135]
41. Li H, Zuo M, Gelb AW, Zhang B, Zhao X, Yao D, et al. Chinese anesthesiologists have high burnout and low job satisfaction: a cross-sectional survey. Anesth Analg 2018;126(3):1004-1012 [FREE Full text] [doi: 10.1213/ANE.0000000000002776] [Medline: 29309320]
42. Li H, Yuan B, Meng Q, Kawachi I. Contextual factors associated with burnout among Chinese primary care providers: a multilevel analysis. Int J Environ Res Public Health 2019;16(19):3555 [FREE Full text] [doi: 10.3390/ijerph16193555] [Medline: 31547533]
43. Wu H, Liu L, Wang Y, Gao F, Zhao X, Wang L. Factors associated with burnout among Chinese hospital doctors: a cross-sectional study. BMC Public Health 2013;13:786 [FREE Full text] [doi: 10.1186/1471-2458-13-786] [Medline: 23985038]
44. Lu S, Zhang L, Klazinga N, Kringos D. More public health service providers are experiencing job burnout than clinical care providers in primary care facilities in China. Hum Resour Health 2020;18(1):95 [FREE Full text] [doi: 10.1186/s12960-020-00538-z] [Medline: 33272284]
45. De la Fuente-Solana EI, Pradas-Hernández L, González-Fernández CT, Velando-Soriano A, Martos-Cabrera MB, Gómez-Urquiza JL, et al. Burnout syndrome in paediatric nurses: a multi-centre study. Int J Environ Res Public Health 2021;18(3):1324 [FREE Full text] [doi: 10.3390/ijerph18031324] [Medline: 33535707]
46. González-Rodríguez R, López-Castedo A, Pastor-Seller E, Verde-Diego C. Burnout syndrome in the health system: the case of social health workers. Enferm Glob 2020;19(2):141-161. [doi: 10.6018/eglobal.382631]
47. Ogdon Lebrón MA, Díaz-Reissner CV. Burnout syndrome in dentists working at the XVIII sanitary region of the ministry of public health and social, Paraguay. Rev Salud Publica Parag 2017;7(1):21-27. [doi: 10.18004/rspp.2017.junio.21-27]
48. de la Cruz SP, Abellán MV. Professional burnout, stress and job satisfaction of nursing staff at a university hospital. Rev Lat Am Enfermagem 2015;23(3):543-552 [FREE Full text] [doi: 10.1590/0104-1169.0284.2586] [Medline: 26155012]
49. Grau A, Suñer R, García MM. [Burnout syndrome in health workers and relationship with personal and environmental factors]. Gac Sanit 2005;19(6):463-470 [FREE Full text] [doi: 10.1016/s0213-9111(05)71397-2] [Medline: 16483524]
50. Cañadas-de la Fuente GA, Albendín-García L, Cañadas GR, San Luis-Costas C, Ortega-Campos E, de la Fuente-Solana EI. Nurse burnout in critical care units and emergency departments: intensity and associated factors. Emergencias 2018;30(5):328-331 [FREE Full text] [Medline: 30260117]
51. Al-Dubai SAR, Rampal KG. Prevalence and associated factors of burnout among doctors in Yemen. J Occup Health 2010;52(1):58-65 [FREE Full text] [doi: 10.1539/joh.o8030] [Medline: 19907108]
52. da Silva RADA, Teixeira AKM, Frota MMA, Maciel JAC, de Farias MR. Job satisfaction and burnout among oral healthcare providers within the unified health system in sobral, Ceará, Brazil. Rev Bras Med Trab 2019;17(3):313-324 [FREE Full text] [doi: 10.5327/Z1679443520190352] [Medline: 32368665]
53. Beltrán CA, Salas JHB, Paredes JJR, Zárate LAQ, Cabrales APC, Ortega GG, et al. Factores psicosociales y síndrome de burnout en médicos generales de Tepic, Nayarit (México). Rev Salud Uninorte 2015;31(2):245-254 [FREE Full text] [doi: 10.14482/sun.31.2.6018]
54. Bojčeski O, Galjak M, Kulić L, Đurić S, Mirković M, Milošević J. Testing the correlation between occupational stress and occupational burnout among nurses working in public health institutions. Praxis Med 2019;48(3-4):13-18. [doi: 10.5937/pramed1904013b]

55. Maciel JAC, de Farias MR, Sampaio JJC, Pérez Guerrero JA, Castro-Silva II. Satisfacción profesional y prevalencia del síndrome de burnout en equipos de salud bucal de atención primaria en el Municipio Sobral, Ceará-Brasil. Salud Trab 2018;26(1):34-44.
56. Engelbrecht MC, Bester CL, Van Den Berg H, Van Rensburg HCJ. A study of predictors and levels of burnout: the case of professional nurses in primary health care facilities in the free state. S Afr J Econ 2008;76(Suppl1):S15-S27. [doi: 10.1111/j.1813-6982.2008.00164.x]
57. Gan Y, Jiang H, Li L, Yang Y, Wang C, Liu J, et al. Prevalence of burnout and associated factors among general practitioners in Hubei, China: a cross-sectional study. BMC Public Health 2019;19(1):1607 [FREE Full text] [doi: 10.1186/s12889-019-7755-4] [Medline: 31791282]
58. Khan N, Palepu A, Dodek P, Salmon A, Leitch H, Ruzycki S, et al. Cross-sectional survey on physician burnout during the COVID-19 pandemic in Vancouver, Canada: the role of gender, ethnicity and sexual orientation. BMJ Open 2021;11(5):e050380 [FREE Full text] [doi: 10.1136/bmjopen-2021-050380] [Medline: 33972345]
59. Tomljenovic M, Kolaric B, Stajduhar D, Tesic V. Stress, depression and burnout among hospital physicians in Rijeka, Croatia. Psychiatr Danub 2014;26(Suppl 3):450-458. [Medline: 25536981]
60. Alotni MA, Elgazzar SE. Investigation of burnout, its associated factors and its effect on the quality of life of critical care nurses working in buraydah central hospital at Qassim Region, Saudi Arabia. Open Nurs J 2020;14:190-202 [FREE Full text] [doi: 10.2174/1874434602014010190]
61. Mackanga JR, Mouloungui EGM, Iba-Ba J, Pottier P, Kombila JBM, Boguikouma JB. Burnout level and associated factors in a sub-Saharan African medical setting: prospective cross-sectional survey. BMC Med Educ 2020;20(1):295 [FREE Full text] [doi: 10.1186/s12909-020-02194-2] [Medline: 32912195]
62. Amiri M, Khosravi A, Eghtesadi AR, Sadeghi Z, Abedi G, Ranjbar M, et al. Burnout and its influencing factors among primary health care providers in the North East of Iran. PLoS One 2016;11(12):e0167648 [FREE Full text] [doi: 10.1371/journal.pone.0167648] [Medline: 27930726]
63. da Silva ATC, de Souza Lopes C, Susser E, Coutinho LMS, Germani ACCG, Menezes PR. Burnout among primary health care workers in Brazil: results of a multilevel analysis. Int Arch Occup Environ Health 2021;94(8):1863-1875. [doi: 10.1007/s00420-021-01709-8] [Medline: 34057590]
64. Lahana E, Papadopoulou K, Roumeliotou O, Tsounis A, Sarafis P, Niakas D. Burnout among nurses working in social welfare centers for the disabled. BMC Nurs 2017;16:15 [FREE Full text] [doi: 10.1186/s12912-017-0209-3] [Medline: 28344515]
65. Marcelino G, Cerveira JM, Carvalho I, Costa JA, Lopes M, Calado NE, et al. Burnout levels among Portuguese family doctors: a nationwide survey. BMJ Open 2012;2(3):e001050 [FREE Full text] [doi: 10.1136/bmjopen-2012-001050] [Medline: 22710131]
66. Nishimura Y, Miyoshi T, Sato A, Hasegawa K, Hagiya H, Kosaki Y, et al. Burnout of healthcare workers amid the COVID-19 pandemic: a follow-up study. Int J Environ Res Public Health 2021;18(21):11581 [FREE Full text] [doi: 10.3390/ijerph182111581] [Medline: 34770095]
67. Osman D, Abdlrheem S. Burnout and job satisfaction among healthcare providers in Aswan university hospital, Upper Egypt. J High Inst Public Health 2019;49(1):64-72. [doi: 10.21608/jhiph.2019.29468]
68. Pantenburg B, Luppa M, König HH, Riedel-Heller SG. Burnout among young physicians and its association with physicians' wishes to leave: results of a survey in Saxony, Germany. J Occup Med Toxicol 2016;11:2 [FREE Full text] [doi: 10.1186/s12995-016-0091-z] [Medline: 26807138]
69. Rajan S, Engelbrecht A. A cross-sectional survey of burnout amongst doctors in a cohort of public sector emergency centres in Gauteng, South Africa. Afr J Emerg Med 2018;8(3):95-99 [FREE Full text] [doi: 10.1016/j.afjem.2018.04.001] [Medline: 30456156]
70. Kosan Z, Calikoglu EO, Guraksin A. Levels of burnout and their associated factors among physicians working in Northeast Anatolia. Niger J Clin Pract 2018;21(7):875-881 [FREE Full text] [Medline: 29984719]
71. Barroso SM, da Rocha Pereira Guerra A. Burnout and quality of life in health comunitary agents of Caetanópolis (MG). Cad Saúde Colet 2013;21(3):338-345. [doi: 10.1590/s1414-462x2013000300016]
72. do Nascimento Ferreira N, de Lucca SR. Síndrome de burnout em técnicos de enfermagem de um hospital público do estado de São Paulo. Rev Bras Epidemiol 2015;18(1):68-79 [FREE Full text] [doi: 10.1590/1980-5497201500010006] [Medline: 25651012]
73. de Vasconcelos EM, De Martino MMF. Preditores da síndrome de burnout em enfermeiros de unidade de terapia intensiva. Rev Gaucha Enferm 2018;38(4):e65354 [FREE Full text] [doi: 10.1590/1983-1447.2017.04.65354] [Medline: 29933422]
74. Koval KW, Lindquist B, Gennosa C, Mahadevan A, Niknam K, Patil S, et al. First look at emergency medical technician wellness in India: application of the Maslach burnout inventory in an unstudied population. PLoS One 2020;15(3):e0229954 [FREE Full text] [doi: 10.1371/journal.pone.0229954] [Medline: 32155192]
75. Marchalik D, Brems J, Rodriguez A, Lynch JH, Padmore J, Stamatakis L, et al. The impact of institutional factors on physician burnout: a national study of urology trainees. Urology 2019;131:27-35. [doi: 10.1016/j.urology.2019.04.042] [Medline: 31199967]

76. Goméz CM, de Vasconcellos LCF, Machado JMH. Saúde do trabalhador: aspectos históricos, avanços e desafios no sistema único de saúde. Ciên Saúde Colet 2018;23(6):1963-1970 [FREE Full text] [doi: 10.1590/1413-81232018236.04922018] [Medline: 29972503]
77. Da Cunha FM, Mendes A. Revisão narrativa sobre a questão da saúde do trabalhador no sistema único de saúde. J Manag Prim Health Care 2021;13:e018. [doi: 10.14295/jmphc.v13.1164]
78. de Souza Julio R, Lourenção LG, de Oliveira SM, Farias DHR, Gazetta CE. Prevalence of anxiety and depression in Brazilian primary health care workers. Cad Bras Ter Ocup 2022;30:e2997 [FREE Full text] [doi: 10.1590/2526-8910.ctoao22712997]
79. de Moura RCD, Chavaglia SRR, Coimbra MAR, Araújo APA, Scárdua SA, Ferreira LA, et al. Common mental disorders in emergency services nursing professionals. Acta Paul Enferm 2022;35:eAPE03032. [doi: 10.37689/acta-ape/2022AO03032]
80. de Campos JJB, Forster AC, Freire Filho JR. Lessons learned from the comparison of Brazilian and Spanish health systems. Espaç Saúde 2016;17(1):121-129.
81. Langbecker A, Castellanos MEP, Catalán-Matamoros D. When public health systems are news: a comparative analysis of the journalistic coverage in Brazil and Spain. Cien Saude Colet 2020;25(11):4281-4292 [FREE Full text] [doi: 10.1590/1413-812320202511.22532018] [Medline: 33175037]
82. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med 2018;283(6):516-529 [FREE Full text] [doi: 10.1111/joim.12752] [Medline: 29505159]
83. Medscape. 2019. URL: https://www.medscape.com/slideshow/2019-lifestyle-burnout-depression-6011056 [accessed 2022-06-18]
84. Yates SW. Physician stress and burnout. Am J Med 2020;133(2):160-164 [FREE Full text] [doi: 10.1016/j.amjmed.2019.08.034] [Medline: 31520624]
85. Rodrigues H, Cobucci R, Oliveira A, Cabral JV, Medeiros L, Gurgel K, et al. Burnout syndrome among medical residents: a systematic review and meta-analysis. PLoS One 2018;13(11):e0206840 [FREE Full text] [doi: 10.1371/journal.pone.0206840] [Medline: 30418984]
86. Moukarzel A, Michelet P, Durand AC, Sebbane M, Bourgeois S, Markarian T, et al. Burnout syndrome among emergency department staff: prevalence and associated factors. Biomed Res Int 2019;2019:6462472 [FREE Full text] [doi: 10.1155/2019/6462472] [Medline: 30800675]
87. Zhou AY, Hann M, Panagioti M, Patel M, Agius R, Van Tongeren M, et al. Cross-sectional study exploring the association between stressors and burnout in junior doctors during the COVID-19 pandemic in the United Kingdom. J Occup Health 2022;64(1):e12311 [FREE Full text] [doi: 10.1002/1348-9585.12311] [Medline: 35025106]
88. Sibeoni J, Bellon-Champel L, Mousty A, Manolios E, Verneuil L, Revah-Levy A. Physicians' perspectives about burnout: a systematic review and metasynthesis. J Gen Intern Med 2019;34(8):1578-1590 [FREE Full text] [doi: 10.1007/s11606-019-05062-y] [Medline: 31147982]
89. Silva KSS, Fogaça JA, Silva SO, Aoyama EA, Lemos LR. Burnout syndrome in nursing professional. ReBIS 2020;2(1):38-42.
90. Rouquayrol M. Epidemiologia & Saúde. 7ª edition. Rio de Janeiro: Medbook; 2013.
91. Medronho RA, Carvalho DM, Bloch KV, Werneck GL, Luiz RR, organizadores. Epidemiologia. São Paulo: Atheneu; 2009.
92. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, et al. Prevalence of burnout among physicians: a systematic review. JAMA 2018;320(11):1131-1150 [FREE Full text] [doi: 10.1001/jama.2018.12777] [Medline: 30326495]
93. Bartholomew AJ, Houk AK, Pulcrano M, Shara NM, Kwagyan J, Jackson PG, et al. Meta-analysis of surgeon burnout syndrome and specialty differences. J Surg Educ 2018;75(5):1256-1263 [FREE Full text] [doi: 10.1016/j.jsurg.2018.02.003] [Medline: 29500145]
94. Grunfeld E, Whelan TJ, Zitzelsberger L, Willan AR, Montesanto B, Evans WK. Cancer care workers in Ontario: prevalence of burnout, job stress and job satisfaction. CMAJ 2000;163(2):166-169 [FREE Full text] [Medline: 10934978]
95. Kalliath TJ, O'Driscoll MP, Gillespie DF, Bluedorn AC. A test of the maslach burnout inventory in three samples of healthcare professionals. Work & Stress 2000;14(1):35-50. [doi: 10.1080/026783700417212]
96. Güler Y, Şengül S, Çaliş H, Karabulut Z. Burnout syndrome should not be underestimated. Rev Assoc Med Bras (1992) 2019;65(11):1356-1360 [FREE Full text] [doi: 10.1590/1806-9282.65.11.1356] [Medline: 31800896]
97. Brito TB, Sousa MDSDC, Rodrigues TS. Síndrome de burnout: estratégias de prevenção e tratamento nos profissionais de enfermagem. Rev Uningá 2019;56(S2):113-122 [FREE Full text] [doi: 10.46311/2318-0579.56.euj2383]
98. Lima SDSFD, Dolabela MF. Estratégias usadas para a prevenção e tratamento da síndrome de burnout. Res Soc Dev 2021;10(5):e11110514500 [FREE Full text] [doi: 10.33448/rsd-v10i5.14500]
99. Matsuzaki PGI, Mariya FA, Ueno LI, Gimenes MJF. Physician burnout: prevention strategies. Rev Bras Med Trab 2021;19(4):511-517 [FREE Full text] [doi: 10.47626/1679-4435-2021-713] [Medline: 35733541]

Abbreviations

BS: burnout syndrome
ICD: International Classification of Diseases
JBI: Joanna Briggs Institute

MBI: Maslach Burnout Inventory
MBI-ES: MBI-Educators Survey
MBI-GS: MBI-General Survey
MBI-GS (S): MBI-General Survey for Students
MBI-HSS: MBI-Human Services Survey
MBI-HSS-(MP): MBI-Human Services Survey for Medical Personnel
PCC: Population, Concept, Context
PHC: primary health care
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
SUS: Unified Health System

Edited by J Torous, G Eysenbach; submitted 09.11.22; peer-reviewed by D Madhusudhan, N Vu, S Pesälä; comments to author 10.12.22; revised version received 21.12.22; accepted 31.12.22; published 21.07.23.

Please cite as:
Soares JP, Lopes RH, Mendonça PBDS, Silva CRDV, Rodrigues CCFM, Castro JLD
Use of the Maslach Burnout Inventory Among Public Health Care Professionals: Scoping Review
JMIR Ment Health 2023;10:e44195
URL: https://mental.jmir.org/2023/1/e44195
doi:10.2196/44195
PMID:37477960

Review

The Involvement of Service Users and People With Lived Experience in Mental Health Care Innovation Through Design: Systematic Review

Lars Veldmeijer[1,2], MA; Gijs Terlouw[2], PhD; Jim Van Os[1], PhD; Olga Van Dijk[3], BA; Job Van 't Veer[2], PhD; Nynke Boonstra[1,2,4], PhD

[1]Department of Psychiatry, Utrecht University Medical Center, Utrecht, Netherlands

[2]Department of Healthcare and Welfare, NHL Stenden University of Applied Sciences, Leeuwarden, Netherlands

[3]NHL Stenden Library, NHL Stenden University of Applied Sciences, Leeuwarden, Netherlands

[4]KieN VIP Mental Health Care Services, Leeuwarden, Netherlands

Corresponding Author:
Lars Veldmeijer, MA
Department of Psychiatry
Utrecht University Medical Center
Heidelberglaan 100
Utrecht, 3584 CX
Netherlands
Phone: 31 6 82874768
Email: lars.veldmeijer@nhlstenden.com

Abstract

Background: Mental health care faces challenges that not only necessitate innovation but also require the involvement of service users and people with lived experience in developing and evaluating mental health care services. As the development of digital interventions is becoming more prevalent, design approaches are increasingly finding their way into mental health. There is evidence that these approaches can successfully integrate user experience into mental health services. However, there is no clear overview of the studies conducted and the lessons learned concerning the involvement of service users and people with lived experience.

Objective: In this systematic review, we aimed to provide an overview of the involvement of service users and people with lived experience in mental health care services through design approaches and to synthesize the advantages of design approaches in mental health care.

Methods: The following 5 databases were searched for relevant abstracts: PsycINFO, PubMed, Web of Science, Scopus, and Embase. In addition, 2 health design journal archives, *Design for Health* and *The Journal of Health Design*, were searched. To categorize the results, we collected the reported added value from the included articles and conducted a thematic synthesis in which the themes were developed from the retrieved data. The themes were discussed, revised, and checked until saturation was achieved.

Results: We included and categorized 33 papers. Most studies involved service users, primarily adults, and used various design approaches. Most of these studies aimed to design or evaluate digital interventions. Service users and people with lived experience were involved in different roles but never as decision makers. Studies that used co-design approaches exhibited the highest levels of involvement. Various added values were reported, including tailoring and testing interventions and digital interventions, improving engagement and collaboration, gathering the needs of stakeholders, and empowering participants as resourceful actors. The challenges reported were maintaining participants' continued participation throughout the study, managing the iterative nature of design, providing a safe space, balancing insights from design and medical science, and navigating design processes in medical environments.

Conclusions: This systematic review provides an overview of the studies that used design approaches to involve service users and people with lived experience in mental health care innovation. Design approaches have advantages in mental health care innovation, offering added value and having manageable challenges. Future studies using design approaches in mental health care should involve participants as partners and decision makers and report on collaboration in a systematic and clear manner.

(JMIR Ment Health 2023;10:e46590) doi:10.2196/46590

KEYWORDS

design approaches; innovation; psychiatry; mental health care; involvement; service users; people with lived experience; cocreation; mobile phone

Introduction

Background

Mental health care services are in need of transformation [1]. However, incremental and iterative integration of advances is recommended rather than a complete paradigm shift [2]. Creative thinking [3] and the involvement of people with lived experience [4] are needed to catalyze these advances. Design and design research hold the potential to drive these incremental and iterative changes in traditional mental health care services, as it builds on creative thinking and doing [5]. Over the last decade, the application of design approaches in mental health care services has increased and has shown to accelerate innovation [6-8]. For example, design or co-design has been effective in designing or redesigning psychiatric facilities [9-11] and has proven successful in engaging end users in the design process of novel therapies [12-14]. The integration of design into mental health services is a logical continuation of the existing solutions to the challenges of the sector. These challenges include the relationship between professionals and patients in the era of value-based competition in health care [15,16], paradigm shifts in the diagnosis of mental health conditions [2,17,18], the integration of empirical evidence of recovery into traditional services [19,20], the efficacy of psychotherapies and pharmacotherapies in adults [21], and the cocreation of novel concepts and language with people with lived experience [22].

Design Approaches

Design can contribute to change and innovation in mental health care because it is a transformative discipline [23]. A scoping review of the extensive evidence on the role of the arts in improving health and well-being concluded that design, as a subcategory of the arts, can help prevent the onset of mental health conditions and support their treatment [24]. Design differs from the rational way of solving problems, that is, designers do not aim for the optimal solution to a given problem but choose a more exploratory approach, where the problem and solution evolve together [25]. Although digital innovation in mental health care has shown promising results, such as ecological momentary assessment and passive sensor tracking [26], the field encounters challenges involving service users and people with lived experience in the design of digital health interventions [27]. Despite these challenges, the involvement of service users and people with lived experience in design is seen as essential to tailor innovations to their needs and balance the power between patients and professionals [4]. Recent developments in design methodologies offer potential approaches to address these issues. For example, in recent years, the scope of design in health has been expanded by introducing various user-centered approaches, such as co-design [28], experience-based co-design [29], participatory design [14], design thinking [8], design research [30], iterative design [13], value-sensitive design [31], experience design [32], and human-centered design [33]. In these design approaches, designers aim to integrate users' tacit knowledge into the design of products and services by assigning them an important role in the design process [34]. These design approaches use various methods from other fields [35], such as natural sciences, sociology, psychology, anthropology, and visual arts [36]. Design approaches have provided opportunities to all the stakeholders involved in the innovation processes to discover the user experience and to place the end users at the center of the design process by involving them as stakeholders [37].

Level of Involvement of Service Users and People With Lived Experience

There is a continuing focus on how service users and people with lived experience can be involved in mental health projects. A recently developed framework, The Involvement Matrix [38], was cocreated with experts by experience and researchers. The Involvement Matrix describes 5 roles (ie, listener, cothinker, adviser, partner, and decision maker) and 3 phases. Mainly, these roles are relevant to assessing the level of involvement of people with lived experience and service users in studies using a design approach. The lowest level of involvement is "the listener," as they only receive information, whereas the second level, "the cothinker," is also asked for an opinion. The third level, "the adviser," gives solicited or unsolicited advice, whereas on the fourth level, "the partner," acts as an equal partner. The "decision maker" can be seen as having the highest level of involvement, as the decision maker takes the initiative and is involved in the (final) decision. Although design approaches are useful in engaging stakeholders in mental health care innovation projects, there is currently no comprehensive overview of the studies that used design approaches, including which specific approaches were used, the roles service users and individuals with lived experience had, and the added value they generated.

Objective

In this systematic review, we aimed to provide an overview of the involvement of service users and people with lived experience in mental health care services through design approaches and to synthesize the advantages of design approaches in mental health care.

Methods

Databases and Search Strategy

The following 5 databases were searched for relevant abstracts: PsycINFO, PubMed, Web of Science, Scopus, and Embase. These databases cover a wide range of published studies in the field of health and design. *Design for Health* and *The Journal of Health Design*, which are both health design journals, were also searched because they publish papers at the intersection of

health and design. The terms used for the PubMed search are presented in Textbox 1.

Owing to the differences in search engine functionality, the method by which the terms were entered differed per database. A complete overview of these terms is provided in Multimedia Appendix 1. Before conducting the definitive search, we contacted an information specialist and performed 4 preliminary searches by using different terms to minimize the possibility of missing relevant studies. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [39] to report this review. The search was conducted between September 6, 2022, and October 28, 2022.

Textbox 1. Terms used for searching PubMed.

A combination of search terms were used to identify relevant papers under the following categories: ((“Mental Health”[Mesh] OR mental-health[tiab] OR mental-hygiene[tiab] OR mental-care[tiab] OR “Psychiatry”[Mesh] OR psychiatr*[tiab]) AND (user-centered development*[tiab] OR user-centred design*[tiab] OR user-centred development*[tiab] OR user-centric design*[tiab] OR user-driven design*[tiab] OR user-driven development*[tiab] OR “User-Centered Design”[Mesh] OR user-centered-design*[tiab] OR usability-testing[tiab] OR Co-design*[tiab] OR Participatory-design*[tiab] OR Experience-based-co-design*[tiab] OR Interaction-design*[tiab] OR Service-design*[tiab] OR Systemic-design*[tiab] OR Patient-centered-design*[tiab] OR Human-centered-design*[tiab] OR Value-sensitive-design*[tiab] OR Design-thinking[tiab] OR Design-research*[tiab] OR Design-method*[tiab] OR Design-session*[tiab] OR design-approach*[tiab] OR design-principle*[tiab] OR design-choice*[tiab] OR universal-design*[tiab] OR Creativ-method*[tiab] OR Creative-session*[tiab] OR Generative-design*[tiab] OR Generative-method*[tiab] OR Generative-session*[tiab] OR Iterative-design*[tiab] OR Design-driven-innovation*[tiab] OR Speculative-design*[tiab] OR Critical-design*[tiab] OR Discursive-design*[tiab] OR Product-design*[tiab])) NOT (“Research Design”[Mesh] OR research-design[tiab] OR research-designs[tiab] OR research-protocol*[tiab] OR research-instrument*[tiab] OR study-design[tiab] OR study-designs[tiab] OR research-method*[tiab] OR methodology[tiab] OR methodological-research[tiab])

Study Selection and Inclusion and Exclusion Criteria

We included studies that discussed the involvement of service users and people with lived experience in mental health care services through design approaches. We included only original reports or papers that (1) involved service users, people with lived experience, or both; (2) mentioned design approaches; (3) involved an empirical study; and (4) conducted the study in settings including mental health care service or psychiatry programs. Papers that met these criteria were selected for full-text screening. We defined service users as participants who used mental health care services at the time of their involvement and people with lived experience as those who had used mental health care services in the past but were not currently using them during the study in which they were involved. We included empirical studies because they could provide insights into the level of involvement. We selected studies conducted in mental health care and psychiatry care service settings to retrieve as many studies as possible. We are aware that there is an overlap between the terms and that the services offered under both contexts may vary across countries.

The following exclusion criteria were used for full-text screening: (1) non–peer-reviewed papers such as abstracts, conference posters, or trade journals; (2) papers with full texts not available; (3) papers in languages other than English; (4) monographs or short reports; and (5) papers with insufficient information in the abstract.

Research Questions

To provide an overview of the involvement of service users and people with lived experience in mental health care services through design approaches and to provide insight into the advantages of design approaches in mental health care innovation, the included studies were analyzed using the following research questions (RQs):

- RQ 1: Who were included (service users, people with lived experience, or both)?
- RQ 2: What design approach was used?
- RQ 3: What was the aim of the innovation?
- RQ 4: What were the roles of service users and people with lived experience?
- RQ 5: What was the added value of involving service users and people with lived experience through design approaches?
- RQ 6: What were the challenges in involving service users and people with lived experience through design approaches?

Screening Process and Study Selection

After removing duplicates, the titles and abstracts of the papers were screened using Rayyan (**Rayyan** Systems Inc) [40]. A total of 2 reviewers (LV and GT) independently reviewed all the titles and abstracts and were double blinded for relevance with the formulated inclusion and exclusion criteria. Papers were only included if both LV and GT agreed, and a plausible argumentation for consideration of inclusion always led to the inclusion. Full-text papers were retrieved after this step. During the full-text screening phase, the first 20% of the papers were randomly selected and double-blind reviewed by 2 reviewers (LV and GT). The random selection ensured that the screening process was unbiased, and the double-blind review increased its reliability. Subsequently, the primary reviewer (LV) reviewed the other included papers for full-text reading, which helped to maintain consistency throughout the review process.

Data Extraction

Data were extracted using structured forms, including the characteristics of participants (service users and people with lived experience), aim of innovation, design approach, and role of the participants in the design approach. In addition, we focused on the added value and challenges of involving service users and people with lived experience through a design approach that the authors reported in the included studies. Findings from all the studies regarding the reported added values and challenges were extracted and collated using a thematic synthesis. Thematic synthesis preserves principles that have traditionally been important to systematic reviewing [41]. This

synthesis was performed by LV and GT. Because our primary aim was to find all possible added values reported (ie, a comprehensive overview), we refrained from using the existing design frameworks to guide data extraction and analysis, as such an approach may have impeded the identification of novel findings that were not aligned with the themes these frameworks provide or other a priori themes that we could have developed [42]. As such, the themes that emerged from the data were discussed and revised to minimize overlap and were checked by researchers JvV and NB. This process was repeated until saturation of the themes was achieved.

Results

Search Results

Our initial database search yielded 2758 records. After removing 745 (27.01%) duplicates from 2758 records, the titles and abstracts of the 2013 (72.99%) records were screened. Next, excluding 1971 (97.91%) papers from the 2013 records after full-text screening, 42 (2.13%) records were sought for retrieval, and finally, 28 (67%) papers were included. We also identified 71 records through a journal search. After removing 17 (24%) duplicates from 71 records, we retrieved 54 (76%) records that were assessed for eligibility. Of these 54 records, 5 (9%) were included. This resulted in 33 included papers in this systematic review (Figure 1 shows a flow diagram of the results in the different selection stages). In both stages, a consensus was reached by the reviewers regarding the inclusion and analysis of the papers. All the included studies were published between 2010 and 2022. An overview of these studies is presented in Table 1. In Table 1, to avoid misinterpretation, we retained the terms that were used by the authors to describe their participants.

The following section outlines the characteristics of the included studies and the results orientated toward the RQs. First, we have described whether the studies included service users, people with lived experience, or both, as well as the age group of the involved participants. Then, we have summarized the aim of the innovation and reported the design approaches used. To address the level of involvement, we have presented the roles of service users and people with lived experiences. Finally, we have elaborated on the included studies to provide a context for the reported added value and challenges.

Figure 1. Selection process: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram.

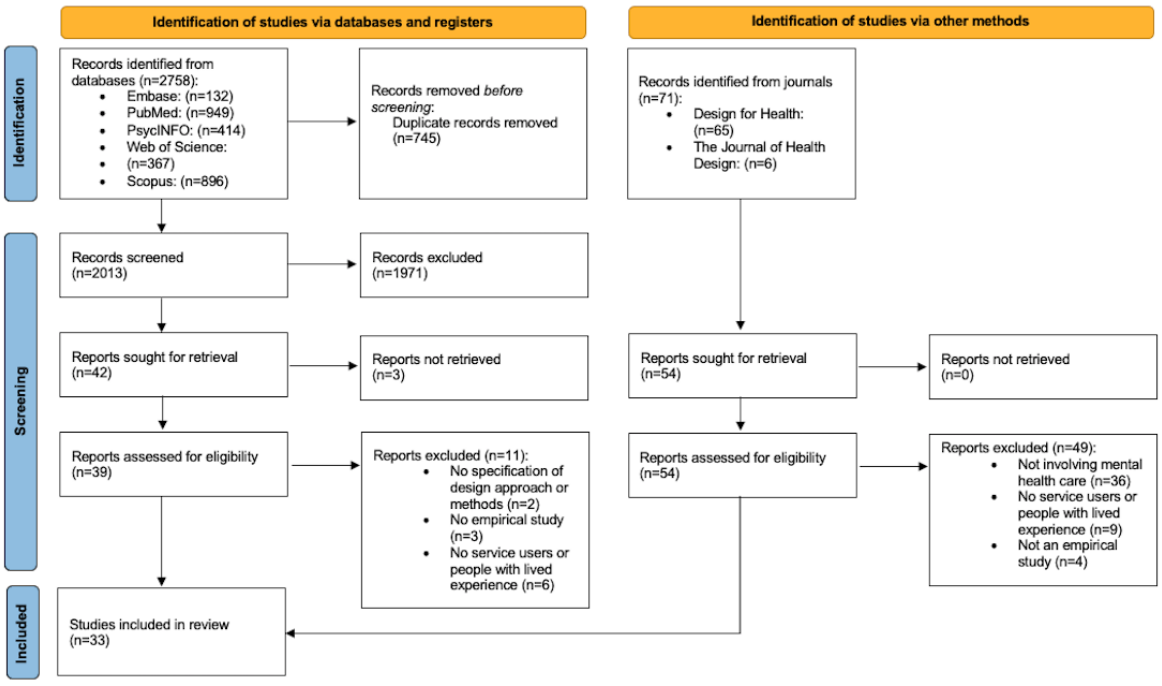


Table 1. Overview and categorization of the included studies.

Study, year	Study title	Participants and age group	Aim or aims of innovation	Approach and role
Owens et al [43], 2010	Involving service users in intervention design: a participatory approach to developing a text - messaging intervention to reduce repetition of self - harm	• Service users • Adults	Digital intervention development	• Participatory workshops • Cothinkers
Gammon et al [44], 2014	Service users' perspectives in the design of an online tool for assisted self-help in mental health: a case study of implications	• Service users • Adults	Digital intervention development—recovery orientated	• Community-based participatory research with iterative and cocreative design • Cothinkers
Terp et al [45], 2016	A room for design: through participatory design young adults with schizophrenia become strong collaborators	• Service users • Adolescents	Digital intervention development	• Co-design • Partners
Grim et al [46], 2017	Development and usability testing of a web-based decision support for users and health professionals in psychiatric services	• Service users • People with lived experience • Adults	Digital intervention development	• Participatory design • Cothinkers
Nakarada-Kordic et al [6], 2017	Co-designing for mental health: creative methods to engage young people experiencing psychosis	• Service users • Young people	Service improvement	• Co-design • Partners
Schmitt and Yarosh [47], 2018	Participatory design of technologies to support recovery from substance use disorders	• Service users • Adults	Digital intervention development—recovery orientated	• Participatory design • Cothinkers
McClelland and Fitzgerald [48], 2018	A participatory mobile application (app) development project with mental health service users and clinicians	• Service users • Not specified	Digital intervention development	• Co-design • Cothinkers
Vilardaga et al [49], 2018	User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness	• Service users • Adults	Digital intervention development	• User-centered design • Cothinkers
Terp et al [50], 2018	A smartphone app to foster power in the everyday management of living with schizophrenia: qualitative analysis of young adults' perspectives	• Service users • Adults	Digital intervention development—recovery orientated	• Participatory design • Partners
Vieira da Silva and Bueno [51], 2018	Compass: a personal organization mobile app for individuals with mental disorders	• Service users and people with lived experience • Adults	Digital intervention development	• User-centered design • Cothinkers
Hackett et al [52], 2018	Co-designing for quality: creating a user-driven tool to improve quality in youth mental health services	• Service users • Young people	Digital intervention development	• Experience-based co-design • Advisers
Mulvale et al [53], 2019	Co-designing services for youth with mental health issues: novel elicitation approaches	• Service users • Young people	Service improvement	• Co-design • Advisers
Romm et al [54], 2019	Designing easy access to care for first-episode psychosis in complex organizations	• Service users • Adults	Service improvement	• Service design • Cothinkers

Study, year	Study title	Participants and age group	Aim or aims of innovation	Approach and role
Derks et al [55], 2019	Development of an ambulatory biofeedback app to enhance emotional awareness in patients with borderline personality disorder: multicycle usability testing study	• Service users • Adults	Digital intervention development—recovery orientated	• User-centered design • Cothinkers
Realpe et al [56], 2019	Co-designing a virtual world with young people to deliver social cognition therapy in early psychosis	• Service users and people with lived experience • Young people	Digital intervention development—recovery orientated	• Co-design • Advisers
Maathuis et al [57], 2019	Exploring human values in the design of a web-based QoL[a] instrument for people with mental health problems: a value-sensitive design approach	• Service users • Not specified	Digital intervention development	• Value-sensitive design • Cothinkers
Fonseka et al [58], 2019	Collaborating with individuals with lived experience to adapt CANMAT[b] clinical depression guidelines into a patient treatment guide: the CHOICE-D[c] co-design process	• People with lived experience • Adults	Service improvement—recovery orientated	• Co-design • Partners
Terlouw et al [59], 2020	Design of a digital comic creator (it's me) to facilitate social skills training for children with autism spectrum disorder: design research approach	• Service users • Young people	Digital intervention development	• Design research • Advisers
Stawarz et al [60], 2020	Design considerations for the integrated delivery of cognitive behavioral therapy for depression: user-centered design study	• Service users • Adults	Digital intervention development	• User-centered design • Advisers
Callan et al [61], 2020	CBT[d] MobileWork: user-centered development and testing of a mobile mental health application for depression	• Service users • Adults	Digital intervention development—recovery orientated	• User-centered design • Cothinkers
Flobak et al [62], 2021	Designing videos with and for adults with ADHD[e] for an online intervention: participatory design study and thematic analysis of evaluation	• People with lived experience • Adults	Digital intervention development	• Participatory design • Cothinkers
van der Meer et al [63], 2021	Targeting personal recovery of people with complex mental health needs: the development of a psychosocial intervention through user-centered design	• Service users • Adults	Recovery orientated	• User-centered design • Advisers
García et al [64], 2021	Co-design of avatars to embody auditory hallucinations of patients with schizophrenia: a study on patients' feeling of satisfaction and psychiatrists' intention to adopt the technology	• Service users • Adults	Digital intervention development—recovery orientated	• Co-design • Advisers
Jonathan et al [65], 2021		• Service users • Adults		

Study, year	Study title	Participants and age group	Aim or aims of innovation	Approach and role
	A smartphone-based self-management intervention for bipolar disorder (livewell): user-centered development approach		Digital intervention development—recovery orientated	• User-centered development approach • Cothinkers
Milton et al [66], 2021	Technology-enabled reform in a nontraditional mental health service for eating disorders: participatory design study	• People with lived experience • Not specified	Digital intervention development—recovery orientated	• Participatory design • Cothinkers
Sanin et al [67], 2021	Creative well-being. prototyping an arts-health practice program for mental health recovery	• Service users • Not specified	Service improvement—recovery orientated	• Participatory design • Autonomous design • Partners
Knight et al [14], 2021	Participatory design to create a VR[f] therapy for psychosis	• People with lived experience • Not specified	Digital intervention development	• Participatory design • Advisers
Kruzan et al [68], 2022	Centering lived experience in developing digital interventions for suicide and self-injurious behaviors: user-centered design approach	• People with lived experience • Adolescents	Digital intervention development—recovery orientated	• User-centered design • Advisers
Bos et al [69], 2022	A web-based application for personalized ecological momentary assessment in psychiatric care: user-centered development of the PETRA[g] application	• Service users • Not specified	Digital intervention development	• User-centered development approach • Cothinkers
Bongers et al [70], 2022	I need to know: using the CeHRes[h] roadmap to develop a treatment feedback tool for youngsters with mental health problems	• People with lived experience • Young people	Digital intervention development	• Co-design • Advisers
Wiberg et al [71], 2022	Internet-based cognitive behavior therapy for eating disorders—development and feasibility evaluation	• Service users and people with lived experience • Adults	Digital intervention development—recovery orientated	• User-centered design • Cothinkers
Illarregi et al [72], 2022	Is designing therapeutic? a case study exploring the experience of co-design and psychosis	• Service user • Adult	Intervention development	• Co-design • Partner
Jenness et al [73], 2022	Lessons learned from designing an asynchronous remote community approach for behavioral activation intervention for teens	• Service users • Teenagers and adolescents	Digital intervention development	• Human-centered design • Advisers

[a]QoL: quality of life.

[b]CANMAT: Canadian Network for Mood and Anxiety Treatments.

[c]CHOICE-D: Canadian Network for Mood and Anxiety Treatments Health Options for Integrated Care and Empowerment in Depression.

[d]CBT: cognitive behavioral therapy.

[e]ADHD: attention-deficit/hyperactivity disorder.

[f]VR: virtual reality.

[g]PETRA: Personalized Treatment by Real-time Assessment.

[h]CeHRes: Centre for eHealth Research roadmap.

Characteristics of the Included Studies

All the included peer-reviewed articles were published between 2010 and 2022.

Participants

Of the 33 studies, 6 (18%) involved individuals with lived experience [14,58,62,66,68,70], 23 (70%) involved service users [6,43-45,47-50,52-55,57,59-61,63-65,67,69,72,73], and 4 (12%) involved both [46,51,56,71]. A total of 18 (55%) studies of the 33 studies included adult participants [43-47,49-51,54,55,58,60-65,71,72], 6 (18%) studies described their participants as young people [6,52,53,56,59,70], 3 (9%) studies described their participants as adolescents [45,68,73], and 6 (18%) did not specify the age group of their participants [14,48,57,66,67,69].

Aim of the Innovation

In total, 27 studies aimed to develop interventions, with 12 focusing on recovery [44,47,50,55,56,61,63-66,68,71], 3 targeting service improvement [6,53,54], and 2 targeting both [58,67]. A total of 26 studies focused on digital innovation design or evaluation [14,43-52,55-57,59-62,64-66,68-71,73] and 1 focused on design as an intervention itself [72].

Design Approaches

Co-design was used as the design approach in 9 studies [6,45,48,53,56,58,64,70,72]. User-centered design was used in 8 studies [49,51,55,60,61,63,68,71], and participatory design was used in 7 studies [14,46,47,50,62,66,67]. Other identified approaches included value-sensitive design [57], experience-based co-design [52], service design [54], design research [59], human-centered design [73], user-centered development [65,69], community-based participatory research [44], and participatory workshops [43].

Level of Involvement

The Involvement Matrix [38] was used to ascertain the roles of service users and individuals with lived experience in the analyzed studies. None of the studies included the participants as "listeners," whereas 16 studies featured participants in the role of a "cothinker" [43,44,46-49,51,54,55,57,61,62,65,66,69,71], providing their opinions on ideas or evaluating test sessions of innovations developed by the researchers. In addition, participants served as "advisers" in 11 studies, offering solicited or unsolicited feedback to researchers and designers [14,52,53,56,59,60,63,64,68,70,73]. In 6 studies, participants functioned as equal "partners" with researchers and designers [6,45,50,58,67,72]. However, none of the studies identified participants in the decision maker role and reported involving participants in the final decision-making processes. The studies in which co-design was used as the approach exhibited the highest levels of participant involvement.

Reported Added Values

Overview

The included studies reported a wide range of added value of the involvement of service users and people with lived experience through design approaches. Not every article mentioned specific design approach–related benefits [49,70]. On the basis of the thematic synthesis, we divided these added values into 4 categories.

Design for Tailoring and Testing (Digital) Applications and Interventions

In total, 6 studies reported on the added value of tailoring and testing existing designs with service users and people with lived experience [51,55,61,64,69,71]. Vieira da Silva and Bueno [51] adopted a user-centered design approach for the development and testing of the Compass app, which aimed to support people with mental health problems in their daily lives. The study found that patient testing provided key learnings for the final prototype, such as improvement of the interface design, resulting in the app meeting the users' needs and expectations. Derks et al [55] cyclically tested an outpatient biofeedback application to increase emotional awareness in patients with borderline personality disorders. Testing with service users provided insights into the usability of the application. Wiberg et al [71] used a user-centered design process for the development and feasibility assessment of an internet-based cognitive behavioral therapy for patients with eating disorders. This approach contributed to improvements and adjustments to the program according to the end users' needs and perceptions. Bos et al [69] described a user-centered development approach for a web-based application for personalized ecological momentary assessment in psychiatric care. The user-centered approach ensured that the developed application, which was tested with service users, was intuitive, user-friendly, and useful for clients and clinicians. García et al [64] co-designed avatars to embody auditory hallucinations. The co-design approach guaranteed that the service users found the developed software to be complete and useful for representing their voices. Callan et al [61] used an iterative design process to develop and test a mobile health app for individuals with depression, incorporating user feedback to improve the app's appearance, navigation, content, and organization.

Design for Increasing Engagement and Collaboration in the Development of Interventions

We observed that 10 studies reported added value on increasing user engagement and collaboration in the development of interventions [14,52-54,58,59,65-68]. Romm et al [54] focused on the use of service design to address known barriers to developing early intervention services in complex health care organizations. By emphasizing cocreation and divergent thinking, service design encouraged experimentation and innovation to improve service development, engaging service users in organizational improvement. Jonathan et al [65] reported on a user-centered design process for developing a smartphone-based self-management intervention for individuals with bipolar disorder. This approach prioritized collaboration among stakeholders, which helped to personalize service users' goals and plans and integrate human support as a self-management tool. Kruzan et al [68] presented a user-centered design approach to engage individuals with self-harming thoughts and behaviors who may be uncomfortable in traditional in-person therapy settings. User-centered design methods allow for flexibility in engaging individuals and

facilitating collaboration in the most comfortable and accessible ways, such as through the internet. Sanin et al [67] collaborated with occupational therapists and inpatients to develop an arts-health program prototype. The cocreation of designers and occupational therapists generated a collaborative design process, expanding the creative practices of occupational therapists. Hackett et al [52] described the experience-based co-design process for improving the quality of youth mental health services. The co-design event emphasized communication and collaboration in delivering quality care, and the feedback tools developed by young participants served as an important starting point for continuous quality improvement.

In another article, Mulvale et al [53] highlighted the effectiveness of elicitation techniques such as experience maps, trigger videos, and prototyping in promoting mutual understanding and shared ideas for system changes. The experience-based co-design process, which prioritized the engagement of people with lived experience in the design of mental health services, was consistent with a recovery orientation that calls for collaboration with service users in the design of services. Terlouw et al [59] described the design of a digital comic maker for children with autism spectrum disorders. The design approach led to finding different problem perceptions among stakeholders and added value to the acceptability of the developed innovation by exploring and sharing different sensemaking processes from different stakeholder perspectives. Milton et al [66] reported on the use of participatory design processes to customize and configure a technological solution for a nontraditional mental health service for people with eating disorders and body image issues. The authors reported that participatory design processes enabled a collaborative approach to the customization and configuration of the platform. Knight et al [14] described a participatory design process for developing virtual reality therapy for people with psychosis, which facilitated collaboration between expert groups, including individuals with lived experience of psychosis. This approach led to design and innovation within a shared understanding of limitations and evidence-based design. Fonseka et al [58] adapted a clinical guideline through a co-design approach to an accessible version for patients and families, emphasizing the importance of informed discussion and partnership between people with lived experience and researchers and designers.

Design for Identifying the Needs of Service Users and People With Lived Experience

There were 9 studies that reported added value in identifying user needs [43,46-48,56,57,60,72,73]. Schmitt and Yarosh [47] reported on the value of participatory design workshops in aiding the recovery of women with substance use disorders. The workshops provided new insights into the participants' needs, emphasizing the importance of anonymity and safety in recovery. Owens et al [43] used a participatory approach to develop a SMS text messaging intervention to reduce repetitive self-harm and found that people with lived experience preferred individualized texting instead of a generic one-size-fits-all approach. Grim et al [46] used participatory design processes to develop a web-based system and found that incorporating user perspectives can tailor an innovation to the needs of the user group. Maathuis et al [57] used a value-sensitive design approach to identify potential value conflicts in the design of an internet-based quality-of-life tool for people with mental health problems. The study showed how the approach made it possible to identify and anticipate potential value conflicts and patients' needs in the design. McClelland and Fitzgerald [48] used a co-design approach to develop a mobile app for service users involved in the early intervention services for psychosis. The authors described key needs that emerged from service user involvement, and they incorporated those needs into the app prototype design, emphasizing that early input is essential to design flexible and relevant contents that meets user needs.

Stawarz et al [60] identified new perspectives and requirements by exploring service users' needs using user-centered design methods. As a result, the authors formulated 12 design considerations for the integrated delivery of cognitive behavioral therapy for people with depression. Jenness et al [73] used a human-centered design process to develop an app to provide an evidence-based psychosocial intervention for teenagers with depression, and they found that it helped teenagers reduce avoidance. Involving service users in the design process led to design changes based on their needs, such as personalizing the timing and frequency of logging reminders based on participants' routines and times when they felt most comfortable. Realpe et al [56] found that the co-design process with young people with psychosis led to unexpected insights regarding their needs and wishes, which challenged the researchers' understanding of what service users needed during recovery. For example, the approach revealed that young people preferred designs that resembled familiar environments with an urban feel, such as local therapeutic spaces or classrooms, rather than stereotypical places of leisure. Illarregi et al [72] focused on the experience of 1 service user in a co-design study, providing insights into how the design activity can support recovery.

Design for Empowering Service Users and People With Lived Experience as Resourceful Actors

In total, 6 studies reported the added values of empowering participants as resourceful actors through design approaches [6,44,45,50,62,63]. Gammon et al [44] highlighted the importance of involving service users in the design process of a web-based tool for assisted self-help in mental health. Their study showed that involving service users through design approaches could reveal gaps in relevance between mainstream research and service users' interests, and the involvement fosters commitment to follow through in the implementation and research phases. The authors found that the approach stimulated to involve service users, as no one is more capable of conveying the intentions and functionalities of the innovation than the service users. Flobak et al [62] reported that the participatory design could balance the perspectives of people with attention-deficit/hyperactivity disorder and clinicians, leading to mutual learning. The authors noted that in their study, participants were not only experts on what they liked or disliked but also very knowledgeable about ADHD, as they had first-hand experience and tacit knowledge of the ADHD diagnosis. This experiential knowledge was initially unavailable to the clinicians and deemed very important for the design process of the intervention. van der Meer et al [63] found that service users with complex mental health problems could

meaningfully participate in and contribute to understanding the problem and reflecting on the form and content of a psychosocial intervention. Even service users who were less able to express their needs and desires regarding the content or design of the intervention could still identify the factors they felt were important to consider, and these factors played crucial roles in developing the psychosocial intervention.

Nakarada-Kordic et al [6] reported that service users identified unique needs and interests that differed significantly from those identified by clinicians. Creative participatory methods meaningfully involved young people experiencing psychosis in the process. The study emphasized the importance of treating young people as equal partners and acknowledging them as experts in their own conditions and experiences. The design workshops showed the authors that the unique needs and interests of the participants posed a greater challenge than their severe mental health issues. Terp et al [45] used a design approach to develop a mobile phone tool that enabled young adults with schizophrenia to be actively involved in designing more participatory mental health services. The authors emphasized the importance of giving the community of practice a name and body from the beginning of the design process so that young adults with schizophrenia feel included solely as experts in their own experience and not as patients. The community of practice functioned as an "identity changer" from "receiver" to "giver," or from a patient in need to a designer of the need. Terp et al [50] concluded in a follow-up study among young adults with schizophrenia that close collaboration between designers, researchers, and users ensured that the app met the target audience's needs.

Reported Challenges

The involvement of service users and people with lived experience through design approaches presents several challenges. Participants' continued participation throughout the study is challenging due to changing mental states, economic constraints, recruitment and panel size, and difficulty in reaching participants [43,53,54,58,64,70,73]. The iterative nature of design makes it difficult to record dynamic individual small group work [43,63,66], and using a design approach is reported to be time-consuming [44,46,48,56]. Providing a safe space for service users and people with lived experience is not only important but also challenging [45,53,68], and designers must be aware of the trade-off between preserving authenticity and reinforcing the stigmatizing characteristics of mental health [62]. Balancing insights from design and medical science can be challenging, as stakeholder recommendations may be inconsistent with evidence-based practices or theory-based principles of change [49,57]. Managing design processes in medical environments is challenging because of the disruptive nature of co-design formats and methods, difficulty in finding a balance between easy navigation in the app and meeting participant requests for additional features, and logistics of capturing inputs from multiple concurrent discussions [6,65,67].

Discussion

Principal Findings

This systematic review highlights the advantages of involving service users and people with lived experience through design approaches for innovation in mental health care. Innovation projects in mental health care involve service users and people with lived experience in design processes, with the first publication occurring in 2010. Most studies involved service users, mostly adults. Co-design is the most frequently used design approach. Innovation primarily aimed to develop digital interventions that promote recovery. Participants were often involved as cothinkers and advisers, and sometimes as partners, but never as decision makers. We identified 4 themes in which we categorized the added values of design approaches, namely, design for tailoring and testing interventions and applications, to increase engagement and collaboration in the development of interventions, to identify the needs of service users and people with lived experience, and to empower them as resourceful actors. The challenges reported were maintaining participants' continued participation throughout the study, managing the iterative nature of design, providing a safe space, balancing insights from design and medical science, and navigating design processes in medical environments.

Many of the included studies emphasized the end results rather than the design processes. Moreover, most studies did not describe whether and how they evaluated the participants' experiences in the design approach. This can be explained by the fact that the objectives of these studies were not primarily aimed at reporting on the involvement of the participants and their experiences but rather on the development of the interventions or applications. In the studies that we categorized as promoting the partner role, co-design was the design approach that was used most often. This shows the promise of the co-design approach for future innovation projects in mental health care, particularly when significant participation of service users and people with lived experience is required. In line with earlier research, further advantages can be obtained by engaging end users early in the design processes [74]. Many studies reported added values on various topics that centered on the engagement and collaboration in innovation projects. However, based on the findings in this review, this collaboration focused usually on the goals of the researchers and not on the benefits that the design approaches can have for service users and people with lived experience. Although multiple challenges were discovered, these seem to be manageable and can be overcome with more methodological and organizational "design mindedness," in line with earlier research on the challenges of design in health [75-79].

Comparison With Previous Findings

Several studies overlap the scope of this systematic review. In an exploratory mapping review, Vial et al [33] examined the literature to understand how human-centered design is considered in e-mental health intervention research. Their findings show that the included studies relied very little on designers and design research, with limited involvement of end users in the design process. In our review, design approaches

were used in most studies as a form of traditional research as opposed to how these approaches are described in the design literature, and the level of involvement of service users and people with lived experience was typically low for design processes. The results of our systematic review are also consistent with the those of 4 other reviews [80-83]: co-design processes involving service users and people with lived experience contribute to person-centered innovation; as a result, design can help shift the power balance in favor of susceptible populations. However, the implementation and explanation of design requirements are lacking, and the reporting offers no substantial ground for definitive conclusions. In our review, many studies that created an intervention or application did not describe a design rationale, whereas previous research has shown that this is essential in determining the thoughts and foundations of an intervention to enable researchers to build on it [74,84,85].

This systematic review leverages the existing literature to show the potential of design approaches to empower service users and people with lived experience as resourceful actors in innovation processes in mental health care. This finding suggests that high levels of involvement through design may not only have beneficial outcomes for the greater good of the developed product but also contribute to the benefits of the participants involved. Although this value could be miniscule, it is an effect that has not been emphasized previously, exhibiting some therapeutic potential. The studies wherein the design approaches display high levels of involvement seem to foster equal cooperation and shared objectives, which share similarities with the fundamental elements of therapy, such as therapeutic alliances and expectation-rich therapeutic rituals [86]. In addition, using design approaches seems to align with the need to prioritize the use of participatory methods that facilitate the collaboration with service users, empowering them to meaningfully influence design decisions as important stakeholders [87].

Strengths and Limitations

The purpose of this systematic review is to provide an overview of the involvement of service users and people with lived experience in mental health care services through design approaches and to identify its advantages. Therefore, we used broad search terms to include as many studies as possible that covered the topic and provide a complete overview of its advantages. To ensure methodological quality, all the included studies were peer reviewed and published in academic journals. To analyze the results, we used a thematic synthesis to have the categories emerge from the data. Although this approach is appropriate for finding new data, the results of the synthesis may appear different from an a priori theoretically driven approach [42]. However, a priori themes could have resulted in overlooking advantages that did not fit within these established categories, such as the potential that design approaches have for empowering service users as resourceful actors. This could have limited the overview of the design approaches and their advantages, which was the main aim of this systematic review.

Future Research

Although the current emphasis on empowerment, autonomy, involvement, and participation in design-based innovation is a positive development, sharing numerous principles with the recovery movement [88], the included studies provide little information on what they consider involvement and participation, how they intend it, and how they integrate these principles into their methodology. Given that most studies that were categorized as promoting the partner role were also the studies that were categorized as empowering, we recommend that future studies examine what partner and decision maker roles can accomplish to yield therapeutic and recovery outcomes. Furthermore, we suggest that future design studies that focus on innovation in mental health care use a framework that helps to report engagement with service users and people with lived experience in a systemic and clear manner, such as the Involvement Matrix [38]. Our review identified relevant data on the involvement of service users and people with lived experience, but these data were reported in various sections of the papers, were often brief, and lacked detail, corresponding with previous research on patient and public involvement in health and social care research [89]. Involvement frameworks can be valuable when patient-researcher partnerships are led by researchers with little experience involving service users in research [90]. Other directions for future research are examining the different uses of experiential knowledge and its usefulness for design research, investigating whether the use of experiential knowledge by mental health professionals can be beneficial, and exploring how research can benefit from cocreation and how these impact service users, people with lived experience, and mental health professionals.

Conclusions

This systematic review provides an overview of the involvement of service users and people with lived experience in mental health care services through design approaches and their advantages. The results show that design approaches add value to tailoring and testing applications, interventions, digital applications, and digital interventions; improving engagement and collaboration in the development of interventions; identifying the needs of stakeholders; and empowering service users and people with lived experience as resourceful actors. The challenges that were identified are maintaining participants' continued participation throughout the study, managing the iterative nature of design, providing a safe space, balancing insights from design and medical science, and navigating design processes in medical environments. Of all the design approaches, co-design was identified as achieving the highest level of involvement and can be useful for innovation in mental health care, wherein significant participation of service users and people with lived experience is required. Future studies using design approaches in mental health care are recommended to systematically and clearly report on involvement and collaboration.

Acknowledgments

The authors appreciate the financial support of the Faith Research Consortium, GGZ-VS University of Applied Science, and the NHL Stenden University of Applied Sciences Doctor of Philosophy program. These funders had no role in the content of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of search terms.
[DOCX File , 18 KB - mental_v10i1e46590_app1.docx]

References

1. Freeman M. The World Mental Health Report: transforming mental health for all. World Psychiatry 2022 Oct 08;21(3):391-392 [FREE Full text] [doi: 10.1002/wps.21018] [Medline: 36073688]
2. Stein DJ, Shoptaw SJ, Vigo DV, Lund C, Cuijpers P, Bantjes J, et al. Psychiatric diagnosis and treatment in the 21st century: paradigm shifts versus incremental integration. World Psychiatry 2022 Oct 08;21(3):393-414 [FREE Full text] [doi: 10.1002/wps.20998] [Medline: 36073709]
3. Borsboom D, Haslbeck JM, Robinaugh DJ. Systems-based approaches to mental disorders are the only game in town. World Psychiatry 2022 Oct 08;21(3):420-422 [FREE Full text] [doi: 10.1002/wps.21004] [Medline: 36073701]
4. Thornicroft G. Psychiatric diagnosis and treatment in the 21st century: paradigm shifts or power shifts? World Psychiatry 2022 Oct 08;21(3):334-335 [FREE Full text] [doi: 10.1002/wps.21000] [Medline: 36073680]
5. Veldmeijer L, Terlouw G, van ‘t Veer J, Van Os J, Boonstra N. Design for mental health: can design promote human-centred diagnostics? Design Health 2023 Feb 14;7(1):5-23 [FREE Full text] [doi: 10.1080/24735132.2023.2171223]
6. Nakarada-Kordic I, Hayes N, Reay SD, Corbet C, Chan A. Co-designing for mental health: creative methods to engage young people experiencing psychosis. Design Health 2017 Oct 24;1(2):229-244. [doi: 10.1080/24735132.2017.1386954]
7. Szücs Johansson L, Vink J, Wetter-Edman K. A Trojan horse approach to changing mental health care for young people through service design. Design Health 2017 Oct 24;1(2):245-255. [doi: 10.1080/24735132.2017.1387408]
8. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
9. Sumartojo S, Pink S, Duque M, Vaughan L. Atmospheres of care in a psychiatric inpatient unit. Design Health 2020 Mar 05;4(1):24-42. [doi: 10.1080/24735132.2020.1730068]
10. Farahnak Majd N, Golembiewski J, Tarkashvand A. The psychiatric facility: how patients with schizophrenia respond to place. Design Health 2020 Nov 28;4(3):384-406. [doi: 10.1080/24735132.2020.1846849]
11. Liddicoat S. Mental health facility codesign: a new research method for integrating the service user voice in design processes using virtual reality. Gen Psychiatr 2019 Jul 23;32(3):e100061 [FREE Full text] [doi: 10.1136/gpsych-2019-100061] [Medline: 31423474]
12. Adams L, Simonoff E, Tierney K, Hollocks MJ, Brewster A, Watson J, et al. Developing a user-informed intervention study of a virtual reality therapy for social anxiety in autistic adolescents. Design Health 2022 Apr 24;6(1):114-133. [doi: 10.1080/24735132.2022.2062151]
13. Terlouw G, Kuipers D, van 't Veer J, Prins JT, Pierie JP. The development of an escape room-based serious game to trigger social interaction and communication between high-functioning children with autism and their peers: iterative design approach. JMIR Serious Games 2021 Mar 23;9(1):e19765 [FREE Full text] [doi: 10.2196/19765] [Medline: 33755023]
14. Knight I, West J, Matthews E, Kabir T, Lambe S, Waite F, et al. Participatory design to create a VR therapy for psychosis. Design Health 2021 Feb 26;5(1):98-119. [doi: 10.1080/24735132.2021.1885889]
15. Berwick DM. Era 3 for medicine and health care. JAMA 2016 Apr 05;315(13):1329-1330. [doi: 10.1001/jama.2016.1509] [Medline: 26940610]
16. Groenewoud AS, Westert GP, Kremer JA. Value based competition in health care's ethical drawbacks and the need for a values-driven approach. BMC Health Serv Res 2019 Apr 27;19(1):256 [FREE Full text] [doi: 10.1186/s12913-019-4081-6] [Medline: 31029136]
17. Frances AJ, Widiger T. Psychiatric diagnosis: lessons from the DSM-IV past and cautions for the DSM-5 future. Annu Rev Clin Psychol 2012 Apr 27;8(1):109-130. [doi: 10.1146/annurev-clinpsy-032511-143102] [Medline: 22035240]
18. Guloksuz S, van Os J. The slow death of the concept of schizophrenia and the painful birth of the psychosis spectrum. Psychol Med 2017 Jul 10;48(2):229-244. [doi: 10.1017/s0033291717001775]
19. Slade M, Longden E. Empirical evidence about recovery and mental health. BMC Psychiatry 2015 Nov 14;15:285 [FREE Full text] [doi: 10.1186/s12888-015-0678-4] [Medline: 26573691]

20. van Weeghel J, van Zelst C, Boertien D, Hasson-Ohayon I. Conceptualizations, assessments, and implications of personal recovery in mental illness: a scoping review of systematic reviews and meta-analyses. Psychiatr Rehabil J 2019 Jun;42(2):169-181. [doi: 10.1037/prj0000356] [Medline: 30843721]
21. Leichsenring F, Steinert C, Rabung S, Ioannidis JP. The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses. World Psychiatry 2022 Feb 11;21(1):133-145 [FREE Full text] [doi: 10.1002/wps.20941] [Medline: 35015359]
22. van Os J, Guloksuz S. Schizophrenia as a symptom of psychiatry's reluctance to enter the moral era of medicine. Schizophr Res 2022 Apr;242:138-140 [FREE Full text] [doi: 10.1016/j.schres.2021.12.017] [Medline: 34991949]
23. Latour B. A cautious Prometheus? A few steps toward a philosophy of design with special attention to Peter Sloterdijk. In: In Medias Res. Amsterdam: Amsterdam University Press; 2011:151-164.
24. Fancourt D, Saoirse F. What is the Evidence on the Role of the Arts in Improving Health and Well-being? A Scoping Review. Geneva, Switzerland: World Health Organization; 2019.
25. Cross N. Designerly Ways of Knowing. London, England: Springer; 2006.
26. Verhagen S, van Os J, Delespaul P. Ecological momentary assessment and other digital technologies for capturing daily life in mental health. In: Mental Health in a Digital World. Cambridge, Massachusetts, United States: Academic Press; 2022:81-108.
27. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun 21;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]
28. Thabrew H, Fleming T, Hetrick S, Merry S. Co-design of eHealth interventions with children and young people. Front Psychiatry 2018 Oct 18;9:481 [FREE Full text] [doi: 10.3389/fpsyt.2018.00481] [Medline: 30405450]
29. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. Qual Saf Health Care 2006 Oct 01;15(5):307-310 [FREE Full text] [doi: 10.1136/qshc.2005.016527] [Medline: 17074863]
30. Veldmeijer L, Wartena B, Terlouw G, van't Veer J. Reframing loneliness through the design of a virtual reality reminiscence artefact for older adults. Design Health 2020 Nov 26;4(3):407-426. [doi: 10.1080/24735132.2020.1848976]
31. Friedman B. Value-sensitive design. Interactions 1996 Dec;3(6):16-23. [doi: 10.1145/242485.242493]
32. Hassenzahl M, Diefenbach S, Göritz A. Needs, affect, and interactive products – facets of user experience. Interact Comput 2010 Sep;22(5):353-362. [doi: 10.1016/j.intcom.2010.04.002]
33. Vial S, Boudhraâ S, Dumont M. Human-centered design approaches in digital mental health interventions: exploratory mapping review. JMIR Ment Health 2022 Jun 07;9(6):e35591 [FREE Full text] [doi: 10.2196/35591] [Medline: 35671081]
34. Sanders EB, Stappers PJ. Convivial Toolbox Generative Research for the Front End of Design. Amsterdam, Netherlands: BIS Publishers; 2012.
35. Dixon BS. Dewey and Design: A Pragmatist Perspective for Design Research. Cham, Switzerland: Springer Nature; 2020.
36. Koskinen I, Zimmerman J, Binder T, Redstrom J, Wensveen S. Design Research Through Practice From the Lab, Field, and Showroom. Burlington, Massachusetts, United States: Morgan Kaufmann; 2011.
37. Brown T. Change by Design How Design Thinking Transforms Organizations and Inspires Innovation. New York, United States: HarperCollins; 2009.
38. Smits DW, van Meeteren K, Klem M, Alsem M, Ketelaar M. Designing a tool to support patient and public involvement in research projects: the Involvement Matrix. Res Involv Engagem 2020 Jun 16;6(1):30 [FREE Full text] [doi: 10.1186/s40900-020-00188-4] [Medline: 32550002]
39. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. J Clin Epidemiol 2021 Jun;134:178-189 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.03.001] [Medline: 33789819]
40. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
41. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
42. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. ESRC Methods Programme. 2006. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2023-04-15]
43. Owens C, Farrand P, Darvill R, Emmens T, Hewis E, Aitken P. Involving service users in intervention design: a participatory approach to developing a text-messaging intervention to reduce repetition of self-harm. Health Expect 2011 Sep;14(3):285-295 [FREE Full text] [doi: 10.1111/j.1369-7625.2010.00623.x] [Medline: 20860777]
44. Gammon D, Strand M, Eng LS. Service users' perspectives in the design of an online tool for assisted self-help in mental health: a case study of implications. Int J Ment Health Syst 2014 Jan 09;8(1):2 [FREE Full text] [doi: 10.1186/1752-4458-8-2] [Medline: 24405765]
45. Terp M, Laursen BS, Jørgensen R, Mainz J, Bjørnes CD. A room for design: through participatory design young adults with schizophrenia become strong collaborators. Int J Ment Health Nurs 2016 Dec 13;25(6):496-506. [doi: 10.1111/inm.12231] [Medline: 27293176]

46. Grim K, Rosenberg D, Svedberg P, Schön UK. Development and usability testing of a web-based decision support for users and health professionals in psychiatric services. Psychiatr Rehabil J 2017 Sep;40(3):293-302. [doi: 10.1037/prj0000278] [Medline: 28737415]
47. Schmitt Z, Yarosh S. Participatory design of technologies to support recovery from substance use disorders. Proc ACM Hum Comput Interact 2018 Nov;2(CSCW):1-27. [doi: 10.1145/3274425]
48. McClelland GT, Fitzgerald M. A participatory mobile application (app) development project with mental health service users and clinicians. Health Educ J 2018 Jun 05;77(7):815-827. [doi: 10.1177/0017896918773790]
49. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
50. Terp M, Jørgensen R, Laursen BS, Mainz J, Bjørnes CD. A smartphone app to foster power in the everyday management of living with schizophrenia: qualitative analysis of young adults' perspectives. JMIR Ment Health 2018 Oct 01;5(4):e10157 [FREE Full text] [doi: 10.2196/10157] [Medline: 30274966]
51. Vieira da Silva AC, Bueno J. Compass. Inf Des J 2019 Sep 23;24(3):220-235. [doi: 10.1075/idj.24.3.02vie]
52. Hackett CL, Mulvale G, Miatello A. Co-designing for quality: creating a user-driven tool to improve quality in youth mental health services. Health Expect 2018 Dec;21(6):1013-1023 [FREE Full text] [doi: 10.1111/hex.12694] [Medline: 29707865]
53. Mulvale G, Moll S, Miatello A, Murray-Leung L, Rogerson K, Sassi RB. Co-designing services for youth with mental health issues: novel elicitation approaches. Int J Qual Methods 2019 Feb 21;18:160940691881624. [doi: 10.1177/1609406918816244]
54. Romm KL, Gardsjord ES, Gjermundsen K, Aguirre Ulloa M, Berentzen LC, Melle I. Designing easy access to care for first-episode psychosis in complex organizations. Early Interv Psychiatry 2019 Oct 27;13(5):1276-1282. [doi: 10.1111/eip.12802] [Medline: 30919597]
55. Derks YP, Klaassen R, Westerhof GJ, Bohlmeijer ET, Noordzij ML. Development of an ambulatory biofeedback app to enhance emotional awareness in patients with borderline personality disorder: multicycle usability testing study. JMIR Mhealth Uhealth 2019 Oct 15;7(10):e13479 [FREE Full text] [doi: 10.2196/13479] [Medline: 31617851]
56. Realpe A, Elahi F, Bucci S, Birchwood M, Vlaev I, Taylor D, et al. Co-designing a virtual world with young people to deliver social cognition therapy in early psychosis. Early Interv Psychiatry 2020 Feb 25;14(1):37-43. [doi: 10.1111/eip.12804] [Medline: 30908854]
57. Maathuis I, Niezen M, Buitenweg D, Bongers IL, van Nieuwenhuizen C. Exploring human values in the design of a web-based QoL-instrument for people with mental health problems: a value sensitive design approach. Sci Eng Ethics 2020 Apr 09;26(2):871-898. [doi: 10.1007/s11948-019-00142-y] [Medline: 31598830]
58. Fonseka TM, Pong JT, Kcomt A, Kennedy SH, Parikh SV. Collaborating with individuals with lived experience to adapt CANMAT clinical depression guidelines into a patient treatment guide: the CHOICE-D co-design process. J Eval Clin Pract 2020 Aug 14;26(4):1259-1269 [FREE Full text] [doi: 10.1111/jep.13308] [Medline: 31729117]
59. Terlouw G, van 't Veer JT, Prins JT, Kuipers DA, Pierie JP. Design of a digital comic creator (It's Me) to facilitate social skills training for children with autism spectrum disorder: design research approach. JMIR Ment Health 2020 Jul 10;7(7):e17260 [FREE Full text] [doi: 10.2196/17260] [Medline: 32673273]
60. Stawarz K, Preist C, Tallon D, Wiles N, Kessler D, Turner K, et al. Design considerations for the integrated delivery of cognitive behavioral therapy for depression: user-centered design study. JMIR Ment Health 2020 Sep 03;7(9):e15972 [FREE Full text] [doi: 10.2196/15972] [Medline: 32880580]
61. Callan JA, Dunbar Jacob J, Siegle GJ, Dey A, Thase ME, DeVito Dabbs A, et al. CBT MobileWork©: user-centered development and testing of a mobile mental health application for depression. Cogn Ther Res 2020 Sep 25;45(2):287-302. [doi: 10.1007/s10608-020-10159-4]
62. Flobak E, Nordby ES, Guribye F, Kenter R, Nordgreen T, Lundervold AJ. Designing videos with and for adults with ADHD for an online intervention: participatory design study and thematic analysis of evaluation. JMIR Ment Health 2021 Sep 14;8(9):e30292 [FREE Full text] [doi: 10.2196/30292] [Medline: 34519666]
63. van der Meer L, Jonker T, Wadman H, Wunderink C, van Weeghel J, Pijnenborg GH, et al. Targeting personal recovery of people with complex mental health needs: the development of a psychosocial intervention through user-centered design. Front Psychiatry 2021 Apr 8;12:635514 [FREE Full text] [doi: 10.3389/fpsyt.2021.635514] [Medline: 33897494]
64. García AS, Fernández-Sotos P, Vicente-Querol MA, Sánchez-Reolid R, Rodriguez-Jimenez R, Fernández-Caballero A. Co-design of avatars to embody auditory hallucinations of patients with schizophrenia. Virtual Real 2021 Jul 08;27(1):217-232. [doi: 10.1007/s10055-021-00558-7]
65. Jonathan GK, Dopke CA, Michaels T, Bank A, Martin CR, Adhikari K, et al. A smartphone-based self-management intervention for bipolar disorder (LiveWell): user-centered development approach. JMIR Ment Health 2021 Apr 12;8(4):e20424 [FREE Full text] [doi: 10.2196/20424] [Medline: 33843607]
66. Milton AC, Hambleton A, Dowling M, Roberts AE, Davenport T, Hickie I. Technology-enabled reform in a nontraditional mental health service for eating disorders: participatory design study. J Med Internet Res 2021 Feb 16;23(2):e19532 [FREE Full text] [doi: 10.2196/19532] [Medline: 33591283]

67. Sanin J, Spong L, McRae C. Creative wellbeing. Prototyping an arts-health practice program for mental health recovery. Design Health 2021 May 20;5(1):61-81. [doi: 10.1080/24735132.2021.1927504]
68. Kruzan KP, Meyerhoff J, Biernesser C, Goldstein T, Reddy M, Mohr DC. Centering lived experience in developing digital interventions for suicide and self-injurious behaviors: user-centered design approach. JMIR Ment Health 2021 Dec 24;8(12):e31367 [FREE Full text] [doi: 10.2196/31367] [Medline: 34951602]
69. Bos FM, von Klipstein L, Emerencia AC, Veermans E, Verhage T, Snippe E, et al. A web-based application for personalized ecological momentary assessment in psychiatric care: user-centered development of the PETRA application. JMIR Ment Health 2022 Aug 09;9(8):e36430 [FREE Full text] [doi: 10.2196/36430] [Medline: 35943762]
70. Bongers IL, Buitenweg DC, van Kuijk RE, van Nieuwenhuizen C. I need to know: using the CeHRes roadmap to develop a treatment feedback tool for youngsters with mental health problems. Int J Environ Res Public Health 2022 Aug 31;19(17):10834 [FREE Full text] [doi: 10.3390/ijerph191710834] [Medline: 36078564]
71. Wiberg AC, Ghaderi A, Danielsson HB, Safarzadeh K, Parling T, Carlbring P, et al. Internet-based cognitive behavior therapy for eating disorders - development and feasibility evaluation. Internet Interv 2022 Dec;30:100570 [FREE Full text] [doi: 10.1016/j.invent.2022.100570] [Medline: 36110307]
72. Illarregi ER, Alexiou K, DiMalta G, Zamenopoulos T. Is designing therapeutic? A case study exploring the experience of co-design and psychosis. Psychosis 2022 Apr 27:1-15. [doi: 10.1080/17522439.2022.2052450]
73. Jenness JL, Bhattacharya A, Kientz JA, Munson SA, Nagar RR. Lessons learned from designing an asynchronous remote community approach for behavioral activation intervention for teens. Behav Res Ther 2022 Apr;151:104065 [FREE Full text] [doi: 10.1016/j.brat.2022.104065] [Medline: 35248749]
74. Terlouw G, Kuipers D, Veldmeijer L, van 't Veer J, Prins J, Pierie JP. Boundary objects as dialogical learning accelerators for social change in design for health: systematic review. JMIR Hum Factors 2022 Feb 03;9(1):e31167 [FREE Full text] [doi: 10.2196/31167] [Medline: 35113023]
75. Groeneveld B, Dekkers T, Boon B, D'Olivo P. Challenges for design researchers in healthcare. Design Health 2019 Jan 12;2(2):305-326. [doi: 10.1080/24735132.2018.1541699]
76. Blandford A, Berndt E, Catchpole K, Furniss D, Mayer A, Mentis H, et al. Strategies for conducting situated studies of technology use in hospitals. Cogn Tech Work 2014 Dec 16;17(4):489-502. [doi: 10.1007/s10111-014-0318-7]
77. Noël G. Health design: mapping current situations, envisioning next steps. Design J 2017 Sep 06;20(sup1):S2304-S2314. [doi: 10.1080/14606925.2017.1352746]
78. Tsekleves E, Cooper R. Emerging trends and the way forward in design in healthcare: an expert's perspective. Design J 2017 Sep 06;20(sup1):S2258-S2272. [doi: 10.1080/14606925.2017.1352742]
79. Rowe A, Knox M, Harvey G. Re-thinking health through design: collaborations in research, education and practice. Design Health 2020 Nov 15;4(3):327-344. [doi: 10.1080/24735132.2020.1841918]
80. Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Serv Res 2022 Jul 07;22(1):877 [FREE Full text] [doi: 10.1186/s12913-022-08079-y] [Medline: 35799251]
81. Mulvale A, Miatello A, Hackett C, Mulvale G. Applying experience-based co-design with vulnerable populations: lessons from a systematic review of methods to involve patients, families and service providers in child and youth mental health service improvement. Patient Exp J 2016 Apr 28;3(1):117-129. [doi: 10.35680/2372-0247.1104]
82. Schouten SE, Kip H, Dekkers T, Deenik J, Beerlage-de Jong N, Ludden GD, et al. Best-practices for co-design processes involving people with severe mental illness for eMental health interventions: a qualitative multi-method approach. Design Health 2022 Nov 26;6(3):316-344. [doi: 10.1080/24735132.2022.2145814]
83. Aryana B, Brewster L, Nocera JA. Design for mobile mental health: an exploratory review. Health Technol 2018 Oct 27;9(4):401-424. [doi: 10.1007/s12553-018-0271-1]
84. Kuipers DA, Terlouw G, Wartena BO, van 't Veer JT, Prins JT, Pierie JP. The role of transfer in designing games and simulations for health: systematic review. JMIR Serious Games 2017 Nov 24;5(4):e23 [FREE Full text] [doi: 10.2196/games.7880] [Medline: 29175812]
85. Veldmeijer L, Terlouw G, van 't Veer J, Kuipers D. Harnessing virtual reality simulation in training healthcare workers in handling patients with suspected COVID-19 infections: results of training and lessons learned about design. Design Health 2022 Apr 21;6(1):44-68. [doi: 10.1080/24735132.2022.2059997]
86. van Os J, Guloksuz S, Vijn TW, Hafkenscheid A, Delespaul P. The evidence-based group-level symptom-reduction model as the organizing principle for mental health care: time for change? World Psychiatry 2019 Feb 02;18(1):88-96 [FREE Full text] [doi: 10.1002/wps.20609] [Medline: 30600612]
87. Jones N, Callejas L, Brown M, Colder Carras M, Croft B, Pagdon S, et al. Barriers to meaningful participatory mental health services research and priority next steps: findings from a national survey. Psychiatr Serv 2023 Mar 20:appips20220514 [FREE Full text] [doi: 10.1176/appi.ps.20220514] [Medline: 36935620]
88. Davidson L. The recovery movement: implications for mental health care and enabling people to participate fully in life. Health Affairs 2016 Jun 01;35(6):1091-1097. [doi: 10.1377/hlthaff.2016.0153] [Medline: 27269027]

89. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect 2014 Oct;17(5):637-650 [FREE Full text] [doi: 10.1111/j.1369-7625.2012.00795.x] [Medline: 22809132]
90. Hamilton CB, Hoens AM, Backman CL, McKinnon AM, McQuitty S, English K, et al. An empirically based conceptual framework for fostering meaningful patient engagement in research. Health Expect 2018 Feb;21(1):396-406 [FREE Full text] [doi: 10.1111/hex.12635] [Medline: 28984405]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question

Edited by J Torous; submitted 17.02.23; peer-reviewed by A Segev, S Markham, N Dippel; comments to author 02.04.23; revised version received 10.05.23; accepted 16.06.23; published 25.07.23.

Please cite as:
Veldmeijer L, Terlouw G, Van Os J, Van Dijk O, Van 't Veer J, Boonstra N
The Involvement of Service Users and People With Lived Experience in Mental Health Care Innovation Through Design: Systematic Review
JMIR Ment Health 2023;10:e46590
URL: https://mental.jmir.org/2023/1/e46590
doi:10.2196/46590
PMID:37490326

Review

Long-Term Efficacy of Internet-Based Cognitive Behavioral Therapy Self-Help Programs for Adults With Depression: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Megi Mamukashvili-Delau[1,2], MSc; Nicole Koburger[3], Dr; Sandra Dietrich[4], Dr; Christine Rummel-Kluge[1,2], Prof Dr med

[1]Department of Psychiatry and Psychotherapy, Medical Faculty, Leipzig University, Leipzig, Germany

[2]Department of Psychiatry and Psychotherapy, Universitätsklinikum Leipzig, Leipzig, Germany

[3]Department of Research and Transfer, Leipzig University, Leipzig, Germany

[4]Leipzig Travel, Leipzig Tourismus und Marketing GmbH, Leipzig, Germany

Corresponding Author:
Christine Rummel-Kluge, Prof Dr med
Department of Psychiatry and Psychotherapy
Medical Faculty
Leipzig University
Klinik und Poliklinik für Psychiatrie und Psychotherapie Semmelweisstraße 10, Haus 13
Leipzig, 04103
Germany
Phone: 49 0341 97 24464
Fax: 49 0341 97 24509
Email: Christine.Rummel-Kluge@medizin.uni-leipzig.de

Abstract

Background: Depression is a worldwide mental disorder and a leading cause of disability. Many people with depression do not want to take medication or have the motivation to seek psychotherapy treatment for many reasons. Guided internet-based self-help programs may be a promising solution for addressing these issues. This kind of intervention has proven to be effective in reducing depression symptoms on a short-term scale. However, as treatment often is a long-term rehabilitation process, it is important to examine not only the short-term effects of internet-based cognitive behavioral therapy (iCBT) self-help treatment but also the follow-up or long-term efficacy of this kind of intervention.

Objective: This systematic review and meta-analysis aimed to identify studies that examined follow-up data ≥8 weeks after posttreatment measurements and thereby examined the long-term efficacy of iCBT self-help programs with minimal weekly guidance for people with depression. It aimed to analyze the long-term efficacy of iCBT treatments compared to control conditions as well as long-term efficacy within the iCBT treatment conditions. Additionally, it aimed to conduct subgroup analyses according to the follow-up time points for each outcome. Finally, it examined long-term improvements in quality of life.

Methods: The Cochrane Collaboration Depression, Anxiety, and Neurosis Controlled Trials Register (CCDANCTR), grey literature, reference lists, and correspondence were used to search for published and unpublished randomized controlled trials (RCTs) that reported the long-term or follow-up efficacy of computer-based or iCBT self-help treatments for depression with minimal guidance of up to 10 min/wk. The search took place between 2015 and 2022 (October).

Results: The search resulted in a total of 2809 study abstracts, of which 15 studies (with 17 samples) met all inclusion criteria and were included in the long-term analysis. The results showed that the depression outcomes of all follow-up time points together in the treatment conditions were favored over the control conditions with a medium effect size of 0.43 (n=1689 participants; 9 RCTs; standardized mean difference [SMD] –0.43, 95% CI –0.67 to –0.20; *P*<.001). The analysis of long-term efficacy within the iCBT treatment conditions showed that the follow-up outcomes of the treatment groups were favored over the posttreatment outcomes with a small effect size of 0.20 (n=2196 participants; 17 RCTs; SMD 0.20, 95% CI 0.07-0.49; *P*=.003). Findings for improving quality of life also showed that the iCBT conditions were favored over the control conditions with a small effect size of 0.19 (n=1345 participants; 3 RCTs; SMD 0.19, 95% CI 0.08-0.30; *P*<.001).

Conclusions: This systematic review and meta-analysis found that iCBT self-help interventions had a superior long-term efficacy for individuals with depressive symptoms compared to control groups. The within-group analysis of iCBT treatment conditions also showed statistically significant improvements in reducing depressive symptoms at follow-up compared to posttreatment measurements.

(JMIR Ment Health 2023;10:e46925) doi:10.2196/46925

KEYWORDS

depression; internet-based cognitive behavioral therapy; iCBT; self-help; minimal guidance; long-term; follow-up; mental health; psychotherapy; cognitive behavioral therapy; CBT; systematic review; meta-analysis; meta-analyses; review method

Introduction

Depression is a worldwide mental disorder and one of the leading causes of disability. According to the World Health Organization (WHO), the incidence of mental disorder conditions increased 13% within the last decade. Approximately 280 million people in the world have depression [1]. COVID-19 Mental Disorders Collaborators [2] estimated that there were approximately 53.2 million additional cases of major depression due to the COVID-19 pandemic across the world. This suggests that the recent pandemic situation increased the urgency for various accessible depression treatments.

Psychotherapy and pharmacotherapy are effective ways to treat depressive disorders [3]. Despite the availability of these evidence-based treatment options, only nearly half of the people with depression receive suitable treatment [4]. Many people with depression are hesitant to take medication or show poor adherence after having taken medication and experiencing side effects [5]. Furthermore, a large amount of individuals with depression do not have the motivation to seek psychotherapy treatment for many reasons, such as perceived stigma, the unavailability of psychotherapists including long waiting lists for the beginning of treatment, probable prohibitive costs, or geographic distance [6,7].

Internet-based self-help programs may be a promising solution for addressing these issues [4]. They offer the people with depression brief and structured therapy with or without any contact with therapists. It can be received at home and is relatively anonymous. It might help avoid stigma and can be used according to the patient's own schedules and needs.

Moreover, web-based self-help treatments can help the individuals with depression develop usable skills to identify and monitor problematic thoughts and emotions and cope with them [8]. During the internet-based cognitive behavioral therapy (iCBT) self-help treatment, the severity of mild to moderate depressive symptoms may improve, or the waiting period until clinical or face-to-face treatment is available may be bridged.

A growing number of randomized controlled trials (RCTs) [9-13] and meta-analyses [4,14-17] are reporting about the effectiveness of computer-based or iCBT self-help treatments for people with depression.

Furthermore, iCBT self-help treatment can be used as a stand-alone intervention as well as with different levels of support, which can be implemented in different forms, such as brief phone calls, short text messages, emails, or postcards [18]. Several studies are reporting a higher efficacy of guided self-help interventions compared to unguided ones [18-23].

As therapy for depression in general usually requires a long-term rehabilitation process, it is important to study not only the short-term effects of iCBT treatments but also the follow-up or long-term efficacy of this kind of self-help intervention.

Although there are some studies or meta-analyses that studied the effectiveness of iCBT on reducing depressive symptoms at follow-up, the results are inconsistent. Some studies did not find any significant effects of iCBT at follow-up [24,25], whereas other studies [11,26,27] and meta-analyses [14,15] reported significant effects of iCBT treatment over the control group in reducing depression symptoms at follow-up. Andersson et al [28] even reported about a tendency for the guided iCBT group to be superior to group-based cognitive behavioral therapy at 3-year follow-up.

Despite the large amount of studies examining the effectiveness of iCBT self-help treatments at the posttreatment or follow-up stage, there is a lack of meta-analyses analyzing the long-term efficacy of iCBT with a weekly minimal guidance up to 10 minutes during the treatment period.

Therefore, this systematic review and meta-analysis aimed to identify studies that examined the follow-up or long-term efficacy of such iCBT self-help programs with minimal weekly guidance. It aimed to analyze the long-term efficacy of iCBT treatments compared to control conditions as well as the long-term efficacy within the iCBT treatment conditions. Additionally, this review and meta-analysis aimed to conduct subgroup analyses according to the follow-up time points for each outcome. Finally, it examined long-term improvements in quality of life for the participants who are randomized to iCBT self-help interventions compared to control conditions.

Methods

Overview

The methods of this long-term meta-analysis refer to an original study published in 2022 [29]; the design and outcomes of posttreatment (short-term) efficacy of iCBT self-help programs for depression with weekly minimal guidance are described there in detail.

The aim of this meta-analysis was to summarize the long-term depression outcomes of the studies that were included in the previous original meta-analysis [29]. We reported the follow-up data, measured at least 8 weeks after posttreatment measurements, and thereby examined the long-term efficacy of

iCBT self-help interventions with a weekly minimal guidance (up to 10 minutes) compared to the control conditions of patients who did not receive any treatment before the time point of follow-up measurements. Furthermore, we analyzed the efficacy within the iCBT intervention conditions. Lastly, we analyzed long-term improvements in quality of life for treatment conditions compared to control groups.

Search Methodology and Study Selection

To identify relevant studies, we searched the Cochrane Collaboration Depression, Anxiety, and Neurosis Controlled Trials Register (CCDANCTR), which contains the searches of MEDLINE (1950 to present), Embase (1974 to present), and PsycINFO (1967 to present); quarterly searches of the Cochrane Central Register of Controlled Trials (CENTRAL); and review-specific searches of additional databases. We also searched international trial registries via the WHO's trials portal (International Clinical Trials Registry Platform [ICTRP]) and ClinicalTrials.gov to identify unpublished or ongoing studies. We searched sources of grey literature, including dissertations and theses, clinical guidelines, and reports from regulatory agencies (where appropriate). We checked the reference lists of all included studies and relevant systematic reviews to identify additional studies missed from the original electronic searches. We also conducted a cited reference search on the Web of Science.

We did not impose any restriction on date, language, or publication status to the searches.

The selection criteria for studies are shown in Textbox 1.

Textbox 1. Selection criteria.

Types of studies

- Published or unpublished randomized controlled trials, as well as crossover trials

Diagnosis

- Studies using one of the following depression questionnaires were accepted: Patient Health Questionnaire (PHQ) [30], Beck Depression Inventory (BDI) [31,32], Hamilton Depression Rating Scale (HDRS) [33], Montgomery Depression Scale (MADRS) [34], The Center for Epidemiologic Studies Depression Scale (CES-D) [35], Hospital Anxiety and Depression Scales (HADS) [36], Kessler Psychological Distress Scale (K-10) [37], Depression Anxiety Stress Scales (DASS) [38], or any other validated depression scale.
- If studies reported more than one type of depression outcome measure, those outcomes were extracted with the highest priority according to the following list: (1) PHQ-9, (2) BDI-II, (3) HDRS, (4) MADRS, (5) CES-D, (6) HADS, and (7) others.

Types of interventions

- Studies with experimental internet-based cognitive behavioral therapy (iCBT) self-help programs with weekly minimal guidance (ie, up to 10 minutes) given by a mental health professional or a therapist
- Eligible control comparisons: treatment as usual, waiting list or delayed treatment condition, not active control condition, attention placebo, and psychological placebo

Types of participants

- Participants from any racial or ethnic groups aged ≥14 years with depression (ie, measured with a validated depression questionnaire) were included.
- For the long-term analyses, only participants from the intervention conditions that completed the follow-up measurements were eligible, as well as the participants from control conditions that had not received any iCBT self-help program until the follow-up stage.

Setting

- Studies conducted in community, primary, secondary, or tertiary services were all eligible for inclusion.

Types of outcome measures

- Primary outcome

1. Long-term efficacy of iCBT with weekly minimal guidance (up to 10 minutes): changes in depressive symptomatology at the follow-up stage (treatment group compared with control group, where the participants did not receive any iCBT treatment before the follow-up measurements).

1.1. Subgroup meta-analysis: changes in depression outcomes compared by the time point of follow-up measurements, such as (1) follow-up assessed <6 months after posttreatment measurements, (2) follow-up assessed between 6-8 months after posttreatment measurements, or (3) follow-up assessed >8 months after posttreatment measurements

2. Long-term efficacy within the iCBT treatment conditions: changes of depression symptomatology at follow-up compared with posttreatment outcomes

2.1. Subgroup meta-analysis within the iCBT treatment conditions: changes in depression outcomes analyzed by the time point of follow-up measurements, such as (1) follow-up assessed <4 months after posttreatment measurements, (2) follow-up assessed between 4-7 months after posttreatment measurements, or (3) follow-up assessed >7 months after posttreatment measurements

- Secondary outcome

3. Improvements in quality of life at the follow-up stage, assessed with the use of validated measures

Data Collection and Analysis

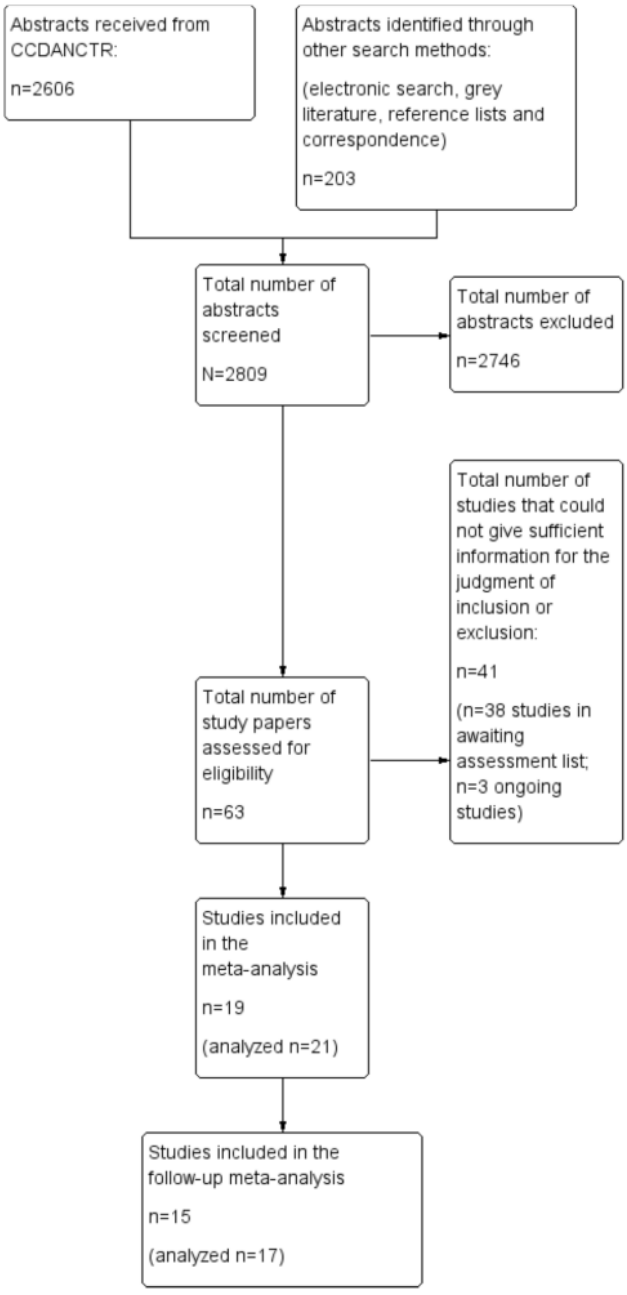
The search took place between 2015 and 2022 (October). The CCDANCTR yielded the abstracts in 2015 and 2018 (to update data). The last update was carried out in 2022. Four independent researchers were involved in the literature search and analysis.

The search resulted in a total of 2809 study abstracts from the CCDANCTR, electronic searches, cross-reference searches, and grey literature. A total of 2756 studies were excluded because they did not meet 1 or more inclusion criteria, and 38 studies could not be included or excluded because of a lack of required information described in the original studies or because there was no publication available to decide on inclusion or exclusion. These study authors were contacted a few times during the process of the meta-analysis. Either there was no response from them, or they could not provide sufficient information for making the decision to exclude or include these studies. Therefore, they are still in the awaiting assessment list. There were also 3 ongoing studies [39-41] that were in process at the time of conducting this meta-analysis.

Finally, 19 studies [6,9,11,27,42-56] met all inclusion criteria, but 4 studies [6,45,51,55] reported only posttreatment measurements and no follow-up measurements. Thus, they could not be included in the long-term analysis.

The results of 2 included studies [42,56] could be used as 2 separate samples due to their 3-arm design. Therefore, in this meta-analysis, a total of 15 studies (with 17 samples) were included. Figure 1 outlines the search process.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart outlining the process of the meta-analysis. CCDANCTR: The Cochrane Depression, Anxiety, and Neurosis Controlled Trials Register.



We used the latest version of Review Manager (RevMan; version 5.4.1; Cochrane Collaboration) software [57], to extract the characteristics of the included studies, such as the number of participants and means and SDs of outcomes at the posttreatment and follow-up stages.

The quality of the individual studies was assessed with the *Cochrane tool for assessing risk of bias* (Cochrane Collaboration) [58]. The studies were assessed using 7 categories of risk of bias: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other bias. Each category was rated as low, high, or unclear.

Statistical Analysis

The continuous outcomes such as mean differences (MDs) were pooled into standardized MDs (SMDs) as different questionnaires were used to assess the severity of depression symptomatology in the original studies. We used 95% Cls [58,59].

Furthermore, we tested statistical heterogeneity between studies using a standard chi-square test. We examined the I^2 value using

the following overlapping bands provided in the *Cochrane Handbook for Systematic Reviews of Interventions* [58,59]: 0% to 40%=might not be important; 30% to 60%=may represent moderate heterogeneity; 50% to 90%=may represent substantial heterogeneity; and 75% to 100%=may represent considerable heterogeneity.

Data synthesis

As studies were estimating different treatment effects, we used the random-effects model of meta-analysis.

Ethical Considerations

All included studies reported having ethics approval. The participants in the original studies provided written informed consent to participate in the studies.

Results

Characteristics of the Included Studies

The final research yielded 15 relevant studies (17 samples) [9,11,27,42-44,46-50,52-54,56] for the long-term analysis. The participants in all included studies were randomized in 2 or more groups, where at least one group was an iCBT self-help intervention group with minimal weekly guidance (up to 10 minutes) and the other group was either a control or waitlist group, treatment as usual (TAU), or a not active control condition.

In this long-term meta-analysis, all participants from the intervention conditions who completed the follow-up measurements were included. From the control conditions, only the participants who had not received any treatment until they completed the follow-up measurements were included.

A total of 3226 participants were included in the posttreatment meta-analysis [29]. From this number, for this study, 1280 participants were excluded either from the treatment condition due to not completing the follow-up measurements or from the control condition due to receiving a self-help program after posttreatment measurements. Thus, a total of 1946 participants were included in the long-term meta-analysis.

Measurements of depression symptoms in treatment conditions—as well as in control conditions—were followed up in 9 samples (7 studies) [11,27,42,48,49,52,56]. Another 8 studies [9,43,44,46,47,50,53,54] reported either the follow-up outcomes only for treatment conditions or, if they reported the follow-up outcomes for control conditions, we could not use them due to the participants receiving self-help treatment before the follow-up measurements.

Depression scores were followed up at 12 months after posttreatment measurements only in 2 samples (1 study) [42], and 8-month follow-up depression outcomes were also reported in 1 study [48]. Participants in 4 studies [9,11,27,43] were followed up after 6 months. Two studies [46,50,52] measured follow-up depression scores approximately 4 months after posttreatment measurements. Another 4 studies [44,47,53,54] assessed 3-month follow-up depression scores. Three samples (2 studies) [49,56] measured depression outcomes after 2 or 2.5 months.

Table 1 provides a detailed overview of the follow-up outcomes in each included study, as well as the time point of follow-up measurements (in weeks) and type of control conditions as described in the original studies.

Table 1. Detailed characteristics of included studies in the long-term analysis.

Study author, (year)	Follow-up outcomes for the intervention group, mean (SD)	Follow-up outcomes for the control group (without any intervention until follow-up measurements), mean (SD)	Time point of follow-up measurements (weeks)	Number of participants included in the follow-up analysis, n[a]	Follow-up depression outcome	Type of control condition
Andersson et al [43], 2005	13.1 (9.1)	—[b]	24	36	BDI-II[c]	Waitlist (at the follow-up assessment, the control group had completed the internet program)
Berger et al [9], 2011	16.24 (11.4)	—	24	25	BDI-II	Waitlist (at the follow-up assessment, the control group had completed the unguided self-help program)
Choi et al [44], 2012	5.68 (5.39)	—	12	21	PHQ-9[d]	Waitlist (follow-up assessment for the control group was not reported)
Farrer et al [27], 2011	18.4 (10.4)	34.2 (13.5)	24	40	CES-D[e]	TAU[f]
Gilbody et al [42], 2015	8.13 (6.13)	8.45 (6.28)	48	331	PHQ-9	Usual GP[g] care
Gilbody et al [42], 2015a	7.39 (5.51)	8.45 (6.28)	48	348	PHQ-9	Usual GP care
Klein et al [11], 2016	8.05 (4.20)	9.52 (4.34)	24	636	PHQ-9	TAU alone
Mantani et al [52], 2017	8.92 (6.00)	8.85 (5.93)	17	117	PHQ-9	Switch alone arm
Mohr et al [46], 2013	5.52 (4.45)	—	16	25	PHQ-9	Waitlist (at the follow-up assessment, the control group had completed the coached or self-directed moodManager)
Newby et al [47], 2013	4.05 (3.79)	—	12	40	PHQ-9	Waitlist (participants in the waitlist condition commenced iCBT[h] immediately after posttreatment assessments)
Newby et al [53], 2017	10.98 (4.49)	—	12	19	PHQ-9	TAU (participants in TAU gained access to the iCBT program after posttreatment measurements)
Proudfoot et al [48], 2004	9.3 (8.5)	14.9 (11.3)	32	186	BDI-II	TAU
Selmi et al [49], 1991	6.17 (5.57)	20.67 (9.89)	8	24	BDI-II	Waitlist (the waitlist control group received the treatment after follow-up measurements)
Smith et al [54], 2017	9.41 (4.71)	—	12	30	PHQ-9	Waitlist (participants in this group had the choice of enrolling in either iCBT or 1 of the 2 self-help books after posttreatment measurements)
Stiles-Shields et al [56], 2019	8.9 (5.88)	11.5 (4.25)	10	20	PHQ-9	Waitlist (the control group began treatment after follow-up measurements)
Stiles-Shields et al [56], 2019a	5.29 (4.46)	11.5 (4.25)	10	18	PHQ-9	Waitlist (the control group began treatment after follow-up measurements)
Titov et al [50], 2010	6.49 (3.94)	—	16	30	PHQ-9	Waitlist (the control group began treatment after the intervention group completed posttreatment assessments)

[a]Total: n=1946.

[b]Not available.

[c]BDI-II: Beck Depression Inventory II.

[d]PHQ-9: Patient Health Questionnaire 9.

[e]CES-D: The Center for Epidemiologic Studies Depression Scale.

[f]TAU: treatment as usual.

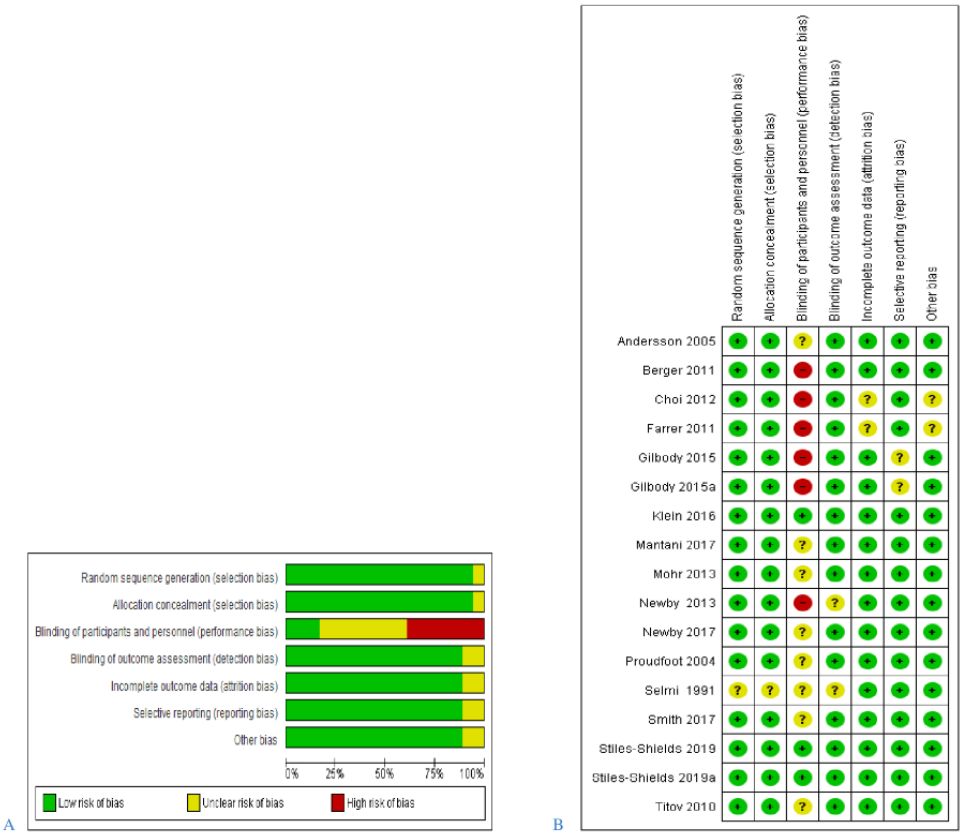
[g]GP: general practitioner.

[h]iCBT: internet-based cognitive behavioral therapy.

Quality of the Included Studies

Figure 2 provides detailed judgments about each risk-of-bias item presented in percentages across all included studies. If there was no sign of bias, it was assessed as "low risk of bias." If the original study authors did not report sufficient information to judge existing bias, it was assessed as "unclear." If we had suspicion of real existing bias, it was assessed as "high risk of bias."

Figure 2. (A) Risk of bias graph: judgements about each risk of bias item presented as percentages. (B) Summary of risk of bias identified for each included study [9,11,27,42-44,46-50,52-54,56].



The participants of all included studies were randomized, and all studies except one [49] described their randomization method in detail. The process of random allocation sequence was circumstantially described in all of the included studies except one [49]. Therefore, there was a low risk of selection bias.

If participants as well as personnel were blinded, the risk of performance bias was assessed as low. If personnel were not blinded, it was assessed as high risk. If there was not sufficient information about the blinding of personnel, performance bias was assessed as unclear. In most of the included studies except 2 studies (3 samples) [11,56], this kind of bias was assessed as high or unclear.

There was a low risk of detection bias in most of the included studies except 2 studies [47,49]. In these 2 studies, there was not enough information provided to permit judgment about the blinding of outcome assessment.

Two studies [27,44] did not report sufficient information to judge the risk of attrition bias. The remaining studies had a low risk of incomplete outcome data.

All included studies except 1 study (2 samples) [42] reported all predefined outcomes. Therefore, there was a low risk of reporting bias.

Finally, there was no sign of high risk of other sources of bias. Two studies [27,44] did not report sufficient information about other bias.

In total, the risk of bias of all included studies could be assessed as "low to moderate," except for performance bias, which could

be assessed as "moderate to high" due to the lack of blinding of participants and personnel in the original studies.

Test of Heterogeneity

We chose the random-effects model to interpret the results of the long-term meta-analysis. The heterogeneity of the effect size samples was automatically tested in RevMan with I^2 values for the first primary outcome.

The results of the heterogeneity test for iCBT treatment efficacy at the follow-up time points showed substantial or considerable heterogeneity (I^2=75%; P<.001).

Primary Outcomes

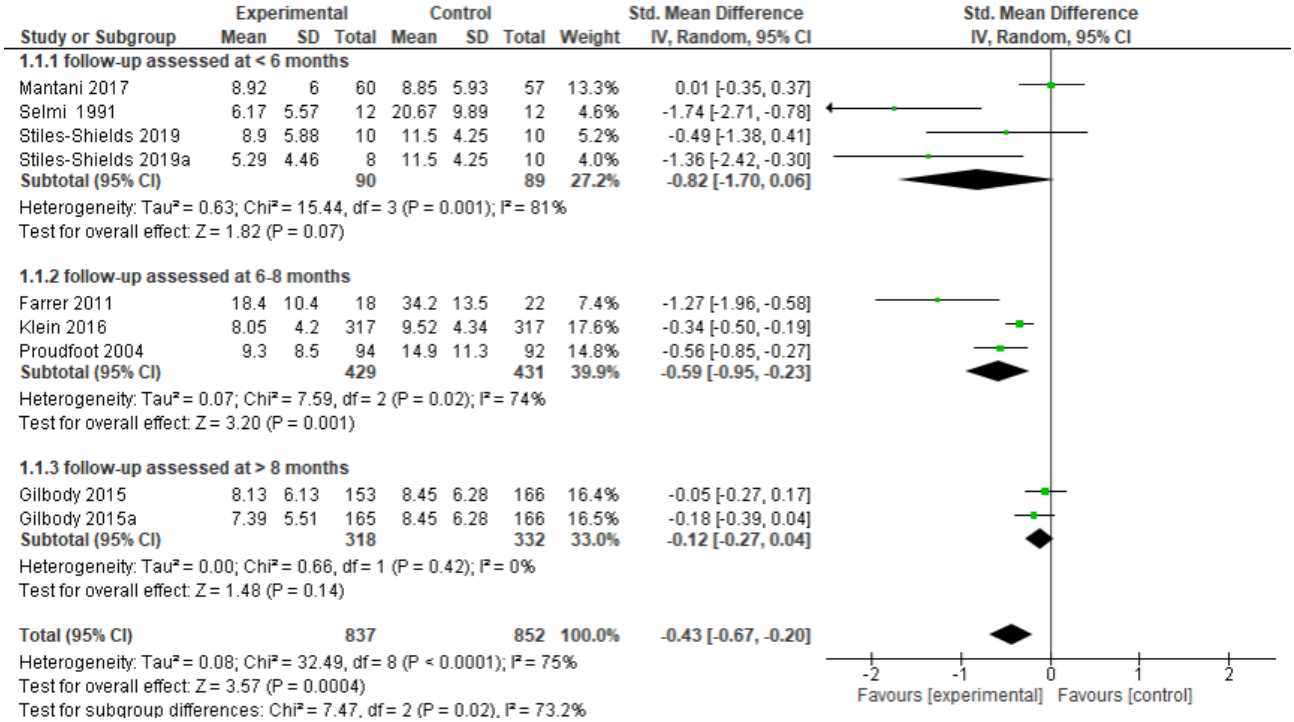
1. Long-Term Efficacy (iCBT Compared to Control Condition)

A total of 7 studies [11,27,42,48,49,52,56] assessed the follow-up outcomes in both conditions: for iCBT interventions as well as for control conditions. For 2 studies [42,56], 2 separate samples were usable; therefore, 9 samples were analyzed with a total 1689 participants using follow-up end point scores of depression symptoms.

The outcomes of long-term efficacy of iCBT self-help programs were assessed in various depression scales: Patient Health Questionnaire 9 (PHQ-9) [30], Beck Depression Inventory II (BDI-II) [31], and The Center for Epidemiologic Studies Depression Scale (CES-D) [35]. Therefore, we had to pool MDs into SMDs.

The long-term analysis of all depression scales and all follow-up time points together showed statistically significant differences between iCBT self-help treatment groups and control conditions that included participants who did not receive any treatment until follow-up measurements. Namely, the follow-up outcomes in the treatment conditions were favored over the control conditions with a medium effect size of 0.43 (n=1689 participants; 9 RCTs; SMD –0.43, 95% CI –0.67 to –0.20; Z=3.57, P<.001; I^2=75%, P<.001; see Figure 3).

Figure 3. Forest plot of standardized mean difference (95% CI) in change of depressive symptoms for intervention and control conditions at follow-up [11,27,42,48,49,52,56].



1.1. Subgroup Analysis (iCBT Compared to Control Condition)

The analysis of depression outcomes, using the subgroups of different follow-up stages (followed up at <6 months, 6-8 months, or >8 months), showed that the iCBT conditions were favored over the control conditions.

The participants in the iCBT treatment groups who were followed up <6 months after posttreatment measurements improved their depression symptoms with a large effect size of 0.82. However, this improvement was not statistically significant (n=179 participants; 4 RCTs; SMD –0.82, 95% CI –1.70 to 0.06; Z=1.82, P=.07; I^2=81%, P<.001; see Figure 3).

The depression outcomes followed up between 6-8 months after treatment also showed statistically significant improvements in depression symptoms in iCBT self-help groups with a moderate effect size of 0.59 (n=860 participants; 3 RCTs; SMD –0.59, 95% CI –0.95 to –0.23; Z=3.20, P=.001; I^2=74%, P=.02; see Figure 3).

The analysis of 2 samples, where the follow-up was assessed >8 months after posttreatment measurements, showed that the iCBT intervention conditions were favored over the control conditions in improving depression symptoms with a small effect of 0.12. However, this improvement was not statistically significant (n=650 participants; 2 RCTs; SMD 0.12, 95% CI –0.27 to 0.04; Z=1.48, P=.14; I^2=0%, P=.42; see Figure 3).

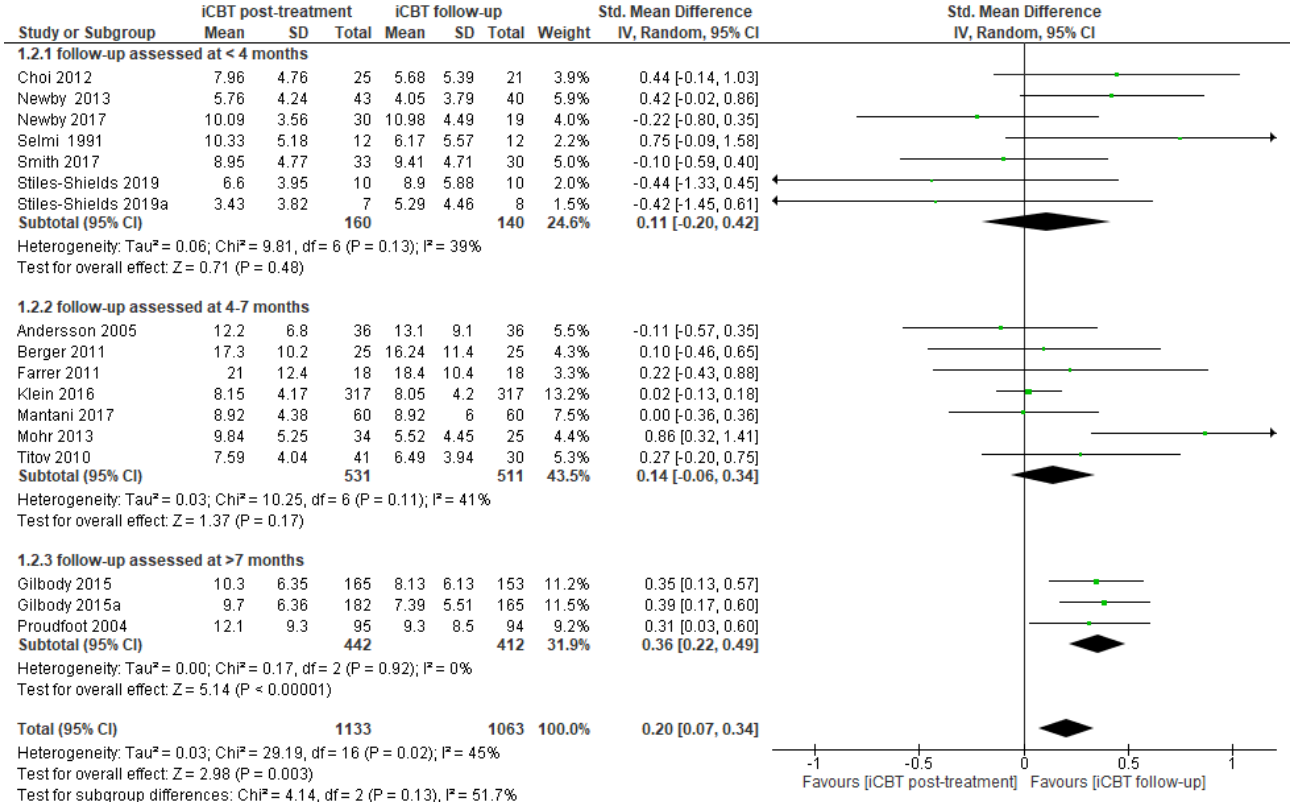
2. Long-Term Efficacy Within iCBT Treatment Conditions

The depressive symptoms of the participants in the iCBT self-help intervention groups were followed up in 17 samples (15 studies) [9,11,27,42-44,46-50,52-54,56].

Depression outcomes were assessed in an iCBT treatment condition among 1133 participants at the posttreatment stage. A total of 1063 (95.5%) out of 1113 participants in the intervention groups completed the follow-up measurements.

The results of this comparison between the depression outcomes at the posttreatment stage and follow-up stage showed statistically significant differences between these 2 time points. Namely, the participants in the iCBT self-help intervention groups continued to improve their depressive symptoms even a few months after they received self-help programs with minimal guidance. Specifically, the follow-up outcomes of the treatment groups were favored over the posttreatment outcomes with a small effect size of 0.20 (n=2196 participants; 17 RCTs; SMD 0.20, 95% CI 0.07-0.49; Z=2.98, P=.003; I^2=45%, P=.02; see Figure 4).

Figure 4. Forest plot of standardized mean difference (95% CI) in change of depressive symptoms for intervention conditions: posttreatment stage compared to follow-up stage [9,11,27,42-44,46-50,52-54,56]. iCBT: internet-based cognitive behavioral therapy.



2.1. Subgroup Analysis of iCBT Treatment Conditions

Moreover, we analyzed the depression outcomes by the time point of follow-up measurements, such as (1) follow-up assessed <4 months after posttreatment measurements, (2) follow-up assessed between 4-7 months after posttreatment measurements, or (3) follow-up assessed >7 months after posttreatment measurements.

The subgroup analysis showed improvements of depressive symptoms in the iCBT intervention groups that were followed up in <4 months after posttreatment measurements with a small effect size of 0.11. However, these changes were not statistically significant (n=300 participants; 7 RCTs; SMD 0.11, 95% CI –0.20 to 0.42; Z=0.71, P=.48; I^2=39%, P=.13; see Figure 4).

The participants of the iCBT intervention groups that were followed up between 4-7 months after posttreatment measurements also showed improvements in depressive symptoms at the follow-up stage with a small effect size of 0.14. However, these changes were not statistically significant (n=1042 participants; 7 RCTs; SMD 0.14, 95% CI –0.06 to 0.34; Z=1.37, P=.17; I^2=41%, P=.11; see Figure 4).

The analysis of the intervention groups where the participants were followed up >7 months after posttreatment measurements showed statistically significant improvements in depressive symptoms with a small effect size of 0.36 (n=854 participants; 3 RCTs; SMD 0.36, 95% CI 0.22-0.49; Z=5.14, P<.001; I^2=0%, P=.92; see Figure 4).

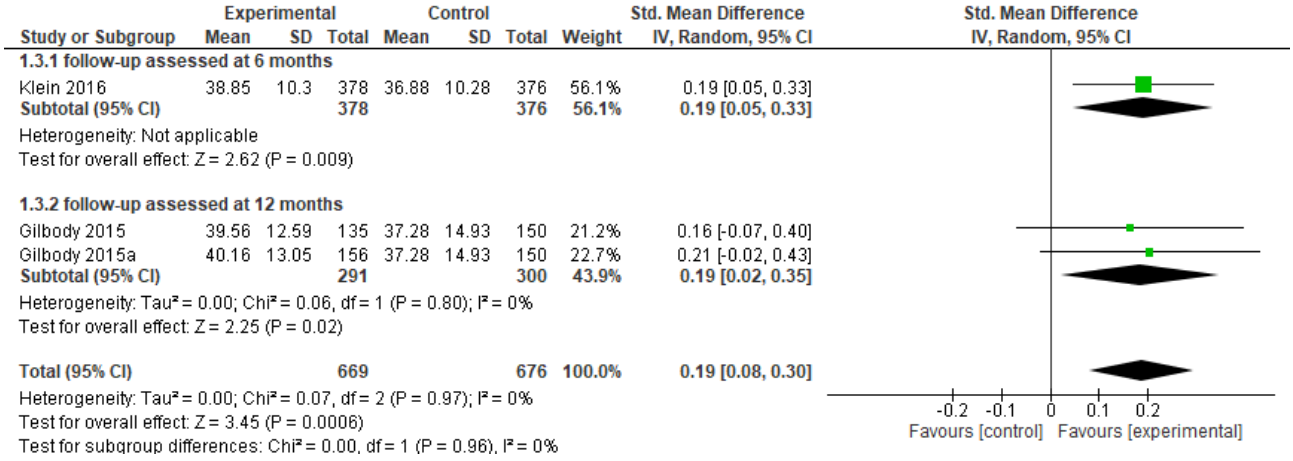
Secondary Outcome: 3. Improvements in Quality of Life at Follow-Up

A total of 3 samples (2 studies) [11,42] assessed improvements in quality of life at follow-up among 1345 participants in the intervention and control conditions together. Klein et al [11] used the Mental Composite Score of SF-12 [60] to assess improvements in quality of life. Gilbody et al [42] assessed the quality of life with the SF-36 [61]. Low scores in this outcome correspond to low improvements in quality of life.

The results showed statistically significant improvements for both follow-up time points: (1) follow-up assessed 6 months after posttreatment measurements and (2) follow-up assessed 12 months after posttreatment measurements.

Namely, the improvement in quality of life among the participants in the iCBT conditions were favored over the participants in the control conditions with a small effect size of 0.19, which is statistically significant (n=1345 participants; 3 RCTs; SMD 0.19, 95% CI 0.08-0.30; Z=3.45, P<.001; I^2=0%, P=.97; see Figure 5).

Figure 5. Forest plot of standardized mean difference (95% CI) in quality of life (low=poor) for intervention and control conditions at follow-up [11,42].



Discussion

Principal Findings

As depression therapy is a long-term rehabilitation process, it is important to examine not only short-term effects of iCBT self-help treatments but also the follow-up or long-term efficacy of this kind of intervention.

This systematic review and meta-analysis identified 17 samples (15 studies) that assessed the long-term efficacy of iCBT self-help interventions with minimal guidance (up to 10 min/wk) for depression at follow-up among 1946 participants.

Long-Term Efficacy of iCBT Compared to Control Condition

The analysis of long-term efficacy in 9 samples (among 1689 participants) revealed that the efficacy of iCBT interventions with a weekly minimal guidance showed statistically significant improvements in reducing depressive symptoms (with a medium effect size of 0.43) compared to the control conditions, where the participants did not receive any treatment until the follow-up measurements. This finding is well supported by a current meta-analysis, where Karyotaki et al [14] reported that guided iCBT self-help interventions reduced depressive symptoms compared with TAU at the 6-month follow-up. Zhou et al [62] reported similarly about the positive effect of iCBT interventions on reducing depression levels, which was significant at the <3-month follow-up.

We also conducted a subgroup analysis for this outcome according to the time point of follow-up measurements and found that the iCBT intervention groups were favored over the control condition in reducing depressive symptoms at every stage of follow-up (at <6 months, 6-8 months, or >8 months), although these improvements were statistically significant only at the stage of 6- to 8-month follow-up. Hence, the results showed the inconsistency of the significance level of the efficacy of iCBT in reducing depressive symptoms on the different stages of follow-up. These findings are supported by the findings of Zhou et al [62]. In addition, the results of this outcome indicate a high degree of heterogeneity (I^2=75%).

These findings suggest that iCBT self-help interventions for depression with minimal weekly guidance (up to 10 minutes) can be useful in reducing depressive symptoms not only at the posttreatment stage but also at the follow-up stage.

Long-Term Efficacy Within iCBT Treatment Conditions

A total of 17 samples (15 studies) with 1133 participants at the posttreatment time point and 1063 participants at the follow-up time point were included in the analysis of efficacy within the iCBT treatment conditions at follow-up.

Our previous meta-analysis [29] reported statistically significant improvements of depressive symptoms in iCBT treatment conditions compared to control conditions at the posttreatment stage. The long-term analysis of iCBT treatment conditions revealed that the participants who received iCBT self-help with weekly minimal guidance improved their depressive symptoms with a statistically significant, small effect size of 0.20 for all follow-up time points together compared to the outcomes at the posttreatment time point. This result is well supported by the findings of a previous meta-analysis [62] that reported about significant within-group effects of iCBT interventions on depression improvements at 3-month follow-up.

A subgroup analysis for this outcome according to the time point of follow-up measurements (ie, at <4 months, 4-7 months, or >7 months) within iCBT intervention groups also showed improvements in reducing depressive symptoms at every stage of follow-up. However, these improvements were statistically significant only at the follow-up time point of >7 months after posttreatment measurements.

Improvements in Quality of Life at Follow-Up

As the secondary outcome, we analyzed long-term improvements in quality of life within 3 samples with a total of 1345 participants in the iCBT intervention and control groups together.

The analysis showed small but statistically significant improvements in quality of life at the follow-up stage (at 6 and 12 months) in the participants of the intervention groups who received iCBT self-help treatments with weekly minimal guidance. This result approximates the findings of the recent meta-analysis by Han and Kim [15], who reported a small effect of internet-based intervention on improving quality of life compared to control groups at follow-up.

Strengths, Limitations, and Implications

Among the strength of this systematic meta-analysis is the clearly defined set of inclusion and exclusion criteria regarding participants, intervention, study design, outcomes, etc. [63].

Moreover, we examined the funnel plots for each outcome to assess the likely presence of publication bias. There was no evidence of possible asymmetry for either outcome.

In addition, we were able to conduct subgroup analyses and examined the efficacy of iCBT with minimal guidance at different stages of follow-up measurements.

Finally, the quality of the included studies was rather high, which allowed us to conclude that this meta-analysis is relatively free from critical bias. There was overall low risk of bias for all included studies. The quality of only 1 case—performance bias—was assessed as moderate to high. However, it is very difficult or sometimes even impossible to achieve total blinding of personnel and participants in such psychotherapeutic studies with minimal guidance.

This systematic review and meta-analysis has also several limitations that should be taken into consideration when interpreting the results.

First, one of the important limitations was the lack of follow-up data reported in the original studies. Only 9 out of 17 included samples assessed follow-up outcomes in both conditions: for iCBT interventions as well as for control conditions. Furthermore, relatively few studies examined improvements in quality of life.

Second, in the included studies, various iCBT programs with different number of sessions were used, which may report different effect sizes and can be a source of high heterogeneity between the included studies. Additionally, the inclusion of studies with a different types of control conditions (eg, waitlist, TAU, or usual general practitioner care), as well as the inclusion and comparation of studies with a different level of technological development (eg, an iCBT self-help program in 1991 compared to iCBT-based multimedia in 2019), could make it hard to interpret the results.

Furthermore, the different time points of follow-up measurements in the included studies may have a role in analyzing the long-term efficacy of iCBT with minimal guidance. Nonetheless, subgroup analysis was carried out to examine these differences.

Finally, this meta-analysis included only published outcomes of follow-up measurements. The potential for studies reporting small or null findings at the follow-up stage and not being published through either reluctance from authors or journal editors dismissing them may be a problem. Publication bias is, however, a problem for all researchers and not only for this meta-analysis.

Conclusions

In conclusion, this systematic review and meta-analysis found that iCBT self-help interventions with weekly minimal guidance of up to 10 minutes had superior long-term efficacy for individuals with depressive symptoms compared to control groups.

The within-group analysis of iCBT treatment conditions showed statistically significant improvements in reducing depressive symptoms at the follow-up stage compared to posttreatment measurements.

In addition, the analysis of improvements in quality of life at follow-up (at 6 and 12 months) showed statistically significant improvements in the participants that received iCBT self-help treatments compared to the control conditions.

However, the statistical significance of the long-term effectiveness of iCBT self-help programs for depression at various follow-up stages was inconsistent. Furthermore, it is important that future studies systematically examine the moderator factors at follow-up for this inconsistency, such as the number of previous depression episodes, severity of depression, symptom duration, etc.

Moreover, further research should be undertaken to develop practicable approaches to include iCBT interventions in health care systems, as it would help patients with mild to moderate depressive symptoms in reducing the severity of their depressive symptoms or to bridge the waiting period until they receive clinical or face-to-face treatment.

Data Availability

The data sets used and analyzed during this review and meta-analysis are available from the corresponding author upon reasonable request.

Authors' Contributions

MM-D, SD, NK, and CR-K developed the search strategy, selected which trials to include, and extracted data from the trials. MM-D and SD entered data into Review Manager (RevMan; Cochrane Collaboration) software. MM-D and CR-K carried out

the analysis, interpreted the analysis, drafted the final review, and kept the review up to date. MM-D wrote the main manuscript text. All authors reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. WHO. Depressive disorder (depression). World Health Organization. 2023 Mar 31. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2023-07-27]
2. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
3. Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry 2020 Feb;19(1):92-107 [FREE Full text] [doi: 10.1002/wps.20701] [Medline: 31922679]
4. Pang Y, Zhang X, Gao R, Xu L, Shen M, Shi H, et al. Efficacy of web-based self-management interventions for depressive symptoms: a meta-analysis of randomized controlled trials. BMC Psychiatry 2021 Aug 11;21(1):398 [FREE Full text] [doi: 10.1186/s12888-021-03396-8] [Medline: 34380440]
5. Arroll B, Chin W, Martis W, Goodyear-Smith F, Mount V, Kingsford D, et al. Antidepressants for treatment of depression in primary care: a systematic review and meta-analysis. J Prim Health Care 2016 Dec;8(4):325-334. [doi: 10.1071/HC16008] [Medline: 29530157]
6. Rosso IM, Killgore WDS, Olson EA, Webb CA, Fukunaga R, Auerbach RP, et al. Internet-based cognitive behavior therapy for major depressive disorder: a randomized controlled trial. Depress Anxiety 2017 Mar;34(3):236-245 [FREE Full text] [doi: 10.1002/da.22590] [Medline: 28009467]
7. Heinz I, Baldofski S, Beesdo-Baum K, Knappe S, Kohls E, Rummel-Kluge C. "Doctor, my back hurts and I cannot sleep." depression in primary care patients: reasons for consultation and perceived depression stigma. PLoS One 2021 Mar 5;16(3):e0248069 [FREE Full text] [doi: 10.1371/journal.pone.0248069] [Medline: 33667268]
8. Hanson K, Webb TL, Sheeran P, Turpin G. Attitudes and preferences towards self-help treatments for depression in comparison to psychotherapy and antidepressant medication. Behav Cogn Psychother 2016 Mar;44(2):129-139 [FREE Full text] [doi: 10.1017/S1352465815000041] [Medline: 25697236]
9. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011 Nov 7;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
10. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]
11. Klein JP, Berger T, Schröder J, Späth C, Meyer B, Caspar F, et al. Effects of a psychological internet intervention in the treatment of mild to moderate depressive symptoms: results of the EVIDENT study, a randomized controlled trial. Psychother Psychosom 2016 May 27;85(4):218-228 [FREE Full text] [doi: 10.1159/000445355] [Medline: 27230863]
12. Proudfoot J, Goldberg D, Mann A, Everitt B, Marks I, Gray JA. Computerized, interactive, multimedia cognitive-behavioural program for anxiety and depression in general practice. Psychol Med 2003 Feb;33(2):217-227. [doi: 10.1017/s0033291702007225] [Medline: 12622301]
13. Sethi S, Campbell AJ, Ellis LA. The use of computerized self-help packages to treat adolescent depression and anxiety. J Technol Hum Serv 2010 Aug 31;28(3):144-160. [doi: 10.1080/15228835.2010.508317]
14. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
15. Han A, Kim TH. Effects of internet-delivered behavioral activation on individuals with depressive symptoms: a systematic review and meta-analysis. J Psychiatr Res 2022 Aug;152:104-118 [FREE Full text] [doi: 10.1016/j.jpsychires.2022.05.031] [Medline: 35717866]
16. Sztein DM, Koransky CE, Fegan L, Himelhoch S. Efficacy of cognitive behavioural therapy delivered over the Internet for depressive symptoms: a systematic review and meta-analysis. J Telemed Telecare 2018 Sep;24(8):527-539. [doi: 10.1177/1357633X17717402] [Medline: 28696153]
17. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
18. Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? Expert Rev Pharmacoecon Outcomes Res 2007 Jun;7(3):291-297. [doi: 10.1586/14737167.7.3.291] [Medline: 20528315]

19. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
20. Levin W, Campbell DR, McGovern KB, Gau JM, Kosty DB, Seeley JR, et al. A computer-assisted depression intervention in primary care. Psychol Med 2011 Jul;41(7):1373-1383. [doi: 10.1017/S0033291710001935] [Medline: 20961474]
21. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
22. Karyotaki E, Furukawa TA, Efthimiou O, Riper H, Cuijpers P. Guided or self-guided internet-based cognitive-behavioural therapy (iCBT) for depression? study protocol of an individual participant data network meta-analysis. BMJ Open 2019 Jun 05;9(6):e026820 [FREE Full text] [doi: 10.1136/bmjopen-2018-026820] [Medline: 31171550]
23. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
24. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
25. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Umanodan R, et al. Effects of an internet-based cognitive behavioral therapy (iCBT) program in manga format on improving subthreshold depressive symptoms among healthy workers: a randomized controlled trial. PLoS One 2014 May 20;9(5):e97167 [FREE Full text] [doi: 10.1371/journal.pone.0097167] [Medline: 24844530]
26. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. J Med Internet Res 2012 Dec 31;15(1):e2 [FREE Full text] [doi: 10.2196/jmir.2240] [Medline: 23302475]
27. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Internet-based CBT for depression with and without telephone tracking in a national helpline: randomised controlled trial. PLoS One 2011 Nov 30;6(11):e28099 [FREE Full text] [doi: 10.1371/journal.pone.0028099] [Medline: 22140514]
28. Andersson G, Hesser H, Veilord A, Svedling L, Andersson F, Sleman O, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord 2013 Dec;151(3):986-994. [doi: 10.1016/j.jad.2013.08.022] [Medline: 24035673]
29. Mamukashvili-Delau M, Koburger N, Dietrich S, Rummel-Kluge C. Efficacy of computer- and/or internet-based cognitive-behavioral guided self-management for depression in adults: a systematic review and meta-analysis of randomized controlled trials. BMC Psychiatry 2022 Nov 24;22(1):730 [FREE Full text] [doi: 10.1186/s12888-022-04325-z] [Medline: 36424570]
30. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
31. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [doi: 10.1001/archpsyc.1961.01710120031004] [Medline: 13688369]
32. Powles WE. Beck, Aaron T. Depression: Causes and Treatment. Philadelphia: University of Pennsylvania Press, 1972. Pp. 370. $4.45. Am J Clin Hypn 1974 Apr;16(4):281-282. [doi: 10.1080/00029157.1974.10403697]
33. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23(1):56-62 [FREE Full text] [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]
34. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979 Apr;134:382-389. [doi: 10.1192/bjp.134.4.382] [Medline: 444788]
35. Radloff LS. The CES-D Scale. Applied Psychological Measurement 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
36. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
37. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]
38. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
39. Online Program to Reduce Depression in MS. ClinicalTrials.gov. 2020 Jul 29. URL: https://clinicaltrials.gov/show/nct02740361 [accessed 2016-01-01]
40. ClinicalTrials.gov. 2018 Sep 10. URL: https://clinicaltrials.gov/ct2/show/nct03188575 [accessed 2017-01-01]
41. Addressing Depression and Anxiety Symptoms in Patients With Inflammatory Bowel Disease. ClinicalTrials.gov. 2018 Apr 17. URL: https://clinicaltrials.gov/ct2/show/nct03327038 [accessed 2017-01-01]

42. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
43. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldo V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. Br J Psychiatry 2005 Nov;187:456-461. [doi: 10.1192/bjp.187.5.456] [Medline: 16260822]
44. Choi I, Zou J, Titov N, Dear BF, Li S, Johnston L, et al. Culturally attuned Internet treatment for depression amongst Chinese Australians: a randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]
45. Clarke G, Eubanks D, Reid E, Kelleher C, O'Connor E, DeBar LL, et al. Overcoming Depression on the Internet (ODIN) (2): a randomized trial of a self-help depression skills program with reminders. J Med Internet Res 2005 Jun 21;7(2):e16 [FREE Full text] [doi: 10.2196/jmir.7.2.e16] [Medline: 15998607]
46. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013 Aug 21;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
47. Newby JM, Mackenzie A, Williams AD, McIntyre K, Watts S, Wong N, et al. Internet cognitive behavioural therapy for mixed anxiety and depression: a randomized controlled trial and evidence of effectiveness in primary care. Psychol Med 2013 Dec;43(12):2635-2648. [doi: 10.1017/S0033291713000111] [Medline: 23419552]
48. Proudfoot J, Ryden C, Everitt B, Shapiro DA, Goldberg D, Mann A, et al. Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:46-54. [doi: 10.1192/bjp.185.1.46] [Medline: 15231555]
49. Selmi PM, Klein MH, Greist JH, Sorrell SP, Erdman HP. Computer-administered therapy for depression. MD Comput 1991 Mar;8(2):98-102. [Medline: 2038242]
50. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun 08;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
51. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008 Mar 25;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
52. Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. J Med Internet Res 2017 Nov 03;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
53. Newby J, Robins L, Wilhelm K, Smith J, Fletcher T, Gillis I, et al. Web-based cognitive behavior therapy for depression in people with diabetes mellitus: a randomized controlled trial. J Med Internet Res 2017 May 15;19(5):e157 [FREE Full text] [doi: 10.2196/jmir.7274] [Medline: 28506956]
54. Smith J, Newby JM, Burston N, Murphy MJ, Michael S, Mackenzie A, et al. Help from home for depression: a randomised controlled trial comparing internet-delivered cognitive behaviour therapy with bibliotherapy for depression. Internet Interv 2017 Sep;9:25-37 [FREE Full text] [doi: 10.1016/j.invent.2017.05.001] [Medline: 30135834]
55. Lambert JD, Greaves CJ, Farrand P, Price L, Haase AM, Taylor AH. Web-based intervention using behavioral activation and physical activity for adults with depression (the eMotion study): pilot randomized controlled trial. J Med Internet Res 2018 Jul 16;20(7):e10112 [FREE Full text] [doi: 10.2196/10112] [Medline: 30012547]
56. Stiles-Shields C, Montague E, Kwasny MJ, Mohr DC. Behavioral and cognitive intervention strategies delivered via coached apps for depression: Pilot trial. Psychol Serv 2019 May;16(2):233-238 [FREE Full text] [doi: 10.1037/ser0000261] [Medline: 30407055]
57. Core software for Cochrane Reviews. Cochrane RevMan. URL: https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman [accessed 2023-07-27]
58. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. West Sussex, England: The Cochrane Collaboration and John Wiley & Sons Ltd; Sep 22, 2018:187-241.
59. Deeks JJ, Higgins JPT, Altman DG, The Cochrane Statistical Methods Group. Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. Hoboken, NJ: The Cochrane Collaboration and John Wiley & Sons Ltd; Sep 20, 2019.
60. Globe DR, Levin S, Chang TS, Mackenzie PJ, Azen S. Validity of the SF-12 quality of life instrument in patients with retinal diseases. Ophthalmology 2002 Oct;109(10):1793-1798. [doi: 10.1016/s0161-6420(02)01124-7] [Medline: 12359596]
61. Terada I, Hyde C. The SF-36: an instrument for measuring quality of life in ESRD patients. EDTNA ERCA J 2002 Apr;28(2):73-76, 83. [doi: 10.1111/j.1755-6686.2002.tb00206.x] [Medline: 12216848]
62. Zhou T, Li X, Pei Y, Gao J, Kong J. Internet-based cognitive behavioural therapy for subthreshold depression: a systematic review and meta-analysis. BMC Psychiatry 2016 Oct 21;16(1):356 [FREE Full text] [doi: 10.1186/s12888-016-1061-9] [Medline: 27769266]

63. Rummel-Kluge C, Dietrich S, Koburger N. Behavioural and cognitive-behavioural therapy based self-help versus treatment as usual for depression in adults and adolescents. Cochrane Database Syst Rev 2015 Jun 16;6:CD011744. [doi: 10.1002/14651858.cd011744]

Abbreviations

BDI: Beck Depression Inventory
CCDANCTR: The Cochrane Depression, Anxiety, and Neurosis Controlled Trials Register
CENTRAL: Cochrane Central Register of Controlled Trials
CES-D: The Center for Epidemiologic Studies Depression Scale
iCBT: internet-based cognitive behavioral therapy
ICTRP: International Clinical Trials Registry Platform
MD: mean difference
PHQ: Patient Health Questionnaire
RCT: randomized controlled trial
RevMan: Review Manager
SMD: standardized mean difference
TAU: treatment as usual
WHO: World Health Organization

Edited by J Torous; submitted 02.03.23; peer-reviewed by K Mathiasen, L Udd-Granat; comments to author 01.05.23; revised version received 20.06.23; accepted 11.07.23; published 22.08.23.

Please cite as:
Mamukashvili-Delau M, Koburger N, Dietrich S, Rummel-Kluge C
Long-Term Efficacy of Internet-Based Cognitive Behavioral Therapy Self-Help Programs for Adults With Depression: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Ment Health 2023;10:e46925
URL: https://mental.jmir.org/2023/1/e46925
doi:10.2196/46925
PMID:37606990

Review

Mobile Technologies for Supporting Mental Health in Youths: Scoping Review of Effectiveness, Limitations, and Inclusivity

Shannon Grace Litke[1], MS; Annie Resnikoff[1], BA; Ashley Anil[1], MS; Meredith Montgomery[2], MS; Rishabh Matta[2], BS; Jina Huh-Yoo[2], PhD; Brian P Daly[1], PhD

[1]Department of Psychological and Brain Sciences, Drexel University, Philadelphia, PA, United States

[2]Department of Information Science, Drexel University, Philadelphia, PA, United States

Corresponding Author:
Shannon Grace Litke, MS
Department of Psychological and Brain Sciences
Drexel University
Stratton Hall Suite 118/119
3201 Chestnut Street
Philadelphia, PA, 19104
United States
Phone: 1 7329938165
Email: sgl46@drexel.edu

Abstract

Background: Over the past decade, there has been growing support for the use of mobile health (mHealth) technologies to improve the availability of mental health interventions. While mHealth is a promising tool for improving access to interventions, research on the effectiveness and efficacy of mHealth apps for youths is limited, particularly for underrepresented populations, including youths of color and economically marginalized youths.

Objective: This scoping review study sought to evaluate the following research questions: (1) What is the extent of the current literature on mHealth apps that provide intervention for mental health problems in children and adolescents? (2) What is known from the existing literature about the effectiveness or efficacy of delivering mental health services via mHealth apps? (3) What are the gaps in the knowledge base in the fields of technology and mental health? (4) Do the reviewed mHealth apps address issues of cultural sensitivity or have they been tested with underrepresented groups (ie, youths of color or economically marginalized groups)?

Methods: An electronic database search was conducted using relevant search terms. Seven independent reviewers screened identified studies, including title and abstract review to determine if studies met the following inclusion criteria: (1) targeted samples with mental health symptomology or disorders, (2) studied youth participants aged 6-17 years, and (3) examined the use of a mobile app–based platform for intervention. Relevant studies were subjected to full-text review to extract and chart relevant data based on a priori research questions.

Results: The initial database search yielded 304 papers published from 2010 to 2021. After screening and selection, the final review included 10 papers on the effectiveness and efficacy of mental health intervention apps for youths aged 8 to 17 years. Identified apps targeted a broad range of mental health challenges in youths (ie, depression, self-harm, autism spectrum disorder, anxiety, and obsessive-compulsive disorder). Results identified only a small number of studies suggesting that current effectiveness and efficacy research in this area are limited. While some studies provided general support for the effectiveness of mHealth apps in improving mental health outcomes in youths, several notable limitations were present across the literature, reducing the generalizability of findings. Additionally, considerations around racial, ethnic, and socioeconomic diversity were scarce across studies.

Conclusions: Although some studies cited in this scoping review provide support for the effectiveness and efficacy of mHealth apps targeting mental health concerns in youths, the overall body of literature remains quite limited. Moreover, mHealth apps expressly developed to be culturally responsive are almost nonexistent. Further efforts are needed to recruit youths who are typically underrepresented in research and invite stakeholder participation and collaborative input in the early stages of the mHealth app development process.

(JMIR Ment Health 2023;10:e46949) doi:10.2196/46949

KEYWORDS

mHealth; mobile app; children; adolescents; mental health; effectiveness; efficacy; scoping review; mobile phone

Introduction

Approximately 35% of youths in the United States will have been diagnosed with at least 1 mental health disorder by the time they reach adolescence [1]. When these children do not receive mental health services during the school-age years, the disorders tend to persist [2] and are associated with considerable social, behavioral, and educational and vocational problems and lower quality of life in adulthood [3]. Mobile health (mHealth) is regarded as an important new tool for the assessment and treatment of mental and physical health conditions because these technologies can help reduce logistical and system-level barriers as well as stigma for children and families seeking resources. “mHealth” is a broad umbrella description that refers to various mobile and wireless apps, including SMS text messaging, apps, wearable devices, remote sensing, and the use of social media in the delivery of health-related services [4]. In this paper, the term “mHealth apps” is used to describe mental health mobile technology apps designed specifically for the treatment of mental health symptoms.

Over the previous 2 decades, there has been a sharp increase in the development and use of mHealth apps to improve remote access to, and delivery of, evidence-based care [5] with over 10,000 apps designed specifically for mental or behavioral health available in the marketplace [6,7]. However, few rigorous studies have examined the effectiveness of mHealth apps for youths [8]. In a systematic review of mental health mobile apps for youths [9], authors found insufficient research evidence to support the effectiveness of mHealth apps for youths with mental health problems. Importantly, mHealth’s efficacy in marginalized youth populations has received even less attention in the empirical literature, even though an mHealth model confers unique advantages, including improved access to mental health care, equity of resources, immediate availability, lower cost, and tailored content [9,10]. Given how quickly mobile technologies evolve and are adopted, it is important to constantly examine data from well-designed studies to support practitioners’ understanding around potential benefits and obstacles to the use and effectiveness of mHealth technologies, including with marginalized youth populations.

The familiarity of smartphone apps, ease of use [11], and near ubiquity among youths [12] suggest that mental health treatment providers may benefit from incorporating mHealth technologies and associated features into treatment strategies and plans. The use of mHealth apps can also improve treatment engagement and quality of care by providing more continuous access to self-guided tools [8,13]. For example, mHealth apps can provide easy and timely access to evidence-based strategies that have been demonstrated to reduce mental health symptoms such as journaling, self-monitoring, use of thought records, and relaxation training through increased awareness of the behaviors and practice of these skills [8]. mHealth apps can also promote engagement with these activities through in-app reminders, instructions, and activity templates [14].

While findings from some studies suggest that mHealth apps for mental health are associated with reduced symptoms of depression and anxiety, reduced frequency of self-harm, and increased coping self-efficacy [15], further research on effectiveness is required [9,16]. For example, many reviews cite the inadequacy of previous research in the field, which has ultimately hampered the wider acceptance of mHealth interventions for mental health in youths. These limitations led to recent calls for methodologically robust studies evaluating the safety, efficacy, and effectiveness of these apps [9,17-20]. Furthermore, considering the socioeconomic “digital divide” that may stymie delivery of care to marginalized populations via an mHealth model [21], it is important that research efforts also seek to examine effectiveness within diverse, representative samples of youths.

Underrepresented groups, including youths of color and economically marginalized youths, are at an increased risk for adverse experiences and stressors, compounded by barriers including poor access to care and substandard mental health treatment [22]. Reduced access to care is further compounded by evidence suggesting that the stigma of mental health problems is far greater among youths of color compared to White youths [23,24]. Service pathways and barriers to accessing mental health care are influenced by many individual, interpersonal, and systemic factors [25,26]. Specifically, poor cross-cultural understanding and communication can lead to decreased detection of mental health problems among youths of color [27]. Disparities in connecting with and completing intervention services may remain present even when needs are identified [28]. Adding to this challenge, distrust of mental health professionals, concerns of stigma, or difficulty identifying and accessing quality care, potentially due to lack of resources (eg, shortage of insurance or transportation), may prevent underserved youths from seeking or locating care [29]. A meta-analysis by Hall and colleagues [30] indicated that culturally adapted interventions were associated with high levels of acceptance and satisfaction, in addition to improved outcomes. However, interventions for children that focus on general symptoms of depression, anxiety, and resiliency without inclusion of cultural considerations are more commonplace [31]. Given the widespread use of mobile phones, mHealth has the capacity to provide effective care for vulnerable populations, irrespective of age, socioeconomic status, or geographic area. Health communication technologies can be tailored to fit specific needs in a culturally competent manner to the target population, which is crucial in developing effective interventions.

The objective of this scoping review study was to evaluate the current literature on the effectiveness and efficacy of mHealth apps that target mental health problems in youths. Additionally, this study aimed to identify gaps in the current literature, particularly with regard to including members of

underrepresented populations in effectiveness research for mHealth apps designed for mental health treatment in youths.

Methods

Study Design

As defined by Arksey and O'Malley [32], a scoping review serves to map the key concepts involved in a particular research area by summarizing research findings and identifying gaps in the evidence base. Compared to a systematic review, which often includes a more narrowly defined research question, a scoping review approach is better suited for characterizing themes from research areas with preliminary evidence or emerging literature, such as the topic of mHealth apps for mental health intervention in youths. To ensure methodological rigor, this review was designed and conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [33]. Additionally, the methodological framework set forth by Arksey and O'Malley [32] was used to conceptualize the development of research questions, search and selection procedures, and qualitative analysis as detailed in the following steps.

Identification of Research Questions

The aim of this scoping review was to identify and synthesize information from the literature according to the following a priori research questions: (1) What is the extent of published evidence on using mHealth apps that provide intervention for mental health problems in children and adolescents? (2) What is known from the existing literature about the effectiveness or efficacy of delivering mental health services via mHealth apps? (3) What are the gaps in the knowledge base in the fields of technology and mental health? (4) Do the reviewed mHealth apps address issues of cultural sensitivity or have they been tested with diverse samples? To explore this final research question in more depth, we examined the following questions: (1) Were there specific considerations in recruitment efforts to include underrepresented groups (ie, youths of color or economically marginalized groups)? (2) For studies that included underrepresented or underserved youths, what did the results suggest were barriers and facilitators to effective use of the mHealth app?

Identification of Relevant Studies

A search of the APA PsycINFO electronic database was conducted on July 21, 2021, using the following terms: "mental health" OR "counseling" OR "psychotherapy" OR "anxiety" OR "depression" OR "CBT" OR "cognitive behavioral") AND ("app" OR "app-based" OR "mobile app" OR "smartphone" OR "mobile application") AND ("children" OR "adolescents" OR "youth" OR "teenagers" OR "high school." Papers were included in the initial search if they were peer-reviewed, available in the English language, and published between January 2010 and the date of search. The year 2010 was chosen because the literature examining technology-based mental health interventions before this time largely focused on web-based programs that lacked mobile app components [34].

Study Selection and Screening Procedures

The study selection procedure included two review stages: (1) title and abstract review and (2) full-text review. First, 7 independent reviewers evenly divided the search results to complete a screening of the title and abstract for each paper. Papers were marked as relevant for inclusion in the full-text review stage if they met the following criteria: (1) targeted samples with mental health symptomology or disorders, (2) studied youth participants aged 6-17 years, and (3) examined the use of a mobile app–based platform for intervention. Of note, studies were excluded if they examined nonmobile technologies (eg, a computer-based platform vs a smartphone app). The age range of 6-17 years was selected to best capture the elementary and secondary school-age range of first grade through high school. In addition, the selected age range aligns with inclusion age criteria for the National Survey of Children's Health study that looked at mental health disorders and disparities of mental health care use in children [35]. Moreover, the age range of 6-17 years is commonly used in cross-cutting symptom measures for youths [36].

To assess the interrater reliability of study selection, a total of 82 papers (20% of total search results) were randomly selected for review by 2 independent reviewers, and the consensus was evaluated by a third reviewer. After full-text review, papers were excluded if they did not directly assess and report outcome data on the effectiveness of the mobile intervention. For example, studies reporting solely on usability, feasibility, or qualitative measures of acceptability were excluded from the final review. Review papers such as meta-analyses, systematic, or scoping reviews were excluded from the final analysis. However, review papers were screened to identify relevant references, which were then entered into the title and abstract review stage for evaluation.

Charting the Data

After a full-text review, the following data were extracted from papers within the final review pool: sample characteristics (ie, size, age range, clinical presentation, or mental health characteristics), name, purpose, and brief description of the mHealth app, demographic characteristics of the sample, outcome measures, and effectiveness findings.

Results

Overview

As displayed in Figure 1, the initial PsycINFO search yielded 304 paper references for further eligibility analysis. Upon title and abstract review, 272 papers were excluded based on the specified criteria. No discrepancies in reviewer's inclusion or exclusion decisions were noted upon consensus review. Of the papers identified from the initial search that were marked as relevant for full-text analysis (n=32), 13 were review papers (eg, meta-analyses, systematic, or scoping reviews), which were excluded from the final analysis. Screening of the references in these review papers resulted in an additional 107 papers that were examined via title and abstract review, after which 88 were excluded. A total of 51 papers were subject to full-text review, 41 of which were excluded because they included participants

outside of the target 6-17 years age range, did not target mental health symptomology, reported no effectiveness data, did not specifically include a mobile- or app-based platform for intervention, or were review papers. The final analysis was conducted on 10 papers. Table 1 presents a summary of the study characteristics.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flowchart of study selection.

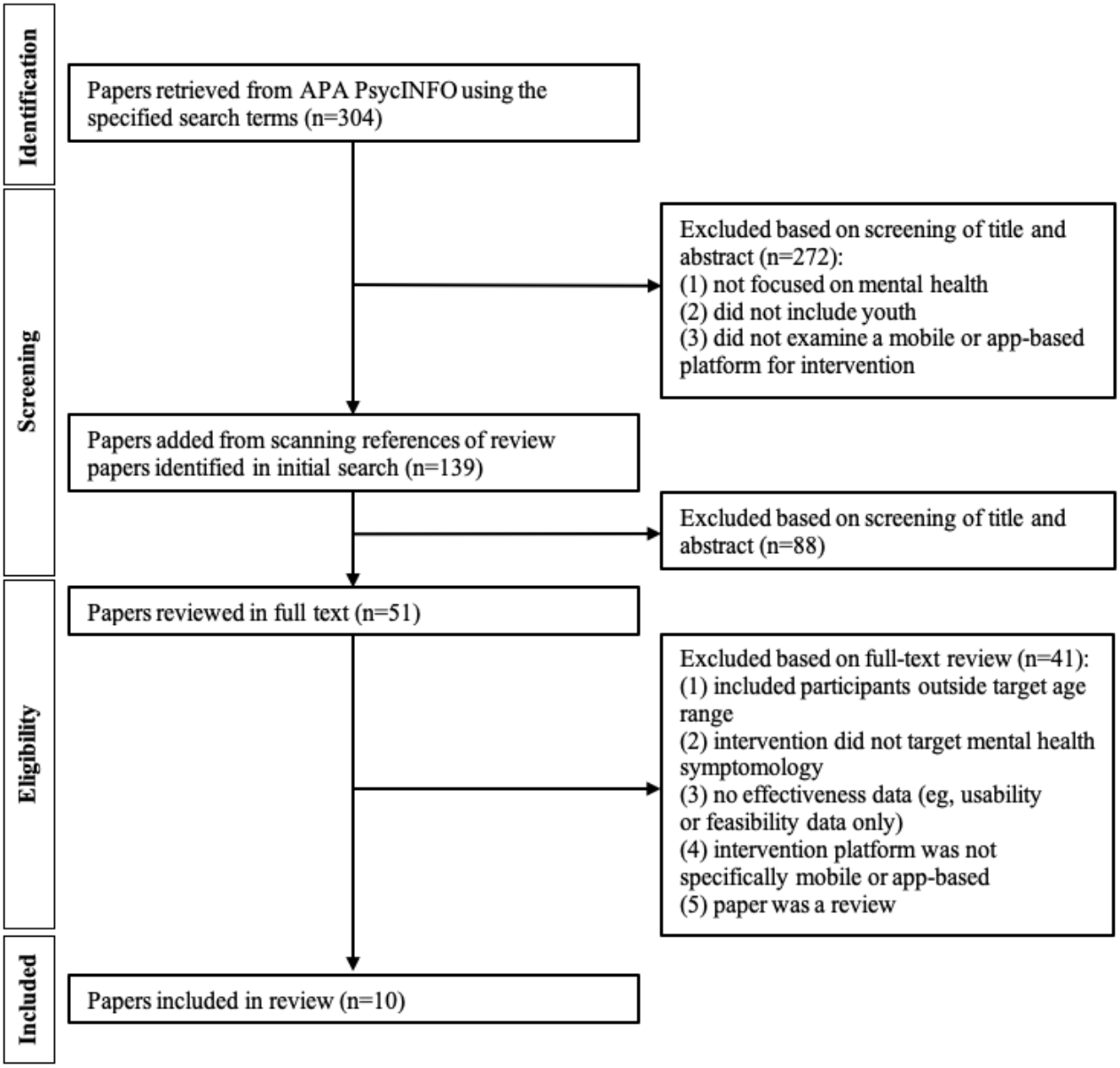


Table 1. Summary of effectiveness studies on mobile interventions for mental health in youths.

App name	Sample	App components	Demographic characteristics	Primary outcome measures	Results
MOSOCO[a] [37]	N=12; 8-to 11-year-old youths; 25% (n=3) with ASD[b]	Social skills training	No gender, racial or ethnic, or socioeconomic demographic data reported	Quantitative coding of target behaviors from video recordings	76.5% reduction in time engaged in social missteps for students with autism using MOSOCO compared to those not using the app (*P*=.002).
CARE mobile app [38]	N=80; 12- to 15-year-old youths with moderate to high rumination	Mindfulness exercises	53.8% boys, 45% girls, 1.2% chose not to answer; 86.25% White, 2.5% Native American, 1.25% Black, 1.25% multiracial, 8.75% chose not to answer; 93.75% non-Hispanic, 3.75% Hispanic, 2.5% chose not to answer; median parental-reported income range US $100,000-US $125,000; 4% reported recipients of government-assisted food program	MASC,[c] CRSQ,[d] CDI,[e] PSC-I[f]	Significant reduction in rumination (η^2=0.112), anxiety (η^2=0.145), and parent-reported internalizing symptoms (η^2=0.370) at 12-week follow-up.
Open Autism Software suite (apps for tablets)—Drawing, Music, Untangle, Photogoo [39]	N=8; 10- to 14-year-old youths with ASD	Social skills training	62.5% boys, 37.5% girls; no racial or ethnic or socioeconomic demographic data reported	Quantitative coding of target behaviors from video recordings	Significantly more verbal (*P*=.001) and physical (*P*<.05) interactions per minute when using app versus not using app. No difference in number of supportive comments, social missteps, or atypical behaviors.
ICBT BiP OCD[g] [40]	N=67; 12- to 17-year-old youths with OCD[h]	Web-based CBT,[i] exposure tasks, ERP[j] reminders	No gender data reported; 93% born in Sweden, 4% born in other European countries, 3% Asian; no socioeconomic demographic data reported	CY-BOCS[k]	Significantly greater reduction in CY-BOCS scores over time for youths using BiP OCD[l] intervention compared to waitlist control at posttreatment (*d*=0.69) and 3-month follow-up (*d*=1.68).
BYOTS[m] [41]	N=72; 12- to 15-year-old youths	Cognitive restructuring	100% female; 100% Black or biracial; 100% enrolled in the federal free breakfast and lunch program	MASC	Significant reduction in MASC total anxiety scores from pre- to postintervention (*d*=0.52).
WeClick [42]	N=193; 12- to 16-year-old youths	Cognitive restructuring, social learning, problem-solving, and conflict resolution skill development	86.53% female; 3.6% Aboriginal or Torres Strait Islander; no socioeconomic demographic data reported	PHQ-A,[n] WEMWS,[o] GHSQ[p]	Decrease in PHQ-A scores from baseline to posttest was not significant (*P*=.138). Significantly greater increase in select secondary outcomes for app users versus control, including well-being (WEMWS; *d*=0.37), help-seeking intention (GHSQ; *d*=0.36), and professional help-seeking intention (GHSQ-P[q]; *d*=0.36).
SmartCAT (version 2.0) [43]	N=34; 9- to 14-year-old youths with anxiety (GAD,[r] SAD,[s] or SocAD[t])	CBT, skill development, and exposure tasks	50% female; 85.3% White, 2.9% Hispanic, 14.7% biracial; total family income mean US $70,001-US $90,000	PARS[u]	Significant reduction in PARS severity scores from pre- to posttreatment (*d*=1.05) and posttreatment to 2-month follow-up (*d*=0.55).

App name	Sample	App components	Demographic characteristics	Primary outcome measures	Results
BlueIce [20]	N=44; 12- to 17-year-old youths with current or past self-harm	CBT and DBT[v] harm-reduction strategies, safety checks, and rerouting to emergency services when necessary	91% girls; no racial or socioeconomic demographic data reported	SDQ,[w] MFQ,[x] RCADS[y]	Reduction in self-harm in 73% of participants as well as a significant reduction in depression (MFQ; *P*=.04) and anxiety (RCADS; *P*=.001) symptoms postintervention.
Mayo Clinic Anxiety Coach [44]	N=2; 10- and 16-year-old youths with OCD (case study)	Exposure tasks, ERP	10-year-old White female, 16-year-old White male; no socioeconomic demographic data reported	CY-BOCS	Decrease in OCD severity ratings from pre- to posttreatment was observed for both participants in this case study.
Mayo Clinic Anxiety Coach [45]	N=8; 8- to 17-year-old youths with a primary diagnosis of an anxiety disorder or OCD	Exposure tasks, ERP	75% girls; 100% White; no socioeconomic demographic data reported	PARS, SCAS,[z] CRS,[aa] CGI-S/I[ab]	Decrease in CGI-S/I, CRS, SCAS, and PARS scores from pre- to posttreatment. Reported large effect sizes.

[a]MOSOCO: Mobile Social Compass.
[b]ASD: autism spectrum disorder.
[c]MASC: Multidimensional Anxiety Scale for Children.
[d]CRSQ: Children's Response Styles Questionnaire.
[e]CDI: Children's Depression Inventory.
[f]PSC-I: Pediatric Symptom Checklist-Internalizing Subscale.
[g]ICBT BiP OCD: Internet-delivered CBT BarnInternetProjektet for obsessive-compulsive disorder.
[h]OCD: obsessive-compulsive disorder.
[i]CBT: cognitive behavioral therapy.
[j]ERP: exposure and response prevention.
[k]CY-BOCS: Children's Yale-Brown Obsessive-Compulsive Scale.
[l]BiP OCD: BarnInternetProjektet for obsessive-compulsive disorder.
[m]BYOTS: Build Your Own Theme Song.
[n]PHQ-A: Patient Health Questionnaire-Adolescents.
[o]WEMWS: Warwick Edinburg Mental Wellbeing Scale.
[p]GHSQ: General Help-Seeking Questionnaire.
[q]GHSQ-P: General Help-Seeking Questionnaire—Professional Help-Seeking Intentions Score.
[r]GAD: generalized anxiety disorder.
[s]SAD: separation anxiety disorder.
[t]SocAD: social anxiety disorder.
[u]PARS: Pediatric Anxiety Rating Scale.
[v]DBT: dialectical behavioral therapy.
[w]SDQ: Strength and Difficulties Questionnaire.
[x]MFQ: Mood and Feelings Questionnaire.
[y]RCADS: Revised Child Anxiety and Depression Scale.
[z]SCAS: Spence Children's Anxiety Scale.
[aa]CRS: clinician's severity rating.
[ab]CGI-S/I: Clinical Global Impression Scale—Severity/Improvement.

Exploring Research Aim 1: Extent of the Literature

Overview

Overall, results of this scoping review reveal that efficacy and effectiveness research on mobile platforms targeting mental health concerns in youths are limited with only a small number of studies examining intervention apps (N=10). It is important to highlight that the paucity of studies in this area renders it inappropriate to draw overarching conclusions about the efficacy or effectiveness of app-based interventions for mental health problems in youths. Although limited in number, the mHealth apps identified across these studies use a variety of innovative strategies for managing mental health concerns that are worthy of description. Participants in these studies ranged from age 8 to 17 years and presented with a variety of clinical symptomatology and disorders, including depression and self-injurious behavior [20], autism spectrum disorder (ASD)

[37,39], anxiety [43], and obsessive-compulsive disorder (OCD) [40,44,45].

Additionally, 2 studies [38,41] examined mobile apps targeting symptoms of anxiety, although they included school-based or community (rather than clinical) samples of youths. Furthermore, one study reported on a mobile app, "WeClick," (Black Dog Institute) designed to improve overall well-being for youths with a broad range of mental health problems [42] but with a specific focus on the primary outcome of depressive symptoms. The next section describes the apps included in the scoping review.

Depression and Self-Harm

This review identified one effectiveness study of an app specifically targeting depression and self-harmful behavior in adolescents. BlueIce (Oxford Health NHS Foundation Trust and University of Bath) is a smartphone app designed to provide youths who have a history of self-harm with 24/7 access to a variety of cognitive behavioral therapy (CBT) and dialectical behavioral therapy intervention tools [20]. Strategies included identifying and challenging negative cognitions, safety planning and routing to appropriate emergency contacts or services, distress tolerance tools, and a menu of personalized activities designed to improve mood [20].

Additionally, the WeClick app [42] was designed as a single-session intervention based on principles of CBT and social learning theory. Specifically, this mobile intervention focuses on relationship-specific issues (eg, family and peer conflict, intimate relationships, and substance use) and promotes cognitive restructuring to help youths overcome these relationship challenges.

Autism Spectrum Disorder

A study by Hourcade and colleagues [39] describes a package of tablet apps available through the Open Autism Software suite (University of Iowa) designed as a mobile intervention to promote social interaction in youths with ASD. Specifically, the apps include activities to encourage face-to-face social interaction and enhance creativity, sharing, turn-taking, and emotion modeling [39]. Similarly, the Mobile Social Compass app (Social & Technological Action Research Group at the University of California, Irvine) was designed to support social skills practice in real-life situations for youths with ASD [37]. This mobile-assistive technology uses the smartphone camera to augment real-life social situations (eg, interaction on the playground at recess) with on-screen visual supports and reminders (eg, cues to make eye contact with the conversation partner). As both platforms highlight, the functionality allowed by mobile intervention modalities (vs table-top or computer-based) offers a unique benefit to youths with ASD, allowing them to access social-skills coaching in real time throughout the day [37,39].

Anxiety

Several app-based interventions targeting anxiety symptoms in youths have been developed and examined for efficacy within the past decade. Silk and colleagues [43] conducted a study examining the SmartCAT (version 2.0; University of Pittsburgh) app, a mobile platform designed as an adjunctive resource to support the practice of CBT skills outside of treatment sessions for youths with anxiety disorders. SmartCAT (version 2.0) uses a web-based interface to engage youths, provide cues and digital rewards for practicing skills at home, and prompt youths to engage in personalized exposures outside of session [43]. A mindfulness mobile app, as described in a recent study by Hilt and Swords [38], specifically targets adolescents' rumination and worry through a variety of mindfulness techniques, including breathing exercises and body scan activities. Further, Neal-Barnett and colleagues [41] developed and evaluated the Build Your Own Theme Song (BYOTS) app (Kent State University), a mobile app designed to deliver a cognitive restructuring intervention to Black middle-school girls. The BYOTS app incorporates a culturally informed approach involving a musical intervention to help users identify and disrupt negative thought cycles by creating their own personal, affirming theme songs [41].

Obsessive-Compulsive Disorder

Lenhard and colleagues [40] evaluated the effectiveness of a clinician- and parent-supported web-based CBT program for OCD, BiP OCD (BarnInternetProjektet for obsessive-compulsive disorder; Barn-och ungdomspsykiatri Region Stockholm), which includes a smartphone app to facilitate the child's completion of exposure tasks, manage progress, and engage parents in psychoeducation on parent-specific topics (eg, family accommodation). Similarly, the Mayo Clinic Anxiety Coach app (Mayo Clinic) [44,45] facilitates parent coaching of exposure therapy for youths with OCD and other anxiety disorders outside of therapy sessions. Additionally, this app includes a variety of clinical content, including tabs with graphics and information to support psychoeducation on different topics, direct access to a tool for youths to build and edit their own fear hierarchies, and built-in forms to support the monitoring of anxiety symptoms. Overall, these apps are designed to support youths in applying the skills learned in treatment for OCD by engaging in out-of-session exposure and response prevention, which is empirically supported as the gold-standard behavioral treatment for youths with OCD [46].

Exploring Research Aim 2: Evidence of Effectiveness and Efficacy

Overview

Although few in number, the intervention apps identified in this review were generally found to be associated with symptom reduction and clinical improvement. Interestingly, the handful of mobile interventions targeting anxiety and OCD consistently demonstrated positive effects, while apps targeting other domains of clinical symptomatology (ie, depressive symptoms, self-harm, and ASD) showed more variability in results. Of note, due to the small number of effectiveness and efficacy studies identified in this scoping review, results have limited generalizability and should be interpreted with caution.

Anxiety and OCD

A study by Hilt and Swords [38] found that use of the mindfulness app intervention was associated with significant

reductions in rumination, as well as other anxiety and internalizing symptoms from pre- to postintervention, with a large effect persisting throughout a 12-week follow-up period. Similarly, Silk and colleagues [43] reported that the SmartCAT (version 2.0) app demonstrated a large effect on reducing anxiety symptom severity from pre- to posttreatment, with a moderate effect persisting at a 2-month follow-up. This change was found to be statistically significant, and clinically meaningful, with 20 out of 30 (66.67%) participants no longer meeting diagnostic criteria for an anxiety disorder at posttreatment [43]. Further, a study by Lenhard and colleagues [40] marks an important contribution to the literature as one of the 2 randomized controlled designs identified in this overall review. Findings revealed that use of the BiP OCD app had a large effect on significantly reducing OCD symptom severity from pre- to posttreatment and at 3-month follow-up [40]. The use of another app targeting OCD and anxiety symptoms, Mayo Clinic Anxiety Coach, was found to be associated with a significant reduction in symptom severity with authors reporting a large effect size; however, conclusions are limited by the small sample sizes in both studies examining this app [44,45]. Additionally, Neal-Barnett and colleagues [41] reported a medium effect size for BYOTS app use, which was associated with significant reductions in total anxiety scores on the Multidimensional Anxiety Scale for Children (MASC) from pre- to postintervention.

Depression and Self-Harm

Examination of the 2 identified studies of intervention apps targeting depressive symptoms and self-harm reveals mixed results. Use of BlueIce was associated with significant reductions in depression and anxiety symptoms as well as reduced frequency of self-injurious behavior in youths presenting with current or past self-harm [20]. In contrast, a randomized controlled trial of the WeClick intervention indicated that the use of this platform was not associated with a significant reduction in depressive symptoms, although some secondary outcomes such as mental health well-being and help-seeking intentions showed significant improvement over time [42].

Autism Spectrum Disorder

For the 2 identified studies of apps targeting the development of social skills in youths with ASD, findings generally support the use of the app in facilitating positive outcomes; however, there were some differences noted between the apps in terms of the specific outcomes. For instance, use of the Mobile Social Compass intervention was associated with a significant reduction in frequency of social missteps for youths with autism [37]. In contrast, the Open Autism Software suite did not seem to impact social missteps, atypical behaviors, or frequency of supportive comments for youths with autism. However, the authors noted a significant increase in number of verbal and physical interactions per minute while using the app [39]. Of note, data in these studies were limited, consisting of quantitative coding of target behaviors in video recordings. As such, without multimodal data from other empirically validated measures or psychometric standards, results should be interpreted with caution in terms of the effectiveness of these apps.

Exploring Research Aim 3: Limitations in the Literature

The mobile intervention apps identified in this review were generally found to have positive effects on mental health outcomes in youths; however, there were several notable limitations among these studies that reduce the generalizability of findings. First, many of the studies included in this review were limited by small sample sizes [37,39,44,45]. Of the larger studies, only 2 were randomized controlled trials [40,42], and the other studies used within-subjects designs to assess intervention effects [20,38,41,43]. As acknowledged by many of these authors, randomized controlled trials mark an important next step in replicating findings from these pilot studies and assessing the effectiveness of these apps while controlling for within-subjects factors.

Most studies in this review involved samples with a limited range of baseline symptom severity, posing an additional threat to generalizability and highlighting an important gap to bridge in future effectiveness studies. For example, results from a study by Hourcade and colleagues [39] suggest that the Open Autism Software suite is a promising skill-development tool for youths with ASD, but the sample was limited to individuals on the high-functioning end of the spectrum. Additionally, Neal-Barnett and colleagues [41] reported that use of the BYOTS app was associated with a statistically significant reduction in the MASC from preintervention to postintervention. While this reduction does mark a shift from “high average” to “average” symptom severity according to cutoff ranges on the MASC [47], it is important to note that participants’ average anxiety severity scores were subclinical at both time points. Therefore, further research is needed in order to determine the effectiveness of this app for youths displaying clinically significant anxiety severity. Similarly, work by Lenhard and colleagues [40] highlights that further investigation of the BiP OCD intervention is needed to identify the clinical appropriateness of this app for youths presenting with greater severity of OCD symptoms. Specifically, the field would benefit from research aimed at pinpointing the symptom profiles of youths that suggest the need for in-person clinical care and to identify whether there is a use case for a stand-alone mobile intervention for a subset of the clinical population whose symptoms are less impairing.

Several authors commented on implications and directions for further research based on gaps in the literature that were left unaddressed in their studies. For instance, Hilt and Swords [38] emphasize the need for further investigation of dosage to understand the minimum amount of user interaction required to achieve meaningful outcomes. Similarly, certain apps were not tested as stand-alone apps but rather in tandem with larger intervention programs [40,41,43-45].

Further dismantling design analyses are required to determine the specific effectiveness or efficacy of these app-based intervention components when used as stand-alone options. Additionally, certain authors [41,43] emphasized the need for continued research on the specific mechanisms of change underlying response to these mHealth interventions as well as investigation of their effectiveness or efficacy across diverse clinical and demographic profiles.

Exploring Research Aim 4: Inclusion of Underrepresented Groups

Considerations around cultural, ethnic, and socioeconomic diversity were alarmingly scarce across studies included in this scoping review. In addition, the level of detail regarding reported sample demographic information was inconsistent across studies. For example, a handful of studies did not report any demographic data on participants' race, ethnicity, or socioeconomic status [20,37,39], while other studies reported limited information [40,42]. In the studies that did include details of the racial, ethnic, and socioeconomic characteristics of youth participants, their samples lacked diversity and the authors did not note procedures to recruit members of historically marginalized populations or those underrepresented in research [38,43-45]. Some authors explicitly mentioned that the generalizability of results was limited by a lack of diverse sampling [38], while others did not acknowledge this limitation. Of the 10 studies included in this review, only 1 study included an explicit focus on testing the app in a sample of underserved, underrepresented youths, who were Black adolescent females from economically marginalized backgrounds [41].

Neal-Barnett and colleagues [41] cite a plethora of evidence to support their emphasis on the use of music as a powerful, culturally significant tool for Black youths. For example, these authors discussed the historical importance of gospel and spiritual music for Black adolescents and young adults during the civil rights era as well as the emergence of hip-hop as a mode of expression for Black youths facing marginalization in underserved, urban areas. Drawing on research that shows Black female youths are (1) disproportionally affected by anxiety disorders, (2) clinically underserved and underrepresented in research, and (3) more likely to turn to music to express their emotions than their male counterparts, the authors exclusively recruited a sample of Black or biracial female adolescents from seventh- and eighth-grade classrooms across a low-income, urban school district where 100% of students were economically underserved.

To examine barriers and facilitators relevant to use of the app, qualitative interviews were conducted. Results revealed that girls in the sample were initially skeptical over whether the app would work for them. Their doubts reportedly stemmed from their prior experiences as participants in other research initiatives, wherein the outcomes did not benefit them as they had expected. By the end of the BYOTS study, however, participants reported being pleased that their expectations of the app matched their actual experience, and "it really worked!" [41]. This qualitative finding holds important implications in terms of researchers' responsibility to build trust and rapport with participants by continuously soliciting participant feedback to collaboratively iterate and improve the user experience. Importantly, these authors acknowledged limitations to generalizability within their own study, namely that Black girls represent a heterogenous population. Further research on the effectiveness of the BYOTS app is needed, including Black females residing in different settings (eg, rural and suburban) and from a variety of socioeconomic status levels [41].

Discussion

Principal Findings

While there are a considerable number of effectiveness and efficacy studies of mobile mental health technologies in adult populations [48], findings from this scoping review highlight the paucity of these types of studies relevant to mHealth apps for youths. Instead, much of the current literature is focused on preliminary studies of feasibility or acceptability. For example, a recent study by Punukollu and colleagues [49] examined student and teacher buy-in to the Safespot app (SafeSpot Limited), a mental health support app designed for use in school settings. Qualitative results of this study demonstrated that this app was regarded as acceptable and likable by stakeholders [49]. Although this study was excluded from this review due to the lack of outcome data on the app's effectiveness, it is important to note that these types of pilot feasibility and acceptability studies often precede larger-scale efficacy and effectiveness trials. As such, the fact that numerous smaller-scale, qualitative studies of mHealth apps are currently underway may be considered a promising marker for future effectiveness trials. However, it is also important to consider that many pilot apps do not reach the efficacy or effectiveness stages of research often due to technological advances outpacing the slower and more deliberate process of psychological research.

From the 10 papers included in this review, several overarching themes emerged, including implications of the narrow range of mental health disorders in the effectiveness studies, considerations surrounding regulation and safety of mHealth technologies, and the limited number of studies that included diverse cultural, ethnic, and socioeconomic samples.

Narrow Clinical Targets

Most studies with larger-scale effectiveness data targeted anxiety-related disorders. This is perhaps not surprising given that anxiety is one of the 3 most diagnosed mental health disorders in children [50] and is the most diagnosed disorder among 12- to 17-year-olds, an age range more likely to own a smartphone or tablet device as compared to younger children [51]. Even in the adult literature, most of the large-scale effectiveness studies of mHealth apps focus on anxiety and depression [52]. Thus, there is a gap in high-quality effectiveness studies of mHealth apps for other mental health disorders frequently seen in youths including eating disorders, disruptive behavior disorders, and OCD. In the absence of reliable outcome data on the use of mHealth apps targeted at specific disorders, clinicians may rightfully feel cautious in recommending mHealth apps to their clients. Similarly, consumers may struggle to discern which mHealth apps provide benefits relative to their specific mental health challenges.

Safety Regulations

In recent years, there have been increased regulatory efforts for mobile medical apps that focus on safety and effectiveness; however, groups such as the Food and Drug Administration [53] do not regulate health and wellness apps that are not intended for medical use. Instead, clinicians and consumers can

access websites such as PsyberGuide [54] that provide ratings and reviews by mental health experts that include any information on published evidence. Although these types of websites provide rating criteria developed by mental health experts, the proliferation of mHealth apps into the marketplace is so rapid that these websites frequently struggle to keep pace with timely reviews. As an alternative, some organizations have published frameworks that clinicians can use to evaluate mental health apps [55]. Even with these frameworks, safety regulations and guidelines for contraindications of mHealth app use are important targets of future research.

Investigations are also needed that examine individual characteristics of populations for whom mHealth apps prove effective. For example, research suggests that mHealth apps have the potential to act as a preventative tool for individuals with subclinical symptoms or as a maintenance tool for individuals who are no longer in need of traditional professional help but are compelled to continue treatment in a more cost-effective manner [56].

Recommendations for Addressing Inclusivity

The surprising lack of studies on the use of mHealth apps in underrepresented populations (namely, youths of color and economically marginalized youths) has been observed in many types of digital health apps across all age groups [57-59]. This failure to include diverse samples must be addressed in the development and evaluation of digital health tools, given the systemic inequities facing youths of color with regard to access and bias in receiving quality health care. Review and perspective papers suggest multiple approaches for addressing the gap [57-62], but we focus on three themes that apply to increasing evidence on mental health apps for underrepresented youths: (1) tailorability, (2) human-centered design, and (3) trust-building.

The increased adaptability of digital approaches can enable tailored content and adjust the type or dosage of the intervention based on individual characteristics (eg, multiphase optimization strategy [63] and sequential, multiple assignment, randomized trials) [64]. For youth populations, the solution should be tailored to factors such as the individual's racial, economic, and educational background, sex, family structure (ie, parenting and guardian contexts), personal hobbies and interests, and types and levels of mentorship available. In other words, digital interventions should be tailored beyond group characteristics (eg, race) to address individual differences.

As more digital mental health apps are developed, incorporating human-centered design and community-centered approaches will help identify the barriers and develop strategies to mitigate those barriers [57,61,65-67]. Recruitment strategies should be culturally sensitive and address community mistrust, participant resource constraints, and potential risks such as community stigma [68]. Participatory design has a long history of effectiveness in addressing community needs [69-72] and is a widely used method in human-computer design for youth populations, given its ability to increase interest and engagement in youths. For example, the BlueIce app [20], one of the mHealth apps included in this review, used a co-design model in the development phase, including a team of youth participants who collaboratively advised on aspects of the app content, design, and layout. Stallard and colleagues [20] found that not only was use of the BlueIce app associated with significant reductions in anxiety, depression, and self-harm but also that this app ranked highly on measures of acceptability with 29 out of 33 (88%) participants indicating that they would like to keep using the app after the study. This type of participatory approach should be adopted as a standard procedure across the development and evaluation of mHealth apps with particular efforts to include underrepresented voices in this process to inform culturally salient adaptations and maximize benefits to recipients of the mHealth intervention.

Even when mental health interventions are available and accessible, trust is still one of the critical barriers to the adoption of interventions by underrepresented populations [68,73]. The process of including end users from the beginning of the design phase, such as participatory design, can increase trust, engagement, and adoption [63,74]. The solution's clinical testing results also should be explicitly available [75] and understandable by the end-user population. Moreover, the solution should have the ability to integrate with existing clinical, community, and school environments to enable broader dissemination and participation.

Limitations

Many efforts were taken to maximize methodological rigor while conducting this review, but there are several important limitations. First, this review used only 1 database (ie, APA PsycINFO) for the initial search and identification of relevant studies rather than replicating the search across multiple databases. Although APA PsycINFO is widely regarded as a reliable and comprehensive index of records spanning the field of psychological science, it is possible that the initial search conducted for this review did not capture all relevant mHealth studies. Additionally, while reasons for exclusion were documented and categorized as part of the process of the full-text review process for determining eligibility, some studies met several exclusion criteria (eg, participants outside of the specified age range as well as lack of effectiveness data). Due to these overlapping exclusion categories, this analysis lacks a precise, quantitative measure of the number of apps excluded for each specific reason. These data may have offered a more precise understanding of the current volume of feasibility or acceptability studies of mHealth apps for youths. Further, it is important to acknowledge that changes in the technological landscape often occur more rapidly than the pace at which research is conducted, reviewed, and published. It is likely that even the more recent apps highlighted in this review will have markedly evolved or perhaps been discontinued. While it is important to take these limitations into account when interpreting the results and implications of this study, the overall integrity and relevance of this review remain intact. Namely, findings serve to clarify the state of the current literature and prompt critical consideration of future directions for research on mHealth apps for youths with mental health concerns.

Conclusions

There is a pressing need for evidence-supported adjunct options to mental health therapy services for youths given the clear gap

between prevalence and treatment rates. For example, only about one-third of children and adolescents with mental health problems access treatment [76,77], and many encounter pervasive barriers such as lengthy waitlists to see mental health providers, and few providers accept insurance [78]. When considering the ubiquity of technology and the comfort with which youths interface with mobile devices, the promise of mHealth apps has always held the potential to be a supplemental option for mental health support, assessment, and intervention for children and adolescents with mental health concerns. mHealth apps provide easy and real-time accessibility for youths to evidence-based tools and can provide some level of support when stigma or treatment barriers impede access to mental health providers. Although some studies cited in this scoping review provide support for the effectiveness and efficacy of mHealth apps targeting mental health concerns in youths, the overall body of literature remains quite limited. Moreover, mHealth apps expressly developed to be culturally responsive are almost nonexistent. These findings suggest the promise of mHealth apps for mental health in youths remains unfulfilled but is conversely not without optimism.

Increased advocacy and engagement by mental health providers in the development, implementation, and evaluation of mHealth apps would go a long way to inspiring more confidence about effectiveness for end users as well as for treatment providers who are interested in recommending apps as supplemental support options. Further efforts are also needed that focus on recruiting diverse samples, including samples of youths who are typically underrepresented in research, and on being inclusive by inviting participation and collaborative input in the early stages of the mHealth app development process. It takes a village to support mental health in youths, and mHealth can play an important role in the overall support system.

Authors' Contributions

SGL contributed to the design of study methodology, data curation and validation, conducting paper review protocols, writing, reviewing, and editing the paper, and project administration; AR contributed to conducting paper review protocols, extracting data, and writing the original paper; AA contributed to conducting paper review protocols, extracting data, data visualization, and writing the original paper; MM and RM contributed to conducting paper review protocols and extracting data; JHY contributed to conceptualization, paper review protocols, extracting data, writing the original paper, and providing supervision; and BPD contributed to conceptualization, paper review protocols, extracting data, writing, reviewing, and editing the paper, and providing supervision.

Conflicts of Interest

None declared.

References

1. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen HU. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
2. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch Gen Psychiatry 2003;60(8):837-844 [FREE Full text] [doi: 10.1001/archpsyc.60.8.837] [Medline: 12912767]
3. Supporting child and student social, emotional, behavioral, and mental health needs. U.S. Department of Education, Office of Special Education and Rehabilitative Services. 2021. URL: https://www2.ed.gov/documents/students/supporting-child-student-social-emotional-behavioral-mental-health.pdf [accessed 2023-07-26]
4. Borrelli B, Ritterband LM. Special issue on eHealth and mHealth: challenges and future directions for assessment, treatment, and dissemination. Health Psychol 2015;34S:1205-1208. [doi: 10.1037/hea0000323] [Medline: 26651461]
5. Gaglani SM, Topol EJ. iMedEd: the role of mobile health technologies in medical education. Acad Med 2014;89(9):1207-1209 [FREE Full text] [doi: 10.1097/ACM.0000000000000361] [Medline: 24892404]
6. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
7. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021;8(11):e31170 [FREE Full text] [doi: 10.2196/31170] [Medline: 34747713]
8. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
9. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
10. Gibson K, Cartwright C. Young people's experiences of mobile phone text counselling: balancing connection and control. Child Youth Serv Rev 2014;43:96-104. [doi: 10.1016/j.childyouth.2014.05.010]

11. Matthews M, Doherty G, Coyle D, Sharry J. Designing mobile applications to support mental health interventions. In: Lumsden J, editor. Handbook of Research on User Interface Design and Evaluation for Mobile Technology. Hershey, PA: IGI Global; 2008:635-656.
12. Lenhart A. Teens, social media & technology overview 2015. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/09/teens-social-media-technology-2015/ [accessed 2022-03-01]
13. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
14. Tang W, Kreindler D. Supporting homework compliance in cognitive behavioural therapy: essential features of mobile apps. JMIR Ment Health 2017;4(2):e20 [FREE Full text] [doi: 10.2196/mental.5283] [Medline: 28596145]
15. Tingley J, Greenhalgh I, Stallard P. Technology matters: BlueIce—using a smartphone app to beat adolescent self-harm. Child Adolesc Ment Health 2020;25(3):192-194. [doi: 10.1111/camh.12397] [Medline: 32526062]
16. Bear HA, Nunes LA, DeJesus J, Liverpool S, Moltrecht B, Neelakantan L, et al. Determination of markers of successful implementation of mental health apps for young people: systematic review. J Med Internet Res 2022;24(11):e40347 [FREE Full text] [doi: 10.2196/40347] [Medline: 36350704]
17. Aboujaoude E, Salame W. Technology at the service of pediatric mental health: review and assessment. J Pediatr 2016;171:20-24 [FREE Full text] [doi: 10.1016/j.jpeds.2015.12.009] [Medline: 26743496]
18. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems—a systematic and meta-review. J Child Psychol Psychiatry 2017;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
19. Melia R, Francis K, Hickey E, Bogue J, Duggan J, O'Sullivan M, et al. Mobile health technology interventions for suicide prevention: systematic review. JMIR Mhealth Uhealth 2020;8(1):e12516 [FREE Full text] [doi: 10.2196/12516] [Medline: 31939744]
20. Stallard P, Porter J, Grist R. A smartphone app (BlueIce) for young people who self-harm: open phase 1 pre-post trial. JMIR Mhealth Uhealth 2018;6(1):e32 [FREE Full text] [doi: 10.2196/mhealth.8917] [Medline: 29382632]
21. Berry RR, Lai B. The emerging role of technology in cognitive–behavioral therapy for anxious youth: a review. J Rat-Emo Cognitive-Behav Ther 2014;32(1):57-66. [doi: 10.1007/s10942-014-0184-5]
22. Wade R, Shea JA, Rubin D, Wood J. Adverse childhood experiences of low-income urban youth. Pediatrics 2014;134(1):e13-e20. [doi: 10.1542/peds.2013-2475] [Medline: 24935995]
23. de Leon Siantz ML, Keltner BR. Mental health and disabilities: what we know about racial and ethnic minority children. Annu Rev Nurs Res 2004;22:265-281. [Medline: 15368774]
24. Price MA, Hollinsaid NL. Future directions in mental health treatment with stigmatized youth. J Clin Child Adolesc Psychol 2022;51(5):810-825 [FREE Full text] [doi: 10.1080/15374416.2022.2109652] [Medline: 36007234]
25. Alegría M, Canino G, Ríos R, Vera M, Calderón J, Rusch D, et al. Inequalities in use of specialty mental health services among Latinos, African Americans, and non-Latino whites. Psychiatr Serv 2002;53(12):1547-1555 [FREE Full text] [doi: 10.1176/appi.ps.53.12.1547] [Medline: 12461214]
26. Cavazos-Rehg P, Min C, Fitzsimmons-Craft EE, Savoy B, Kaiser N, Riordan R, et al. Parental consent: a potential barrier for underage teens' participation in an mHealth mental health intervention. Internet Interv 2020;21:100328 [FREE Full text] [doi: 10.1016/j.invent.2020.100328] [Medline: 32528858]
27. Guo S, Kataoka SH, Bear L, Lau AS. Differences in school-based referrals for mental health care: understanding racial/ethnic disparities between Asian American and Latino youth. School Mental Health 2013;6(1):27-39. [doi: 10.1007/s12310-013-9108-2]
28. Reinke WM, Stormont M, Herman KC, Puri R, Goel N. Supporting children's mental health in schools: teacher perceptions of needs, roles, and barriers. Sch Psychol Q 2011;26(1):1-13. [doi: 10.1037/a0022714]
29. Kataoka S, Stein BD, Nadeem E, Wong M. Who gets care? Mental health service use following a school-based suicide prevention program. J Am Acad Child Adolesc Psychiatry 2007;46(10):1341-1348. [doi: 10.1097/chi.0b013e31813761fd] [Medline: 17885576]
30. Hall GCN, Ibaraki AY, Huang ER, Marti CN, Stice E. A meta-analysis of cultural adaptations of psychological interventions. Behav Ther 2016;47(6):993-1014. [doi: 10.1016/j.beth.2016.09.005] [Medline: 27993346]
31. Huey SJ, Polo AJ. Evidence-based psychosocial treatments for ethnic minority youth. J Clin Child Adolesc Psychol 2008;37(1):262-301 [FREE Full text] [doi: 10.1080/15374410701820174] [Medline: 18444061]
32. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
33. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
34. Reyes-Portillo JA, Mufson L, Greenhill LL, Gould MS, Fisher PW, Tarlow N, et al. Web-based interventions for youth internalizing problems: a systematic review. J Am Acad Child Adolesc Psychiatry 2014;53(12):1254-1270.e5 [FREE Full text] [doi: 10.1016/j.jaac.2014.09.005] [Medline: 25457924]

35. Whitney DG, Peterson MD. US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. JAMA Pediatr 2019;173(4):389-391 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.5399] [Medline: 30742204]
36. DSM-5 parent/guardian-rated level 1 cross-cutting symptom measure—child age 6–17. American Psychiatric Association. 2013. URL: https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/APA_DSM5_Level-1-Measure-Parent-Or-Guardian-Of-Child-Age-6-to-17.pdf [accessed 2023-07-25]
37. Escobedo L, Nguyen DH, Boyd L, Hirano S, Rangel A, Garcia-Rosas D, et al. MOSOCO: a mobile assistive tool to support children with autism practicing social skills in real-life situations. 2012 Presented at: CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, Texas, USA p. 2589-2598. [doi: 10.1145/2207676.2208649]
38. Hilt LM, Swords CM. Acceptability and preliminary effects of a mindfulness mobile application for ruminative adolescents. Behav Ther 2021;52(6):1339-1350 [FREE Full text] [doi: 10.1016/j.beth.2021.03.004] [Medline: 34656190]
39. Hourcade JP, Williams SR, Miller EA, Huebner KE, Liang LJ. Evaluation of tablet apps to encourage social interaction in children with autism spectrum disorders. 2013 Presented at: CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 27, 2013-May 2, 2013; Paris, France p. 3197-3206. [doi: 10.1145/2470654.2466438]
40. Lenhard F, Andersson E, Mataix-Cols D, Rück C, Vigerland S, Högström J, et al. Therapist-guided, internet-delivered cognitive-behavioral therapy for adolescents with obsessive-compulsive disorder: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2017;56(1):10-19.e2 [FREE Full text] [doi: 10.1016/j.jaac.2016.09.515] [Medline: 27993223]
41. Neal-Barnett A, Stadulis R, Ellzey D, Jean E, Rowell T, Somerville K, et al. Evaluation of the effectiveness of a musical cognitive restructuring app for Black inner-city girls: survey, usage, and focus group evaluation. JMIR Mhealth Uhealth 2019;7(6):e11310 [FREE Full text] [doi: 10.2196/11310] [Medline: 31188130]
42. O'Dea B, Han J, Batterham PJ, Achilles MR, Calear AL, Werner-Seidler A, et al. A randomised controlled trial of a relationship-focussed mobile phone application for improving adolescents' mental health. J Child Psychol Psychiatry 2020;61(8):899-913 [FREE Full text] [doi: 10.1111/jcpp.13294] [Medline: 32683737]
43. Silk JS, Pramana G, Sequeira SL, Lindhiem O, Kendall PC, Rosen D, et al. Using a smartphone app and clinician portal to enhance brief cognitive behavioral therapy for childhood anxiety disorders. Behav Ther 2020;51(1):69-84 [FREE Full text] [doi: 10.1016/j.beth.2019.05.002] [Medline: 32005341]
44. Whiteside SPH, Ale CM, Vickers Douglas K, Tiede MS, Dammann JE. Case examples of enhancing pediatric OCD treatment with a smartphone application. Clinic Case Stud 2013;13(1):80-94. [doi: 10.1177/1534650113504822]
45. Whiteside SPH, Biggs BK, Tiede MS, Dammann JE, Hathaway JC, Blasi ME, et al. An online- and mobile-based application to facilitate exposure for childhood anxiety disorders. Cogn Behav Pract 2019;26(3):478-491 [FREE Full text] [doi: 10.1016/j.cbpra.2019.01.002] [Medline: 33828399]
46. Frances A, Docherty JP, Kahn DA. Treatment of obsessive-compulsive disorder. J Clin Psychiatry 1997;58(4):5-72.
47. March JS. Multidimensional Anxiety Scale for Children (MASC2) 2nd Edition. Toronto: Multi-Health Systems Incorporated; 2013.
48. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
49. Punukollu M, Leighton EL, Brooks AF, Heron S, Mitchell F, Regener P, et al. SafeSpot: an innovative app and mental health support package for Scottish schools—a qualitative analysis as part of a mixed methods study. Child Adolesc Ment Health 2020;25(2):110-116 [FREE Full text] [doi: 10.1111/camh.12375] [Medline: 32307842]
50. Ghandour RM, Sherman LJ, Vladutiu CJ, Ali MM, Lynch SE, Bitsko RH, et al. Prevalence and treatment of depression, anxiety, and conduct problems in US children. J Pediatr 2019;206:256-267.e3 [FREE Full text] [doi: 10.1016/j.jpeds.2018.09.021] [Medline: 30322701]
51. Rideout VJ, Robb MB. The Common Sense Census: Media Use by Tweens and Teens. San Francisco, CA: Common Sense Media; 2019.
52. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
53. Device software functions including mobile medical applications. U.S. Food & Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/device-software-functions-including-mobile-medical-applications [accessed 2022-05-04]
54. Apps and digital health resources reviewed by experts. One Mind PsyberGuide. URL: https://onemindpsyberguide.org/ [accessed 2022-05-04]
55. The app evaluation model. American Psychiatric Association. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model [accessed 2022-05-04]
56. Temkin AB, Schild J, Falk A, Bennett SM. Mobile apps for youth anxiety disorders: a review of the evidence and forecast of future innovations. Prof Psychol Res Pract 2020;51(4):400-413. [doi: 10.1037/pro0000342]

57. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
58. Huh J, Koola J, Contreras A, Castillo AK, Ruiz M, Tedone KG, et al. Consumer health informatics adoption among underserved populations: thinking beyond the digital divide. Yearb Med Inform 2018;27(1):146-155 [FREE Full text] [doi: 10.1055/s-0038-1641217] [Medline: 30157518]
59. Stowell E, Lyson MC, Saksono H, Wurth RC, Jimison H, Pavel M, et al. Designing and evaluating mHealth interventions for vulnerable populations: a systematic review. 2018 Presented at: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC, Canada p. 1-17. [doi: 10.1145/3173574.3173589]
60. Mahmud A, Islam MR. Social stigma as a barrier to Covid-19 responses to community well-being in Bangladesh. Int J Comm WB 2021;4(3):315-321 [FREE Full text] [doi: 10.1007/s42413-020-00071-w] [Medline: 34723103]
61. Sanchez-Flack J, Buscemi J, O'Donnell A, Withington MHC, Fitzgibbon M. Black American and Latinx parent/caregiver participation in digital health obesity interventions for children: a systematic review. Front Digit Health 2021;3:687648 [FREE Full text] [doi: 10.3389/fdgth.2021.687648] [Medline: 34713158]
62. Jackson DN, Trivedi N, Baur C. Re-prioritizing digital health and health literacy in to affect health equity. Health Commun 2021;36(10):1155-1162. [doi: 10.1080/10410236.2020.1748828] [Medline: 32354233]
63. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
64. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A "SMART" design for building individualized treatment sequences. Annu Rev Clin Psychol 2012;8:21-48 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032511-143152] [Medline: 22224838]
65. Lyon AR, Munson SA, Renn BN, Atkins DC, Pullmann MD, Friedman E, et al. Use of human-centered design to improve implementation of evidence-based psychotherapies in low-resource communities: protocol for studies applying a framework to assess usability. JMIR Res Protoc 2019;8(10):e14990 [FREE Full text] [doi: 10.2196/14990] [Medline: 31599736]
66. Chen E, Leos C, Kowitt SD, Moracco KE. Enhancing community-based participatory research through human-centered design strategies. Health Promot Pract 2020;21(1):37-48 [FREE Full text] [doi: 10.1177/1524839919850557] [Medline: 31131633]
67. Kia-Keating M, Santacrose DE, Liu SR, Adams J. Using community-based participatory research and human-centered design to address violence-related health disparities among Latino/a youth. Fam Community Health 2017;40(2):160-169 [FREE Full text] [doi: 10.1097/FCH.0000000000000145] [Medline: 28207679]
68. UyBico SJ, Pavel S, Gross CP. Recruiting vulnerable populations into research: a systematic review of recruitment interventions. J Gen Intern Med 2007;22(6):852-863 [FREE Full text] [doi: 10.1007/s11606-007-0126-3] [Medline: 17375358]
69. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-being interventions. JMIR Hum Factors 2015;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]
70. Ospina-Pinillos L, Davenport T, Diaz AM, Navarro-Mancilla A, Scott EM, Hickie IB. Using participatory design methodologies to co-design and culturally adapt the Spanish version of the mental health eClinic: qualitative study. J Med Internet Res 2019;21(8):e14127 [FREE Full text] [doi: 10.2196/14127] [Medline: 31376271]
71. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a mental health eClinic to improve access to and quality of mental health care for young people: using participatory design as research methodologies. J Med Internet Res 2018;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
72. Strang JF, Knauss M, van der Miesen A, McGuire JK, Kenworthy L, Caplan R, et al. A clinical program for transgender and gender-diverse neurodiverse/autistic adolescents developed through community-based participatory design. J Clin Child Adolesc Psychol 2021;50(6):730-745. [doi: 10.1080/15374416.2020.1731817] [Medline: 32375521]
73. Corbie-Smith G, Thomas SB, St George DMM. Distrust, race, and research. Arch Intern Med 2002;162(21):2458-6243 [FREE Full text] [doi: 10.1001/archinte.162.21.2458] [Medline: 12437405]
74. Armaou M, Araviaki E, Musikanski L. eHealth and mHealth interventions for ethnic minority and historically underserved populations in developed countries: an umbrella review. Int Journal Com WB 2019;3(2):193-221 [FREE Full text] [doi: 10.1007/s42413-019-00055-5]
75. Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. JMIR Mhealth Uhealth 2016;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
76. Merikangas KR, He JP, Burstein M, Swendsen J, Avenevoli S, Case B, et al. Service utilization for lifetime mental disorders in U.S. adolescents: results of the National Comorbidity Survey-Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 2011;50(1):32-45 [FREE Full text] [doi: 10.1016/j.jaac.2010.10.006] [Medline: 21156268]
77. Sadler K, Vizard T, Ford T, Goodman A, Goodman R, McManus S. Mental health of children and young people in England, 2017: trends and characteristics. NHS Digital. 2018. URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017 [accessed 2023-07-26]

78. Reardon T, Harvey K, Baranowska M, O'Brien D, Smith L, Creswell C. What do parents perceive are the barriers and facilitators to accessing psychological treatment for mental health problems in children and adolescents? A systematic review of qualitative and quantitative studies. Eur Child Adolesc Psychiatry 2017;26(6):623-647 [FREE Full text] [doi: 10.1007/s00787-016-0930-6] [Medline: 28054223]

Abbreviations

ASD: autism spectrum disorder
BiP OCD: BarnInternetProjektet for obsessive-compulsive disorder
BYOTS: Build Your Own Theme Song
CBT: cognitive behavioral therapy
MASC: Multidimensional Anxiety Scale for Children
mHealth: mobile health
OCD: obsessive-compulsive disorder
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by J Torous; submitted 02.03.23; peer-reviewed by A AL-Asadi, M O'Sullivan; comments to author 19.05.23; revised version received 01.07.23; accepted 07.07.23; published 23.08.23.

Please cite as:
Litke SG, Resnikoff A, Anil A, Montgomery M, Matta R, Huh-Yoo J, Daly BP
Mobile Technologies for Supporting Mental Health in Youths: Scoping Review of Effectiveness, Limitations, and Inclusivity
JMIR Ment Health 2023;10:e46949
URL: https://mental.jmir.org/2023/1/e46949
doi:10.2196/46949
PMID:37610818

Review

Efficacy, Safety, and Evaluation Criteria of mHealth Interventions for Depression: Systematic Review

Andrea Duarte-Díaz[1,2,3], MD; Lilisbeth Perestelo-Pérez[2,3,4], PhD; Estel Gelabert[5], PhD; Noemí Robles[2,6], PhD; Antoni Pérez-Navarro[7,8], PhD; Josep Vidal-Alaball[9,10,11], PhD; Oriol Solà-Morales[12,13,14], PhD; Ariadna Sales Masnou[15], MD; Carme Carrion[2,16,17], PhD

[1]Canary Islands Health Research Institute Foundation (FIISC), El Rosario, Spain

[2]Network for Research on Chronicity, Primary Care and Health Promotion (RICAPPS), Madrid, Spain

[3]The Spanish Network of Agencies for Health Technology Assessment and Services of the National Health System (RedETS), Madrid, Spain

[4]Evaluation Unit (SESCS), Canary Islands Health Service (SCS), El Rosario, Spain

[5]Department of Clinical and Health Psychology, Universitat Autònoma de Barcelona, Bellatera (Barcelona), Spain

[6]eHealth Center, Universitat Oberta de Catalunya (UOC), Barcelona, Spain

[7]Faculty of Computer Sciences, Multimedia and Telecommunication, Universitat Oberta de Catalunya (UOC), Barcelona, Spain

[8]eHealth Lab Research Group, Universitat Oberta de Catalunya (UOC), Barcelona, Spain

[9]Unitat de Suport a la Recerca de la Catalunya Central, Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina, Sant Fruitós de Bages, Spain

[10]Health Promotion in Rural Areas Research Group, Gerencia Territorial de la Catalunya Central, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[11]Faculty of Medicine, University of Vic-Central University of Catalonia, Vic, Spain

[12]Fundació HiTT, Barcelona, Spain

[13]Universitat Internacional de Catalunya (UIC), Barcelona, Spain

[14]Office of Health Economics (OHE), London, United Kingdom

[15]Estudis de Ciències de la Salut, Universitat Oberta de Catalunya (UOC), Barcelona, Spain

[16]eHealth Lab Research Group, School of Health Sciences and eHealth Center, Universitat Oberta de Catalunya (UOC), Barcelona, Spain

[17]School of Medicine, Universitat de Girona (UdG), Girona, Spain

Corresponding Author:
Andrea Duarte-Díaz, MD
Canary Islands Health Research Institute Foundation (FIISC)
Camino Candelaria, 44
El Rosario, 38109
Spain
Phone: 34 922478320
Email: andrea.duartediaz@sescs.es

Abstract

Background: Depression is a significant public health issue that can lead to considerable disability and reduced quality of life. With the rise of technology, mobile health (mHealth) interventions, particularly smartphone apps, are emerging as a promising approach for addressing depression. However, the lack of standardized evaluation tools and evidence-based principles for these interventions remains a concern.

Objective: In this systematic review and meta-analysis, we aimed to evaluate the efficacy and safety of mHealth interventions for depression and identify the criteria and evaluation tools used for their assessment.

Methods: A systematic review and meta-analysis of the literature was carried out following the recommendations of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement. Studies that recruited adult patients exhibiting elevated depressive symptoms or those diagnosed with depressive disorders and aimed to assess the effectiveness or safety of mHealth interventions were eligible for consideration. The primary outcome of interest was the reduction of depressive symptoms, and only randomized controlled trials (RCTs) were included in the analysis. The risk of bias in the original RCTs was assessed using version 2 of the Cochrane risk-of-bias tool for randomized trials.

Results: A total of 29 RCTs were included in the analysis after a comprehensive search of electronic databases and manual searches. The efficacy of mHealth interventions in reducing depressive symptoms was assessed using a random effects meta-analysis. In total, 20 RCTs had an unclear risk of bias and 9 were assessed as having a high risk of bias. The most common element in mHealth interventions was psychoeducation, followed by goal setting and gamification strategies. The meta-analysis revealed a significant effect for mHealth interventions in reducing depressive symptoms compared with nonactive control (Hedges g=−0.62, 95% CI −0.87 to −0.37, I^2=87%). Hybrid interventions that combined mHealth with face-to-face sessions were found to be the most effective. Three studies compared mHealth interventions with active controls and reported overall positive results. Safety analyses showed that most studies did not report any study-related adverse events.

Conclusions: This review suggests that mHealth interventions can be effective in reducing depressive symptoms, with hybrid interventions achieving the best results. However, the high level of heterogeneity in the characteristics and components of mHealth interventions indicates the need for personalized approaches that consider individual differences, preferences, and needs. It is also important to prioritize evidence-based principles and standardized evaluation tools for mHealth interventions to ensure their efficacy and safety in the treatment of depression. Overall, the findings of this study support the use of mHealth interventions as a viable method for delivering mental health care.

Trial Registration: PROSPERO CRD42022304684; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=304684

(JMIR Ment Health 2023;10:e46877) doi:10.2196/46877

KEYWORDS

mobile health; mHealth; apps; depression; systematic review; meta-analysis

Introduction

Background

Depression is the most common mental health condition in the general population and is one of the leading causes of the global burden of disease and disability [1-3]. The worldwide incidence of depression increased by 49.86% between 1990 and 2017, from 172 million cases to 25.8 million [2]. Unipolar depression is predicted to be the leading cause of disability in high-income countries by 2030, surpassing other health conditions such as ischemic heart disease, dementia, alcohol use disorders, and diabetes [3].

Although there is strong clinical evidence that depression can be treated with a variety of pharmacological and psychological interventions [4], human resources for mental health are inadequate, especially in low- and middle-income countries [5-7], and a global shortage of over 15 million health workers is expected by 2030 [8]. Given the rapid advancement and adoption of technology, digital interventions—particularly mobile health (mHealth) interventions—have the potential to provide novel and viable methods of delivering population-scale mental health care [9,10].

The World Health Organization defines mHealth as “the term used for medical and public health practices supported by mobile devices, such as phones, patient monitoring devices, personal digital assistants, and other wireless devices” [11]. Smartphone apps can especially be powerful vectors for mHealth interventions because of their high connectivity, 24-hour availability, and ubiquitous nature [12]. Compared with most traditional treatment services, smartphone-based interventions offer several advantages, including high accessibility and scalability; relatively low costs; minimal contact; patient anonymity; flexibility of use; and the possibility of self-monitoring activity, symptoms, and progression in real time as well as providing motivational support and targeted care [10,13,14].

Self-management features are commonly found in mHealth interventions aimed at mental health problems, enabling clients to manage symptoms by monitoring their own symptoms and behavior [15]. In addition, mHealth apps for mood disorder management often provide stress-relieving games, meditation instructions, mood trackers, and psychoeducational materials. Despite the abundance of apps available in the commercial market for managing depressive symptoms, only a limited number incorporate a cognitive behavioral therapy (CBT) approach, despite CBT being widely recognized as a first-line psychological treatment [16].

Previous systematic reviews and meta-analyses have shown that smartphone-based interventions can have beneficial effects on clinical and nonclinical depressive symptoms in both general and clinical populations [9,17]. Moreover, digital interventions have been shown to be particularly effective, acceptable, feasible, and user friendly when embedded in a therapeutic context involving social interaction with mental health professionals to monitor progress and provide additional support [18]. A recent meta-review of meta-analyses concluded that apps for anxiety and depression produce definite clinical benefits, whether used for self-management or alongside professional guidance [12].

Several mHealth apps are currently available [19-21]. However, despite increased interest and use, no international standards or apps exist to evaluate mHealth apps in a simple and effective manner. Furthermore, although the number of mobile mental health apps is increasing owing to their convenience and high demand, many of these apps do not apply evidence-based principles or have not been tested for efficacy [16,22]. Therefore, selecting an app that is likely to be effective is problematic for users [9]. Health professionals and services are also increasingly using digital tools to facilitate disease management and need to be sure that the apps they recommend meet the minimum quality requirements [23]. Although several initiatives have been launched to define how mHealth apps

should be assessed, these initiatives only address a part of the evaluation process and are mostly concerned with developing a methodology for evaluating all types of mHealth apps. As every health condition has specific needs, new tools and methodologies are required to evaluate apps targeting each condition.

Objectives

This systematic review is part of the EvalDepApps research project [24], the primary objective of which is to develop and pilot an assessment tool for mobile apps aimed at treating and monitoring people with depressive symptoms. To that end, it is critical to comprehensively understand the effectiveness and safety of mHealth interventions based on available scientific evidence as well as the evaluation criteria used to measure these outcomes. Accordingly, the aims of this systematic review are (1) to assess how effective and safe mHealth interventions are in the treatment of depression and (2) to identify the criteria and evaluation tools used to assess these mHealth interventions.

Methods

A systematic review and meta-analysis of the literature was performed following recommendations in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [25] (Multimedia Appendix 1). The protocol for this systematic review and meta-analysis was prospectively registered on PROSPERO on February 19, 2022 (CRD42022304684).

Search Strategy

A scoping search conducted to identify relevant search terms resulted in the following: “apps,” “mHealth,” “eHealth,” and “depression.” These were applied individually or combined according to Medical Subject Headings keyword terms in 3 electronic databases from inception to February 2022: MEDLINE, PsycINFO, and Embase. In addition, the reference lists of all eligible studies were screened to identify additional studies meeting the inclusion criteria.

Inclusion and Exclusion Criteria

We considered studies recruiting adult patients with elevated depressive symptoms (ie, scoring above the cutoff criteria on a validated depression screening instrument) or diagnosed with depressive disorder (ie, diagnosed by a clinician or using any recognized diagnostic criteria). Studies recruiting children or adolescents aged ≤18 years were excluded. Studies assessing the effectiveness or safety of mHealth-based interventions for treating depression were included, whereas those using no mobile tools or relating to diagnosis or prevention were excluded. Studies referring to the management of other conditions, such as cancer, stroke, Alzheimer disease, epilepsy, social anxiety, alcoholism, or pain, were also excluded. Any comparator other than mHealth interventions was considered, including passive (eg, no intervention or waiting list) or active (eg, antidepressants or face-to-face psychotherapy) groups. The primary outcome was the reduction of depressive symptoms, and secondary outcomes included undesirable effects of the mHealth intervention and the criteria and evaluation tools used to assess the effectiveness and safety of mHealth interventions. Randomized controlled trials (RCTs) with at least 10 participants were included in the study design. Nonrandomized studies, uncontrolled studies, observational studies, conference abstracts, letters, commentaries, essays, book chapters, qualitative studies, study protocols, and reviews were excluded. We included studies published in English and Spanish, without imposing any restrictions on the publication year. Studies conducted in any country and clinical setting were considered.

Risk-of-Bias Assessment

The risk of bias in the original RCTs was assessed using version 2 of the Cochrane risk-of-bias tool for randomized trials [26]. Quality assessment was performed by 2 independent reviewers, and any disagreements were resolved by consulting a third reviewer.

Study Selection and Data Extraction

All citations extracted from electronic databases were imported into Rayyan, a web-based software program for systematic reviews, and duplicates were removed. Two members of the research team independently reviewed all titles and abstracts to preselect those systematic reviews meeting the inclusion criteria. The full texts of potentially relevant studies were screened for eligibility by 2 reviewers. Any disagreement was resolved by discussion and consensus, and a third reviewer was consulted, if required. Two reviewers then independently extracted data from each included RCT using a standardized data extraction form in Microsoft Excel using the following variables: (1) first author, (2) year of publication, (3) country, (4) number of participants, (5) study design, (6) study period, (7) study population, (8) intervention and control details, (9) outcome measures, and (10) main results. To gather information about the intervention details and elements included, we primarily relied on the descriptions of the interventions provided in the included studies. Furthermore, we also referred to other publications related to the same study, which offered a more comprehensive description of the intervention’s development process. In addition, when necessary, we consulted public descriptions available through websites or app stores.

Data Synthesis and Analysis

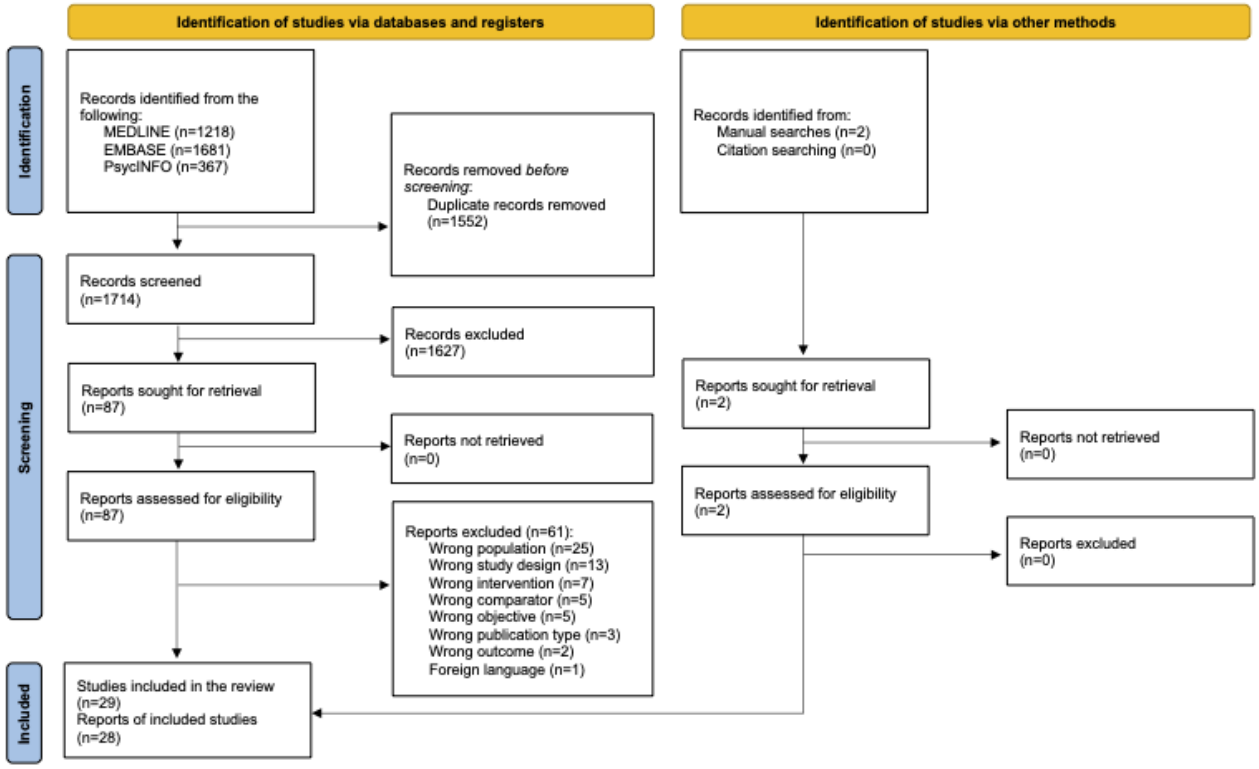
Meta-analyses were performed using the inverse variance method [27] and were visually displayed using forest plots. A random effects model using the Sidik-Jonkman method as the tau estimator was applied [28]. Statistical heterogeneity between the different studies included in the meta-analyses was assessed using the Higgins I^2 value [29]. For each meta-analysis, 2-tailed 95% prediction intervals were calculated. The following post hoc subgroup analyses were carried out: type of nonactive control, intervention length, depression severity at baseline, mHealth intervention framework, delivery mode, mood monitoring, goal setting, and gamification. Furthermore, the Galbraith plot was used to identify possible outliers, and a sensitivity analysis was performed using the leave-one-out function, which performs multiple meta-analyses excluding a single study at a time. We evaluated the publication bias using the Egger test [30], and the trim-and-fill method was used to correct for possible funnel plot asymmetry. All analyses were performed using Stata (version 17; StataCorp).

Results

Overview

The initial search of the electronic databases yielded 3203 references. After removing duplicates, 1714 records were screened by title and abstract and 87 full-text articles were assessed for eligibility. Two additional records were identified through manual searches. Finally, 29 RCTs reported in 28 articles were included [31-58]. A flowchart of our selection process is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection process.



Characteristics of the Included Studies

The study included a total sample of 5594 participants, with an average age of 41.33 years and a majority of participants identifying as female (72%). Most studies were performed in Asia (8/29, 28%) [33-36,41,50,53,58], Europe (8/29, 28%) [37-40,44,51,52,56], and North America (8/29, 28%) [32,42,43,45,46,54,55,57], followed by Australia (3/29, 10%) [46,48,49] and South America (2/29, 7%) [31]. In most studies (18/29, 62%), participants had moderate depressive symptoms at baseline. The intervention period of the included studies ranged from 2 to 24 weeks, with an average of 8 weeks. A complete description of study characteristics is presented in Table 1.

Table 1. Characteristics of the included studies.

Study, year	Country	Sample size (intervention/control)[a]	Age (years), mean	Gender (women), %	Depression at baseline	Intervention	Control	Length (wk)
Araya et al [31], 2021	Brazil	880 (440/440)	56	86.5	Moderately severe depressive symptoms (PHQ-9[b])	• CONEMO	TAU[c]	6
Araya et al [31], 2021	Peru	432 (217/215)	59.7	81.5	Moderate depressive symptoms (PHQ-9)	• CONEMO	TAU	6
Arean et al [32], 2016	United States	626 (211/209/206)	33.9	78.9	Moderate depressive symptoms (PHQ-9)	• EVO • Problem-Solving Therapy App (iPST)	Minimal intervention (health information)	4
Birney et al [43], 2016	United States	300 (150/150)	40.7	76.7	Moderate depressive symptoms (PHQ-9)	• MoodHacker	Minimal intervention (health information)	6
Bruhns et al [52], 2021	Germany	423 (208/215)	23.0	78.5	Moderate depressive symptoms (PHQ-9)	• Metacognitive Training (MCT) and more	Waiting list	4
Chan et al [53], 2021	Hong Kong	320 (167/153)	27.3	72.7	Severe depressive symptoms (CES-D[d])	• proACT-S	Waiting list	6
Dahne et al [54], 2019	United States	52 (24/19/9)	43.8	84.6	Moderately severe depressive symptoms (BDI-II[e])	• Moodivate • MoodKit	TAU	8
Dahne et al [55], 2019	United States	42 (22/9/11)	36.1	66.7	Moderately severe depressive symptoms (BDI-II)	• Aptívate • iCouch CBT[f]	TAU	8
Ebert et al [56], 2018	Germany	204 (102/102)	44.2	80.4	Severe depressive symptoms (CES-D)	• GET.ON Mood Enhancer	Waiting list	12
Graham et al [57], 2020	United States	146 (74/72)	42.3	82	Moderate depressive symptoms (PHQ-9)	• IntelliCare	Waiting list	8
Guo et al [58], 2020	China	300 (150/150)	28.3	7.7	Moderate depressive symptoms (PHQ-9)	• Run4Love (WeChat)	TAU	12
Ham et al [33], 2019	South Korea	80 (28/26/26)	44.2	85.7	Moderate depressive symptoms (BDI-II)	• HARUToday (CBT) • HARUToday (general)	Waiting list	10
Jannati et al [34], 2020	Iran	78 (39/39)	27	100	Moderate depressive symptoms (EPDS[g])	• Happy Mom	Waiting list	8
Kageyama et al [35], 2021	Japan	32 (16/16)	20.1	34.4	Moderate depressive symptoms (CES-D)	• SPSRS	Waiting list	5
Lüdtke et al [37], 2018	Germany	90 (45/45)	42.9	78.4	Moderate depressive symptoms (PHQ-9)	• Be Good to Yourself	Waiting list	4
Liu et al [36], 2022	China	83 (41/42)	23.1	55.4	Moderate depressive symptoms (PHQ-9)	• XiaoNan (WeChat)	Bibliotherapy	16
Lukas et al [38], 2021	Germany	16 (5/11)	24.7	81	Moderately severe depressive symptoms (PHQ-9)	• MT-Phoenix+face-to-face psychoeducation	Waiting list	2

Study, year	Country	Sample size (intervention/control)[a]	Age (years), mean	Gender (women), %	Depression at baseline	Intervention	Control	Length (wk)
Lukas et al [39], 2021	Germany	77 (40/37)	29.9	82	Moderate depressive symptoms (PHQ-9)	• MT-Phoenix+face-to-face psychoeducation	Waiting list	2
Ly et al [40],2015	Sweden	93 (46/47)	30.6	69.9	Moderately severe depressive symptoms (PHQ-9)	• Face-to-face behavioral activation (4 sessions) + smartphone app	Face-to-face behavioral activation (10 sessions)	10
Mantani et al [41], 2017	Japan	164 (81/83)	40.9	53.5	Moderate depressive symptoms (PHQ-9)	• Kokoro-App+TAU	TAU	9
Pratap et al [42], 2018	United States	274 (112/83/79)	34.9	77.1	Moderate depressive symptoms (PHQ-9)	• EVO • iPST	Minimal intervention (health tips)	4
Raevuori et al [44], 2021	Finland	124 (63/61)	25.1	72.6	Moderate depressive symptoms (PHQ-9)	• Meru Health Program + TAU	TAU	8
Roepke et al [45], 2015	United States	283 (93/97/93)	40.2	69.6	Severe depressive symptoms (CES-D)	• SuperBetter • CBT-PPT SuperBetter	Waiting list	4
Sawyer et al [46], 2019	Australia	133 (72/61)	31.1	100	Moderate depressive symptoms (EPDS)	• eMums Plus	TAU	16
Stiles-Shields et al [47], 2019	United States	30 (10/10/10)	NR[h]	NR	Moderately severe depressive symptoms (PHQ-9)	• Boost Me • Thought Challenger	Waiting list	6
Tighe et al [48] (2017)	Australia	61 (31/30)	26.5	64	Moderately severe depressive symptoms (PHQ-9)	• Ibobbly	Waiting list	6
Tønning et al [51], 2021	Denmark	120 (61/59)	43.9	31.6	Mild depressive symptoms (HDRS-17[i])	• MONSENSO	TAU	24
Watts et al [49], 2013	Australia	52 (22/30)	41	80	Moderate depressive symptoms (PHQ-9)	• Mobile Therapy (Get Happy Program)	Computer therapy	8
Wong et al [50], 2021	China	79 (39/40)	32.9	84.8	Moderate depressive symptoms (PHQ-9)	• Lifestyle Hub	Waiting list	9

[a]If there are 3 numbers, the first 2 numbers are intervention groups and the third one is the control group.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]TAU: treatment as usual.

[d]CES-D: Center for Epidemiologic Studies Depression Scale.

[e]BDI-II: Beck Depression Inventory–II.

[f]CBT: cognitive behavioral therapy.

[g]EPDS: Edinburgh Postnatal Depression Scale.

[h]NR: not reported.

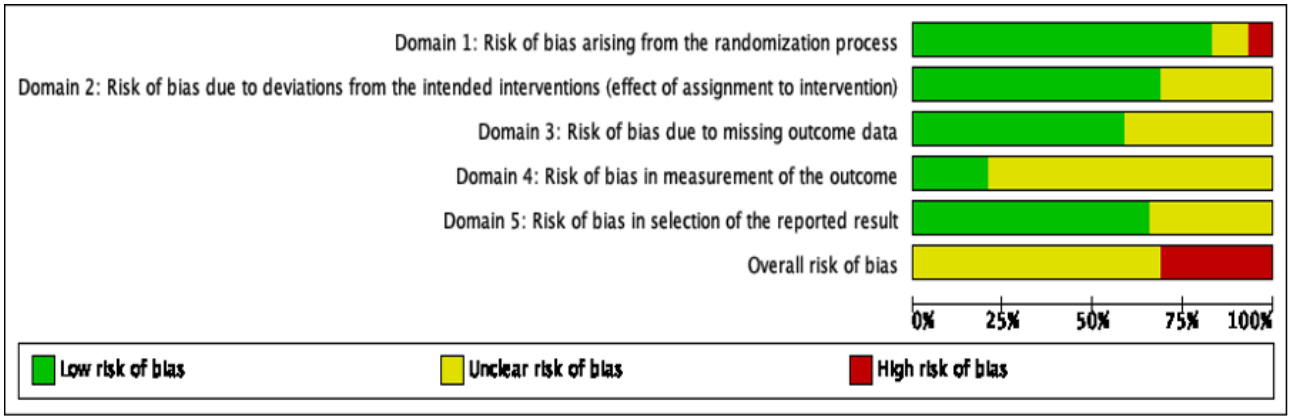
[i]HDRS-17: 17-item Hamilton Depression Rating Scale.

Quality Assessment of the Included Studies

In total, 20 RCTs were identified as having an unclear risk of bias [31,32,34-36,40-42,44,46-53,56-58] and the overall risk of bias in the remaining 9 RCTs was assessed as high [33,37-39,43,45,49,54,55]. Depression symptoms were self-reported, and participants were mostly unblinded; therefore, the main sources of bias were the methods used to assess outcomes. A total of 12 RCTs [31,33,38,39,42,43,49-53,57,58] were judged to have an unclear risk of bias owing to missing outcome data. Most of the studies described treatment allocation as random, but 5 studies [38,39,43,54,55] did not provide

enough details on the methods used to generate or conceal the sequence. Blinding is difficult with psychological mHealth interventions as participants are likely to be aware of what they are receiving. Nine studies reported in 8 references [31,33,42,45,47,49,54,55] that did not provide enough information about blinding or the method used to estimate the effect of assignment on the intervention were deemed to have unclear risks of bias due to deviations from the intended interventions. Most studies were reported in accordance with a prespecified plan and judged as having a low risk of bias in the selection of the reported result. A summary of the evaluation of risk of bias for each study is presented in Figure 2 in the form of a risk-of-bias graph with the opinions of review authors about each risk-of-bias item presented as percentages across all included studies.

Figure 2. Risk-of-bias graph.



Elements Included in the mHealth Interventions

Psychoeducation about depression (17/31, 54%) was the most common element included in the mHealth interventions through videos, informative sheets, and even chatbots. The capability to set goals (13/31, 41%) and gamified strategies such as reward systems, challenges, and badges (12/31, 38%) were also prevalent. A substantial number of mHealth interventions (12/31, 38%) enabled users to receive professional support if required, and several interventions provided feedback on progress (10/31, 32%) and self-monitoring of mood (10/31, 32%). However, only 2 mHealth interventions (2/31, 6%) included components that encouraged user interaction through forums, chats, and other means. It is worth mentioning that most mHealth interventions (20/31, 64%) used regular reminders to increase adherence, retention, and ultimately effectiveness in reducing depressive symptoms. The specific characteristics and elements included in each mHealth intervention are presented in Multimedia Appendix 2 [31-58].

Effectiveness of mHealth Interventions

mHealth Versus Nonactive Control

The meta-analysis of the effectiveness of mHealth for reducing depressive symptoms compared with nonactive controls included 26 effect sizes from 22 RCTs:16 compared mHealth with waiting list (n=1354), 6 with minimal intervention (n=145), and 4 with treatment as usual (TAU; n=620). The random effects meta-analysis showed a significant effect for mHealth (Hedges g=−0.62, 95% CI −0.87 to −0.37; P<.001; Figure 3). Heterogeneity across studies was high and statistically significant (I^2=87%, Q=131.08, P=.001).

A subgroup analysis by type of nonactive control was not statistically significant (P=.12). However, the effect was higher in those studies comparing mHealth with minimal intervention or waiting list than in those comparing with TAU. A subgroup analysis by the severity of depressive symptoms at baseline was not statistically significant, although the effect was higher in people with moderately severe and severe depressive symptoms than in those with moderate depressive symptoms. Similarly, a univariate meta-regression using the baseline Patient Health Questionnaire-9 score as a moderator also displayed a trend toward significance, suggesting that people with higher depressive symptoms would benefit more from mHealth interventions (β=−.15, P=.08, k=14). Neither age nor gender was found to be significantly associated with higher effectiveness. In a subgroup analysis using the mHealth content framework, there were no statistically significant differences (P=.73), but CBT-based interventions were the most effective for reducing depressive symptoms, followed by acceptance-based interventions. Regarding the characteristics of mHealth interventions, only subgroup analysis by delivery mode was statistically significant (P=.03), with hybrid interventions—those combining mHealth with face-to-face sessions—showing the highest effect on reducing depressive symptoms. Univariate meta-regression by number of elements in the mHealth intervention was not statistically significant. More details on the subgroup analyses performed are presented in Table 2.

The funnel plot was symmetrical (Figure 4), trim-and-fill did not need to impute any additional study, and Egger tests showed no evidence of a small-study effect (P=.17).

The leave-one-out analyses suggest that the findings are robust, and neither the direction nor significance of the pooled effect changed after excluding any single study (Multimedia Appendix 3 [33-39,41,43-50,52-58]). However, the Galbraith plot identified 8 outliers that may have contributed to heterogeneity (Multimedia Appendix 4). The subsequent exclusion of outliers yielded a slightly lower pooled effect (Hedges g=−0.54, 95% CI −0.74 to −0.34, k=18), and heterogeneity was nonsignificant

(I^2=51%, Q=20.22, P=.26). None of the subgroup analyses or meta-regressions changed after the exclusion of outliers.

Four studies reported in 3 articles [31,32,43] compared mHealth interventions against nonactive controls but did not provide means and SDs and therefore were not included in the meta-analysis. In the 2 RCTs reported in the study by Araya et al [31], a digital intervention delivered over a 6-week period significantly improved depressive symptoms at 3 months when compared with usual care, but the magnitude of the effect was small in 1 trial, and the effects were not sustained at 6 months. According to Arean et al [32], mHealth apps designed to engage the cognitive correlates of depression have the greatest effect on reducing depressed mood in people with moderate levels of depression. In addition, Birney et al [43] found that the MoodHacker app produced significant effects on depression symptoms at the 6-week follow-up when compared with minimal intervention.

Figure 3. Random effects meta-analysis (mobile health vs nonactive control).

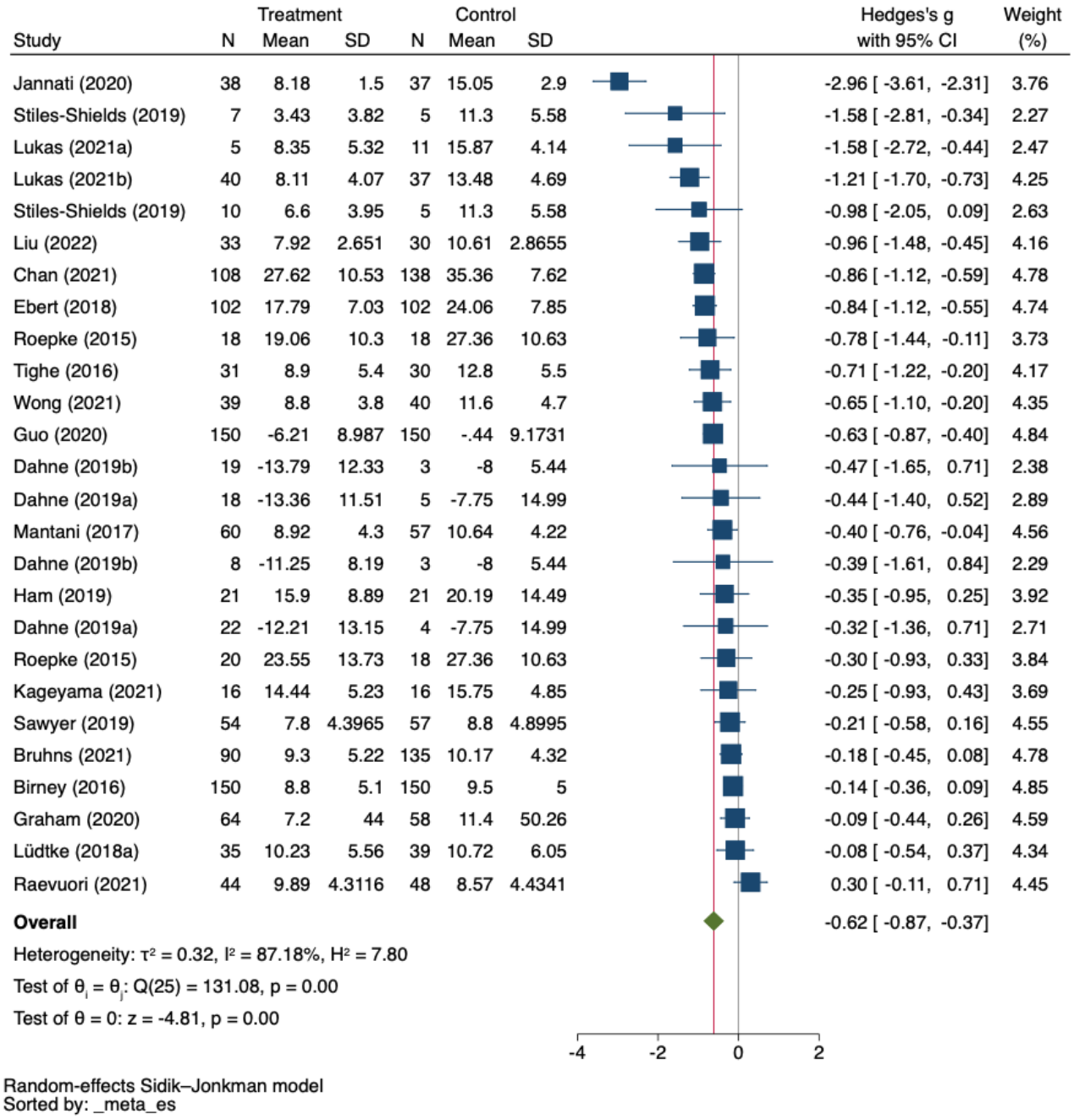


Table 2. Random effects models and subgroup analyses with depressive symptoms as the outcome.

Group	k	Hedges *g* (95% CI)	I^2 (%)	Test for subgroup differences
Type of nonactive control				Q=4.19; *P*=.12
Minimal intervention	6	−0.41 (−0.77 to −0.09)	27.45	
TAU[a]	4	−0.26 (−0.64 to 0.13)	81.0	
Waiting list	16	−0.79 (−1.15 to −0.42)	89.17	
Intervention length (weeks)				Q=0.38; *P*=.54
2-8	20	−0.66 (−0.98 to −0.33)	88.71	
9-16	6	−0.53 (−0.76 to −0.30)	49.73	
Depression severity at baseline				Q=0.95; *P*=.62
Moderate	14	−0.57 (−0.94 to −0.14)	93.81	
Moderately severe	8	−0.77 (−1.19 to −0.35)	31.06	
Severe	4	−0.77 (−1.02 to −0.51)	40.89	
Measure				Q=2.27; *P*=.52
BDI-II[b]	5	−0.38 (−0.78 to 0.02)	0.00	
CES-D[c]	5	−0.69 (−0.69 to −0.42)	48.83	
EPDS[d]	2	−1.57 (−4.24 to 1.09)	98.02	
PHQ-9[e]	14	−0.53 (−0.53 to −0.24)	85.09	
Framework				Q=2.02; *P*=.73
Acceptance based	2	−0.73 (−1.14 to −0.33)	0.03	
BA[f]	3	−0.59 (−1.25 to 0.06)	7.48	
CBT[g] only	10	−0.76 (−1.27 to −0.26)	94.56	
CBT and others	8	−0.38 (−0.78 to 0.02)	85.07	
Other	3	−0.70 (−1.40 to 0.01)	64.83	
Delivery mode				Q=6.87; *P*=.03
App only	22	−0.57 (−0.85 to −0.30)	86.04	
Hybrid intervention	2	−1.28 (−1.75 to −0.80)	4.47	
Web and app	2	−0.48 (−1.15 to 0.19)	92.57	
Psychoeducation				Q=0.75; *P*=.39
Yes	14	−0.69 (−1.12 to −0.27)	93.24	
No	12	−0.48 (−0.71 to −0.25)	48.78	
Mood monitoring				Q=2.39; *P*=.12
Yes	8	−0.37 (−0.59 to −0.15)	23.49	
No	18	−0.69 (−1.04 to −0.35)	91.33	
In-app feedback				Q=0.97; *P*=.33
Yes	8	−0.76 (−1.01 to −0.50)	70.56	
No	18	−0.54 (−0.89 to −0.20)	87.62	
Setting goals				Q=0.13; *P*=.72
Yes	10	−0.67 (−1.18 to −0.17)	95.37	
No	16	−0.57 (−0.84 to −0.30)	68.12	
Gamification				Q=0.46; *P*=.50
Yes	11	−0.52 (−0.79 to −0.25)	62.87	

Group	k	Hedges *g* (95% CI)	I^2 (%)	Test for subgroup differences
No	15	−0.68 (−1.07 to −0.29)	92.37	
Professional support				Q=0.02; *P*=.89
Yes	9	−0.64 (−1.06 to −0.22)	87.81	
No	17	−0.60 (−0.92 to −0.28)	86.61	

[a]TAU: treatment as usual.

[b]BDI-II: Beck Depression Inventory–II.

[c]CES-D: Center for Epidemiologic Studies Depression Scale.

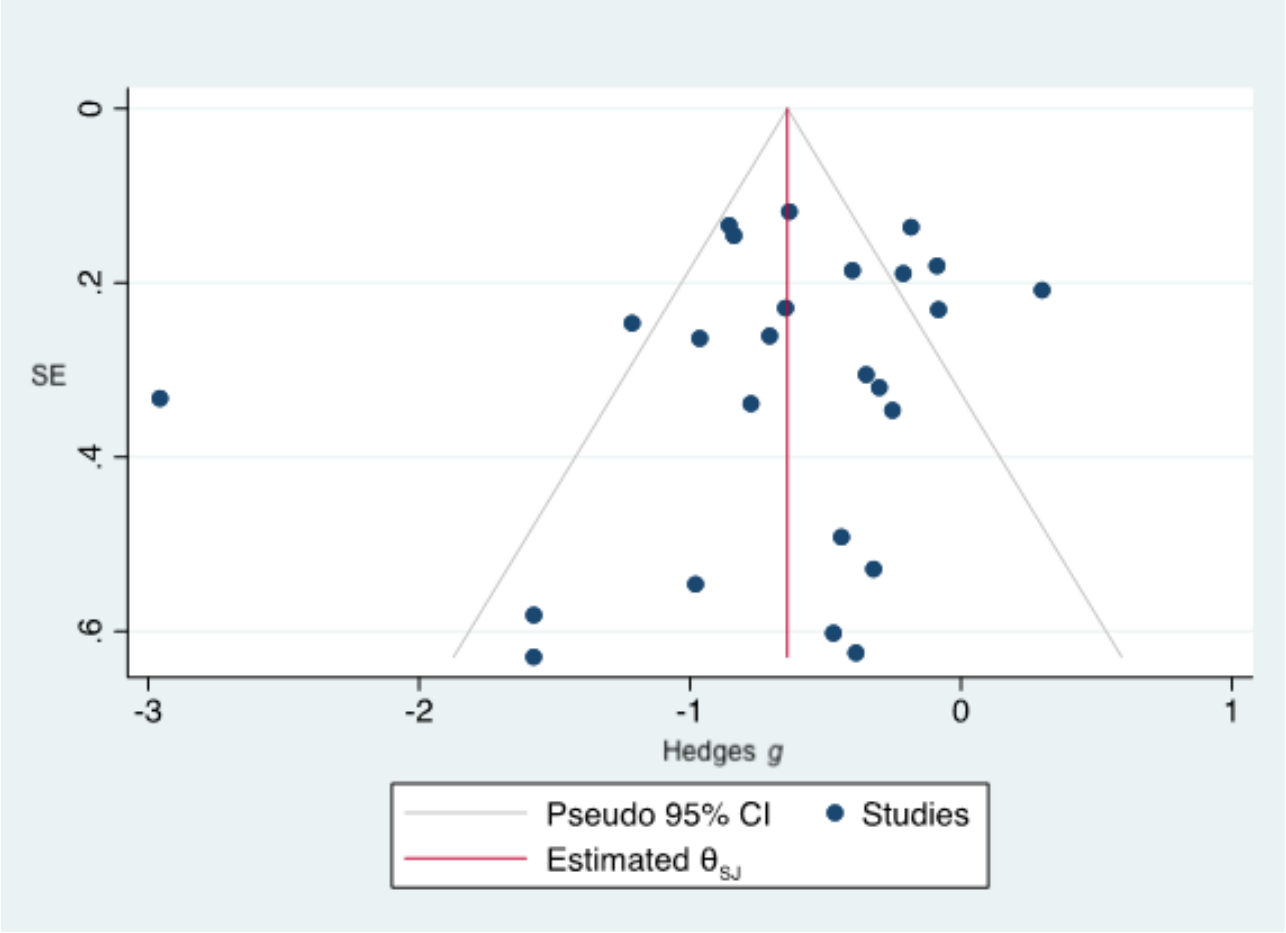
[d]EPDS: Edinburgh Postnatal Depression Scale.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]BA: behavioral activation.

[g]CBT: cognitive behavioral therapy.

Figure 4. Funnel plot.



mHealth Versus Active Control

Three studies compared mHealth interventions with active controls such as bibliotherapy [36], computerized CBT [49], and face-to-face behavioral activation [40]. Liu et al [36] found that a chatbot-delivered self-help depression intervention was superior to bibliotherapy in reducing depression. Watts et al [49] investigated whether a previously validated computerized program would be effective when delivered via a mobile phone app. Both the mobile and computer groups showed significant reductions in depressive symptoms at the 3-month follow-up. The authors concluded that the study provided preliminary evidence of clinically significant improvements in depressive symptoms when CBT is delivered via a mobile app. Ly et al [40] compared a hybrid treatment combining face-to-face behavioral activation and a smartphone app with a 10-session behavioral activation in people with major depression. Although both groups displayed significant improvements after 6 months of treatment, the hybrid intervention had reduced therapist time.

Safety of mHealth Interventions

A total of 14 studies reported in 13 references [31,35,36,41-44,47,50,52,53,57,58] provided information on the safety of mHealth interventions. Most of these studies

assessed safety by monitoring adverse events. Only Bruhns et al [52] and Mantani et al [41] included specific questionnaires: the Inventory for Assessing Negative Effects of Psychotherapy and Frequency, Intensity, and Burden of Side Effects Ratings, respectively. Overall, 9 of the 13 studies (69%) did not report any study-related adverse events. A complete description of the safety results in the included studies can be found in Multimedia Appendix 5 [31,35,36,41-44,47,50,52,53,57,58].

Outcome Tools and Measures

The main end point outcome in the included studies was a reduction in depressive symptoms. However, several studies included secondary outcomes related to the effectiveness of mHealth interventions, such as quality of life, behavioral activation, and anxiety.

Quality of Life

A total of 11 studies out of 29 (37%) attempted to measure participants' quality of life [31,33,37,40,44,50-53,58]. The World Health Organization Quality of Life—Abbreviated Version (4/11, 36%) was the most frequently used outcome tool, followed by the EQ-5D-3L (2/11, 18%), and the 36-item Short Form Survey (2/11, 18%). The Quality of Life Inventory, European Health Interview Survey–Quality of Life 8-Item Index, and 12-Item Short Form Survey were identified among the outcome tools in one study each. Overall, the results were inconsistent across studies, with 5 studies in 4 references reporting significant differences between groups in favor of the mHealth intervention for quality of life [31,50,51,58] and 6 studies reporting no significant differences between groups [33,37,40,44,52,53].

Anxiety

Nine of 29 studies (31%) also included anxiety as an outcome measure [33,35,36,40,44,45,50,53,56]. The most frequently used tool was the Generalized Anxiety Disorder-7 (5/9, 56%), followed by the anxiety subscale of the Hospital Anxiety and Depression Scale (2/9, 22%). The State-Trait Anxiety Inventory (1/9, 11%) and Beck Anxiety Inventory (1/9, 11%) were used in one study each. Six studies found that mHealth interventions significantly reduced anxiety symptoms compared with the waiting list [33,35,40,44,45,50,53,56] or bibliotherapy [36].

Perceived Stress

Perceived stress was assessed in 7 studies (24%) [35,44,46,48,49,51,58]. Three of these used the Perceived Stress Scale, another 3 used 6- or 10-item versions of the Kessler Screening Scale for Psychological Distress, and 1 assessed parenting stress through the Parenting Stress Index. Results were inconsistent, with 4 observing significant stress reductions with the mHealth interventions compared with the control group [44,48,49,58], and 3 indicating no significant effects [35,46,51].

Disability

According to the World Health Organization, depression is a leading cause of disability worldwide and a major contributor to the overall global burden of disease. Disability was measured in 6 (20%) out of 29 studies [31,32,42,49,50]: the Sheehan Disability Scale was used in 3 studies (N=6, 50%), the World Health Organization Disability Assessment Schedule II was used in 2 (N=6, 33%), and 1 used the Disability Symptom Severity (N=6, 16%). Three studies in 2 references found significant effects [31,51], whereas 3 others did not [32,42,49]. Therefore, mHealth interventions have not been conclusively proven to reduce depression-related disability.

Behavioral Activation

As a person with depression may withdraw from their surroundings and disengage from their routines, thus reducing opportunities for positive reinforcement, many depression interventions have included behavioral activation as a goal. Four of the 29 studies (13%) [31,51,56] assessed behavioral activation using the Short Form of the Behavioral Activation for Depression Scale, and 3 of these found statistically significant differences between mHealth interventions and control groups.

Insomnia

In 4 of 29 studies (13%), insomnia was measured using the Insomnia Severity Index [44,50,53,56]. Significant between-group differences favoring the mHealth intervention were found in 3 of these studies (low to large effect sizes compared with the waiting list) [50,53,56]. In contrast, Raevuori et al [44] found no significant differences in sleep disturbance between a group receiving mHealth plus usual care and a control group receiving usual care alone.

Self-Efficacy

Three studies (N=29, 10%) assessed the effectiveness of mHealth interventions on self-efficacy [45,46,58]. Measures used included the General Self-Efficacy Scale, Self-Efficacy Scale, and Parental Sense of Competence Scale. Both studies using general self-efficacy measures found significant results favoring mHealth interventions [45,58], but no effect on parental competence was found [46].

Self-Esteem

The Rosenberg Self-esteem Scale was used in 2 studies (N=29, 6.9%) that compared mHealth interventions with waiting list controls. Although Bruhns et al [52] found a medium to large effect size favoring smartphone self-help apps, Lüdtke et al [37] found no statistically significant differences between groups.

Other Outcome Tools and Measures

Each of the following outcome measures was assessed and described in a single study (1/29, 3%): knowledge of depression (self-developed scale) [43]; problem-solving (Social Problem-Solving Inventory–Revised) [56]; mastery (Pearlin Mastery Scale) [56]; negative thinking (Automatic Thoughts Questionnaire–Revised) [43]; coping (Simplified Ways of Coping Questionnaire) [58]; physical activity (Global Physical Activity Questionnaire) [58]; dysfunctional attitudes (Dysfunctional Attitudes Scale) [33]; affect (The Positive and Negative Affect Schedule) [36]; well-being (World Health Organization Well-being Index) [39]; psychological inflexibility and experiential avoidance (Acceptance and Action Questionnaire) [40]; resilience (Resilience Scale) [44]; satisfaction with life (Satisfaction with Life Scale) [45]; impulsivity (The Barratt Impulsivity Scale) [48]; suicidality (Depressive Symptom Inventory–Suicidality Subscale) [48];

psychological functioning (Functional Assessment Short Test) [51]; empowerment (Roger's Empowerment Scale) [51]; and worry (Penn State Worry Questionnaire) [51]. The results for these outcomes can be found in Table S1 in Multimedia Appendix 6 [31-33,35-40,42-46,48-53,56,58].

Output Tools and Measures

Although the main aim of the selected studies was to measure the effectiveness of mHealth interventions in reducing depressive symptoms, most also measured other outputs that could be relevant in determining primary outcome measures, such as adherence and app use, acceptability, and usability. The results for these outputs can be found in Table S2 in Multimedia Appendix 6.

Discussion

Principal Findings

Our review assessed 29 studies reported in 28 articles involving a substantial number of adult patients with elevated depressive symptoms or diagnosed depression. The meta-analysis of 26 studies comparing the effectiveness of mHealth interventions with the waiting list, minimal intervention, and TAU found moderate positive effects (Hedges g=−0.62) for mHealth, despite high levels of heterogeneity. These results align with those of 2 earlier meta-analyses comparing the efficacy of mHealth interventions and nonactive controls on reducing depressive symptoms, which showed effects of Hedges g=−0.56 and Hedges g=−0.51 [9,59]. However, these are higher than findings from other studies that included patients with any mental health issue (Hedges g=−0.33) [60] and compared mHealth with active treatments (Hedges g=−0.22) [9]. Owing to high heterogeneity and the small number of studies, conducting a meta-analysis to compare mHealth with other active interventions was not feasible.

The dynamic between health care professionals and patients is undergoing transformation owing to the influence of numerous technological, social, and environmental factors, leading to an evolving and changing relationship [61]. As mental health care delivery evolves toward a hybrid model incorporating both in-person and online interventions for diagnosis, therapy, and monitoring, the use of mobile devices becomes increasingly crucial, serving as an integral component in the assessment and intervention of mental health problems [62,63]. Although the number of studies assessing this type of intervention is small, the available evidence suggests that a combination of these 2 modalities can lead to better outcomes for individuals with depression. A potential explanation for the superior efficacy of hybrid therapy is the synergistic combination of app-based and face-to-face interventions. Although app-based interventions provide access to therapeutic content and activities at any time, face-to-face therapy has the advantages of personal interaction, direct guidance, and a supportive environment. An integration of these modalities provides a comprehensive treatment experience for individuals with depression, which may improve the therapeutic process and lead to better outcomes. Furthermore, the complementary nature of the 2 interventions may enhance the reinforcement of skills and strategies learned in face-to-face therapy as well as provide ongoing support and accountability, thereby potentially improving long-term symptom management. As highlighted by Ly et al [40], this could be explained by the dose-response effect, wherein lower doses of psychotherapy have been associated with poorer outcomes [64]. Moreover, hybrid therapy has the potential to be more cost-effective than traditional face-to-face treatments by combining in-person and on the web or app-based sessions, reducing medical costs per patient and increasing the capacity of therapists to treat more individuals with depression, thereby expanding access to treatment. Despite the crucial importance of implementation costs and cost-effectiveness in determining the feasibility and scalability of digital and hybrid interventions, there is a lack of sufficient evidence to date, and additional research is required to inform public and private reimbursement systems and enable investment in digital interventions.

In terms of app design, our findings suggest that incorporating CBT and acceptance frameworks can lead to a greater reduction in depressive symptoms. However, subgroup analyses by theoretical framework did not show statistically significant differences. This is consistent with existing evidence on the effectiveness of psychological interventions. Although CBT is one of the main nonpharmacological treatment options for depressive disorders, a recent network meta-analysis covering efficacy, acceptability, and long-term outcomes found little difference in results from various types of psychotherapy and concluded that most are effective and acceptable for treating adult depression [65]. Clearly, it is essential to design mHealth interventions based on evidence-based frameworks to guarantee their foundation in robust and reliable scientific evidence, and studies have highlighted the need for future research to better characterize the app features that maximize therapeutic effects [66]. However, we found that none of the individual elements in the apps (ie, psychoeducation, mood monitoring, in-app feedback, goal setting, gamification, and professional support) was significantly associated with a greater reduction in depressive symptoms. Moreover, mHealth interventions with a larger number of components are not always more successful: in some cases, simpler interventions that focus on a limited number of well-implemented and user-centered elements can be more effective. It is thus necessary to move beyond "one-size-fits-all" approaches in the design and delivery of mHealth interventions and prioritize tailored approaches that consider individual differences, needs, and preferences [67,68].

With the goal of identifying which sociodemographic and clinical characteristics of patients were associated with greater app effectiveness, we performed subgroup analyses and meta-regressions for gender, age, and baseline depression symptom severity variables. Our results show that mHealth interventions have been effective across demographic factors but may be more effective for individuals with moderate to severe depressive symptoms than for those with lower symptom levels. This is consistent with the results of a previous systematic review [59]. Furthermore, it is in line with the findings of other studies that have concluded that individuals with severe burden benefit equally or to a greater extent from low-intensity internet- and mobile-based interventions [69-71]. There are several potential explanations for these findings. One possibility is that patients with more symptoms have a greater capacity for

definable and noticeable improvement. In addition, people with moderate to severe depressive symptoms may be more motivated to engage in psychological interventions and more likely to adhere to a treatment plan.

The disparity between RCT data and individual patient characteristics encountered in real-world health care settings is a widely acknowledged challenge in daily clinical practice [72]. To ensure the ultimate success of the mental health technology revolution, it is imperative to bolster the path toward the evaluation of implementation, bridging the gap between research findings and the unique features of each patient [73]. Although RCTs have demonstrated the effectiveness of digital interventions for addressing common mental health issues, it is crucial to shift our focus beyond these controlled settings. Unfortunately, there is a scarcity of reported data regarding the implementation of these interventions in the real-world context. The limited available data suggest that uptake and engagement vary widely among the handful of implemented digital self-help apps and programs that have reported this and that use may vary from that reported in trials [74]. It is essential to assess how these mHealth tools are used in real-life scenarios and to determine the extent to which their effectiveness endures beyond the controlled environment of research studies. This exploration beyond RCTs will provide valuable insights into their practical impact, accessibility, and overall contribution to enhancing the mental health of the population.

The increase in the use of mHealth apps has outstripped the development of international standards or practical evaluation tools to assess their effectiveness in a comprehensive and efficient manner. Despite a plethora of mHealth interventions, few have undergone rigorous scientific evaluation. In addition, most mHealth apps that have encountered any evaluation have only undergone a single study, typically with a small sample size. Only a minority of the mHealth interventions identified in our review have been subjected to evaluation in more than one study. Our results do indicate consistency in the assessment of depressive symptoms, as most studies use established and validated measurement tools, such as the Beck Depression Inventory–II, Patient Health Questionnaire-9, and Hospital Anxiety and Depression Scale. However, given the high heterogeneity of the identified measures, there appears to be a lack of consensus on how to assess other important outputs that are crucial in determining primary outcome measures, such as adherence, acceptability, usability, and app use. Furthermore, the absence of adequate regulatory bodies to oversee and regulate app development and availability has made accessing trustworthy and validated mHealth interventions a challenging process [21]. Accordingly, there remains a requirement for the development of new tools and methodologies that facilitate the assessment of various aspects of mHealth interventions intended to manage specific conditions. The results of this SR enable us to understand the effectiveness and safety of apps targeting depression that have been evaluated in RCTs, as well as the evaluation criteria used, and will serve as a starting point for the design of an evaluation tool within the context of the EvalDepApps research project [24].

Strengths and Limitations of This Study

Our study has several key strengths, including a rigorous and systematic search and selection process that ensured comprehensive coverage of the available evidence. Furthermore, the use of validated quality assessment tools facilitated a robust evaluation of the risk of bias in the included studies. Clear and transparent reporting of methods and results enhances the reproducibility of the findings and strengthens the validity and reliability of the results. However, there are also several limitations that should be considered when interpreting our results. Our search for studies was limited to those published in English or Spanish and did not incorporate gray literature, which may have excluded some relevant studies. It should also be noted that most of the reviewed studies were conducted in Western high-income countries; thus, it is unclear whether these results can be generalized to low- and middle-income countries. Our analyses revealed moderate heterogeneity that could not be fully accounted for through subgroup analyses. This heterogeneity may be due to differences in populations, interventions (including the framework, elements included, and definitions of these elements), and comparators across the trials. For example, we compared mHealth interventions with a variety of control conditions, including waiting list, minimal intervention, and TAU. Although we found no significant differences between these control conditions, the variability among them may have contributed to the overall heterogeneity. Another noteworthy limitation of our review was the exclusion of studies that did not present results from RCTs. Although observational studies and nonrandomized trials could potentially offer valuable insights into the practical use and effectiveness of mHealth in the real-world context, we decided to exclude them because of the higher susceptibility of these trial designs to various biases, which may compromise the reliability of the findings. Finally, there are important limitations associated with the small sample sizes and moderate to high risk of bias present in most of the studies reviewed.

Conclusions

This study suggests that mHealth interventions directed toward adults experiencing elevated symptoms of depression result in moderate decreases in these symptoms, regardless of age and gender, with hybrid interventions achieving the best results. However, it should be noted that most of the studies in this review had small sample sizes and were associated with a moderate to high risk of bias. In addition, a high level of heterogeneity was observed in the characteristics and components of the mHealth interventions, with no singular element found to be associated with improved outcomes. Hence, it is imperative to move beyond generic solutions when designing and delivering mHealth interventions and prioritize individualized approaches that take into consideration individual differences, needs, and preferences.

Acknowledgments

The EvalDepApp Project received financial support from the Instituto de Salud Carlos III, grant number PI21/00234, funded by FEDER. AP-N wants to thank CYTED network Geolibero, the network RED2022-134355-T, financed by MCIN/AEI/10.13039/501100011033/FEDER, UE; and the project PID2021-122642OB-C44.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - mental_v10i1e46877_app1.docx]

Multimedia Appendix 2
Characteristics and elements included in the mobile health interventions.
[DOCX File , 16 KB - mental_v10i1e46877_app2.docx]

Multimedia Appendix 3
Sensitivity analysis.
[DOCX File , 182 KB - mental_v10i1e46877_app3.docx]

Multimedia Appendix 4
Galbraith plot.
[DOCX File , 68 KB - mental_v10i1e46877_app4.docx]

Multimedia Appendix 5
Safety of the mobile health interventions.
[DOCX File , 16 KB - mental_v10i1e46877_app5.docx]

Multimedia Appendix 6
Secondary outcome results from the selected studies: output tools and results from the selected studies.
[DOCX File , 35 KB - mental_v10i1e46877_app6.docx]

References

1. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of depression in the community from 30 countries between 1994 and 2014. Sci Rep 2018 Feb 12;8(1):2861 [FREE Full text] [doi: 10.1038/s41598-018-21243-x] [Medline: 29434331]
2. Liu Q, He H, Yang J, Feng X, Zhao F, Lyu J. Changes in the global burden of depression from 1990 to 2017: findings from the Global Burden of Disease study. J Psychiatr Res 2020 Jul;126:134-140 [FREE Full text] [doi: 10.1016/j.jpsychires.2019.08.002] [Medline: 31439359]
3. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]
4. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF3. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry 2013 Jun;12(2):137-148 [FREE Full text] [doi: 10.1002/wps.20038] [Medline: 23737423]
5. Jacob KS, Sharan P, Mirza I, Garrido-Cumbrera M, Seedat S, Mari JJ, et al. Mental health systems in countries: where are we now? Lancet 2007 Sep 22;370(9592):1061-1077. [doi: 10.1016/S0140-6736(07)61241-0] [Medline: 17804052]
6. Eaton J, McCay L, Semrau M, Chatterjee S, Baingana F, Araya R, et al. Scale up of services for mental health in low-income and middle-income countries. Lancet 2011 Oct 29;378(9802):1592-1603. [doi: 10.1016/S0140-6736(11)60891-X] [Medline: 22008429]
7. Kakuma R, Minas H, van Ginneken N, Dal Poz MR, Desiraju K, Morris JE, et al. Human resources for mental health care: current situation and strategies for action. Lancet 2011 Nov 05;378(9803):1654-1663. [doi: 10.1016/S0140-6736(11)61093-3] [Medline: 22008420]
8. Liu JX, Goryakin Y, Maeda A, Bruckner T, Scheffler R. Global health workforce labor market projections for 2030. Hum Resour Health 2017 Feb 03;15(1):11 [FREE Full text] [doi: 10.1186/s12960-017-0187-2] [Medline: 28159017]

9. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
10. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
11. mHealth programmes are sponsored by government. World Health Organization. URL: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/4774 [accessed 2022-12-23]
12. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
13. Economides M, Ranta K, Nazander A, Hilgert O, Goldin PR, Raevuori A, et al. Long-term outcomes of a therapist-supported, smartphone-based intervention for elevated symptoms of depression and anxiety: quasiexperimental, pre-postintervention study. JMIR Mhealth Uhealth 2019 Aug 26;7(8):e14284 [FREE Full text] [doi: 10.2196/14284] [Medline: 31452521]
14. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2012 Nov 14;11:CD006611. [doi: 10.1002/14651858.CD006611.pub3] [Medline: 23152238]
15. Ha S, Kim J. Designing a scalable, accessible, and effective mobile app based solution for common mental health problems. Int J Hum Comput Interact 2020 Apr 26;36(14):1354-1367 [FREE Full text] [doi: 10.1080/10447318.2020.1750792]
16. Huguet A, Rao S, McGrath P, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016 May 02;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
17. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
18. Rauschenberg C, Schick A, Hirjak D, Seidler A, Paetzold I, Apfelbacher C, et al. Evidence synthesis of digital interventions to mitigate the negative impact of the COVID-19 pandemic on public mental health: rapid meta-review. J Med Internet Res 2021 Mar 10;23(3):e23365 [FREE Full text] [doi: 10.2196/23365] [Medline: 33606657]
19. Torous J, Firth J, Huckvale K, Larsen ME, Cosco TD, Carney R, et al. The emerging imperative for a consensus approach toward the rating and clinical recommendation of mental health apps. J Nerv Ment Dis 2018 Aug;206(8):662-666. [doi: 10.1097/NMD.0000000000000864] [Medline: 30020203]
20. Neary M, Schueller S. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537 [FREE Full text] [doi: 10.1016/j.cbpra.2018.01.002] [Medline: 33100810]
21. Eis S, Solà-Morales O, Duarte-Díaz A, Vidal-Alaball J, Perestelo-Pérez L, Robles N, et al. Mobile applications in mood disorders and mental health: systematic search in Apple App store and Google Play store and review of the literature. Int J Environ Res Public Health 2022 Feb 15;19(4):2186 [FREE Full text] [doi: 10.3390/ijerph19042186] [Medline: 35206373]
22. Kertz S, MacLaren Kelly J, Stevens K, Schrock M, Danitz S. A review of free iPhone applications designed to target anxiety and worry. J Technol Behav Sci 2017 Jan 11;2(2):61-70 [FREE Full text] [doi: 10.1007/s41347-016-0006-y]
23. López Seguí F, Pratdepàdua Bufill C, Rius Soler A, de San Pedro M, López Truño B, Aguiló Laine A, et al. Prescription and integration of accredited mobile apps in Catalan health and social care: protocol for the Appsalut site design. JMIR Res Protoc 2018 Dec 21;7(12):e11414 [FREE Full text] [doi: 10.2196/11414] [Medline: 30578234]
24. Carrion C, Sales-Masnou A, Eis S, Robles N, Puigdomènech E, Duarte-Díaz A, et al. PP40 health apps to manage depression: can we separate the grain from the chaff? EvalDepApps project. Int J Technol Assess Health Care 2022 Dec 23;38(S1):S53-S54 [FREE Full text] [doi: 10.1017/s0266462322001817]
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
26. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
27. Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. 2nd edition. Hoboken, NJ: John Wiley & Sons; 2019:241-284.
28. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol 2014 Feb 18;14:25 [FREE Full text] [doi: 10.1186/1471-2288-14-25] [Medline: 24548571]
29. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558 [FREE Full text] [doi: 10.1002/sim.1186] [Medline: 12111919]
30. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]

31. Araya R, Menezes PR, Claro HG, Brandt LR, Daley KL, Quayle J, et al. Effect of a digital intervention on depressive symptoms in patients with comorbid hypertension or diabetes in Brazil and Peru: two randomized clinical trials. JAMA 2021 May 11;325(18):1852-1862 [FREE Full text] [doi: 10.1001/jama.2021.4348] [Medline: 33974019]
32. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
33. Ham K, Chin S, Suh YJ, Rhee M, Yu ES, Lee HJ, et al. Preliminary results from a randomized controlled study for an app-based cognitive behavioral therapy program for depression and anxiety in cancer patients. Front Psychol 2019 Jul 25;10:1592 [FREE Full text] [doi: 10.3389/fpsyg.2019.01592] [Medline: 31402881]
34. Jannati N, Mazhari S, Ahmadian L, Mirzaee M. Effectiveness of an app-based cognitive behavioral therapy program for postpartum depression in primary care: a randomized controlled trial. Int J Med Inform 2020 Sep;141:104145 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104145] [Medline: 32480319]
35. Kageyama K, Kato Y, Mesaki T, Uchida H, Takahashi K, Marume R, et al. Effects of video viewing smartphone application intervention involving positive word stimulation in people with subthreshold depression: a pilot randomized controlled trial. J Affect Disord 2021 Mar 01;282:74-81 [FREE Full text] [doi: 10.1016/j.jad.2020.12.104] [Medline: 33401126]
36. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. Internet Interv 2022 Jan 06;27:100495 [FREE Full text] [doi: 10.1016/j.invent.2022.100495] [Medline: 35059305]
37. Lüdtke T, Pult LK, Schröder J, Moritz S, Bücker L. A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. Psychiatry Res 2018 Nov;269:753-762 [FREE Full text] [doi: 10.1016/j.psychres.2018.08.113] [Medline: 30273901]
38. Lukas C, Eskofier B, Berking M. A gamified smartphone-based intervention for depression: randomized controlled pilot trial. JMIR Ment Health 2021 Jul 20;8(7):e16643 [FREE Full text] [doi: 10.2196/16643] [Medline: 34283037]
39. Lukas CA, Berking M. Blending group-based psychoeducation with a smartphone intervention for the reduction of depressive symptoms: results of a randomized controlled pilot study. Pilot Feasibility Stud 2021 Feb 24;7(1):57 [FREE Full text] [doi: 10.1186/s40814-021-00799-y] [Medline: 33627198]
40. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015 May 26;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
41. Mantani A, Kato T, Furukawa T, Horikoshi M, Imai H, Hiroe T, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. J Med Internet Res 2017 Nov 03;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
42. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
43. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
44. Raevuori A, Vahlberg T, Korhonen T, Hilgert O, Aittakumpu-Hyden R, Forman-Hoffman V. A therapist-guided smartphone app for major depression in young adults: a randomized clinical trial. J Affect Disord 2021 May 01;286:228-238 [FREE Full text] [doi: 10.1016/j.jad.2021.02.007] [Medline: 33743385]
45. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of SuperBetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246 [FREE Full text] [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
46. Sawyer A, Kaim A, Le H, McDonald D, Mittinty M, Lynch J, et al. The effectiveness of an app-based nurse-moderated program for new mothers with depression and parenting problems (eMums Plus): pragmatic randomized controlled trial. J Med Internet Res 2019 Jun 04;21(6):e13689 [FREE Full text] [doi: 10.2196/13689] [Medline: 31165715]
47. Stiles-Shields C, Montague E, Kwasny MJ, Mohr DC. Behavioral and cognitive intervention strategies delivered via coached apps for depression: pilot trial. Psychol Serv 2019 May;16(2):233-238 [FREE Full text] [doi: 10.1037/ser0000261] [Medline: 30407055]
48. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
49. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013 Feb 07;13(1):a [FREE Full text] [doi: 10.1186/1471-244x-13-49]
50. Wong VW, Ho FY, Shi NK, Tong JT, Chung KF, Yeung WF, et al. Smartphone-delivered multicomponent lifestyle medicine intervention for depressive symptoms: a randomized controlled trial. J Consult Clin Psychol 2021 Dec;89(12):970-984 [FREE Full text] [doi: 10.1037/ccp0000695] [Medline: 35025538]

51. Tønning ML, Faurholt-Jepsen M, Frost M, Martiny K, Tuxen N, Rosenberg N, et al. The effect of smartphone-based monitoring and treatment on the rate and duration of psychiatric readmission in patients with unipolar depressive disorder: the RADMIS randomized controlled trial. J Affect Disord 2021 Mar 01;282:354-363 [FREE Full text] [doi: 10.1016/j.jad.2020.12.141] [Medline: 33421863]
52. Bruhns A, Lüdtke T, Moritz S, Bücker L. A mobile-based intervention to increase self-esteem in students with depressive symptoms: randomized controlled trial. JMIR Mhealth Uhealth 2021 Jul 12;9(7):e26498 [FREE Full text] [doi: 10.2196/26498] [Medline: 34255711]
53. Chan CS, Wong CY, Yu BY, Hui VK, Ho FY, Cuijpers P. Treating depression with a smartphone-delivered self-help cognitive behavioral therapy for insomnia: a parallel-group randomized controlled trial. Psychol Med 2023 Apr;53(5):1799-1813. [doi: 10.1017/S0033291721003421] [Medline: 37310329]
54. Dahne J, Lejuez CW, Diaz VA, Player MS, Kustanowitz J, Felton JW, et al. Pilot randomized trial of a self-help behavioral activation mobile app for utilization in primary care. Behav Ther 2019 Jul;50(4):817-827 [FREE Full text] [doi: 10.1016/j.beth.2018.12.003] [Medline: 31208690]
55. Dahne J, Collado A, Lejuez CW, Risco CM, Diaz VA, Coles L, et al. Pilot randomized controlled trial of a Spanish-language behavioral activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among united states Latinx adults with limited English proficiency. J Affect Disord 2019 May 01;250:210-217 [FREE Full text] [doi: 10.1016/j.jad.2019.03.009] [Medline: 30870770]
56. Ebert DD, Buntrock C, Lehr D, Smit F, Riper H, Baumeister H, et al. Effectiveness of web- and mobile-based treatment of subthreshold depression with adherence-focused guidance: a single-blind randomized controlled trial. Behav Ther 2018 Jan;49(1):71-83. [doi: 10.1016/j.beth.2017.05.004] [Medline: 29405923]
57. Graham AK, Greene CJ, Kwasny MJ, Kaiser SM, Lieponis P, Powell T, et al. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. JAMA Psychiatry 2020 Sep 01;77(9):906-914 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1011] [Medline: 32432695]
58. Guo Y, Hong YA, Cai W, Li L, Hao Y, Qiao J, et al. Effect of a WeChat-based intervention (Run4Love) on depressive symptoms among people living with HIV in China: a randomized controlled trial. J Med Internet Res 2020 Feb 09;22(2):e16715 [FREE Full text] [doi: 10.2196/16715] [Medline: 32044751]
59. Serrano-Ripoll MJ, Zamanillo-Campos R, Fiol-DeRoque MA, Castro A, Ricci-Cabello I. Impact of smartphone app-based psychological interventions for reducing depressive symptoms in people with depression: systematic literature review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 Jan 27;10(1):e29621 [FREE Full text] [doi: 10.2196/29621] [Medline: 35084346]
60. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 02;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
61. Yellowlees P, Richard Chan S, Burke Parish M. The hybrid doctor-patient relationship in the age of technology - telepsychiatry consultations and the use of virtual space. Int Rev Psychiatry 2015;27(6):476-489 [FREE Full text] [doi: 10.3109/09540261.2015.1082987] [Medline: 26493089]
62. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
63. Chan SR, Torous J, Hinton L, Yellowlees P. Mobile tele-mental health: increasing applications and a move to hybrid models of care. Healthcare (Basel) 2014 May 06;2(2):220-233 [FREE Full text] [doi: 10.3390/healthcare2020220] [Medline: 27429272]
64. Harnett P, O'Donovan A, Lambert M. The dose response relationship in psychotherapy: implications for social policy. Clin Psychol 2010;14(2):39-44 [FREE Full text] [doi: 10.1080/13284207.2010.500309]
65. Cuijpers P, Quero S, Noma H, Ciharova M, Miguel C, Karyotaki E, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry 2021 Jun;20(2):283-293 [FREE Full text] [doi: 10.1002/wps.20860] [Medline: 34002502]
66. Six SG, Byrne KA, Tibbett TP, Pericot-Valverde I. Examining the effectiveness of gamification in mental health apps for depression: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 29;8(11):e32199 [FREE Full text] [doi: 10.2196/32199] [Medline: 34847058]
67. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020 Feb 20;6:2055207620905411 [FREE Full text] [doi: 10.1177/2055207620905411] [Medline: 32128233]
68. Dugas M, Crowley K, Gao G, Xu T, Agarwal R, Kruglanski AW, et al. Individual differences in regulatory mode moderate the effectiveness of a pilot mHealth trial for diabetes management among older veterans. PLoS One 2018 Mar 07;13(3):e0192807 [FREE Full text] [doi: 10.1371/journal.pone.0192807] [Medline: 29513683]
69. Weisel KK, Lehr D, Heber E, Zarski AC, Berking M, Riper H, et al. Severely burdened individuals do not need to be excluded from internet-based and mobile-based stress management: effect modifiers of treatment outcomes from three

randomized controlled trials. J Med Internet Res 2018 Jun 19;20(6):e211 [FREE Full text] [doi: 10.2196/jmir.9387] [Medline: 29921562]
70. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 2013 Feb 26;346(feb26 2):f540 [FREE Full text] [doi: 10.1136/bmj.f540] [Medline: 23444423]
71. Karyotaki E, Ebert D, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev 2018 Jul;63:80-92 [FREE Full text] [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
72. Kibbelaar RE, Oortgiesen BE, van der Wal-Oost AM, Boslooper K, Coebergh JW, Veeger NJ, et al. Bridging the gap between the randomised clinical trial world and the real world by combination of population-based registry and electronic health record data: a case study in haemato-oncology. Eur J Cancer 2017 Nov;86:178-185 [FREE Full text] [doi: 10.1016/j.ejca.2017.09.007] [Medline: 28992561]
73. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 01;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
74. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]

Abbreviations

CBT: cognitive behavioral therapy
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
TAU: treatment as usual

Edited by J Torous; submitted 02.03.23; peer-reviewed by S Rennick-Egglestone, P Yellowlees; comments to author 31.05.23; revised version received 14.07.23; accepted 15.07.23; published 27.09.23.

Please cite as:
Duarte-Díaz A, Perestelo-Pérez L, Gelabert E, Robles N, Pérez-Navarro A, Vidal-Alaball J, Solà-Morales O, Sales Masnou A, Carrion C
Efficacy, Safety, and Evaluation Criteria of mHealth Interventions for Depression: Systematic Review
JMIR Ment Health 2023;10:e46877
URL: https://mental.jmir.org/2023/1/e46877
doi:10.2196/46877
PMID:37756042

Review

The Safety of Digital Mental Health Interventions: Systematic Review and Recommendations

Rayan Taher[1], MSc; Che-Wei Hsu[1], PhD; Chloe Hampshire[2], MSc; Carolina Fialho[1], MSc; Clare Heaysman[3], PhD; Daniel Stahl[1], PhD; Sukhi Shergill[4], MD, PhD; Jenny Yiend[1], PhD

[1]Psychosis Studies Department, Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London, London, United Kingdom

[2]Department of Psychology, University of Bath, Bath, United Kingdom

[3]London Institute for Healthcare Engineering, King's College London, London, United Kingdom

[4]Kent and Medway Medical School, Canterbury, Kent, United Kingdom

Corresponding Author:
Jenny Yiend, PhD
Psychosis Studies Department, Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London
16 De Crespigny Park
London, SE5 8AB
United Kingdom
Phone: 44 020 7848 0002
Email: jenny.yiend@kcl.ac.uk

Abstract

Background: Evidence suggests that digital mental health interventions (DMHIs) for common mental health conditions are effective. However, digital interventions, such as face-to-face therapies, pose risks to patients. A *safe* intervention is considered one in which the measured benefits outweigh the identified and mitigated risks.

Objective: This study aims to review the literature to assess how DMHIs assess safety, what risks are reported, and how they are mitigated in both the research and postmarket phases and building on existing recommendations for assessing, reporting, and mitigating safety in the DMHI and standardizing practice.

Methods: PsycINFO, Embase, and MEDLINE databases were searched for studies that addressed the safety of DMHIs. The inclusion criteria were any study that addressed the safety of a clinical DMHI, even if not as a main outcome, in an adult population, and in English. As the outcome data were mainly qualitative in nature, a meta-analysis was not possible, and qualitative analysis was used to collate the results. Quantitative results were synthesized in the form of tables and percentages. To illustrate the use of a single common safety metric across studies, we calculated odds ratios and CIs, wherever possible.

Results: Overall, 23 studies were included in this review. Although many of the included studies assessed safety by actively collecting adverse event (AE) data, over one-third (8/23, 35%) did not assess or collect any safety data. The methods and frequency of safety data collection varied widely, and very few studies have performed formal statistical analyses. The main treatment-related reported AE was symptom deterioration. The main method used to mitigate risk was exclusion of high-risk groups. A secondary web-based search found that 6 DMHIs were available for users or patients to use (postmarket phase), all of which used indications and contraindications to mitigate risk, although there was no evidence of ongoing safety review.

Conclusions: The findings of this review show the need for a standardized classification of AEs, a standardized method for assessing AEs to statically analyze AE data, and evidence-based practices for mitigating risk in DMHIs, both in the research and postmarket phases. This review produced 7 specific, measurable, and achievable recommendations with the potential to have an immediate impact on the field, which were implemented across ongoing and future research. Improving the quality of DMHI safety data will allow meaningful assessment of the safety of DMHIs and confidence in whether the benefits of a new DMHI outweigh its risks.

Trial Registration: PROSPERO CRD42022333181; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=333181

(JMIR Ment Health 2023;10:e47433) doi:10.2196/47433

KEYWORDS

digital; digital therapeutics; mental health; psychological; safety; risk; negative effects; harm; adverse event; risk mitigation; mobile phone

Introduction

Digital mental health interventions (DMHIs) are therapeutic interventions delivered via digital technologies, such as mobile apps, websites, or virtual reality (VR), that aim to improve patients' mental health [1]. The effectiveness of some of these interventions has been well established and is comparable with that of face-to-face therapies for the treatment of common mental health conditions such as depression and anxiety [1,2]. A meta-analysis of 66 randomized controlled trials (RCTs) found that DMHIs are effective in treating social anxiety and general anxiety disorders [3]. There is also evidence of their effectiveness in reducing paranoia [4].

DMHIs have the potential to improve access to evidence-based mental health therapies and reach individuals who find the traditional mode of delivery for mental health care difficult [3]. Such interventions can help patients overcome the barrier of stigma by giving them the opportunity to privately seek evidence-based care without having to talk to a professional [2]. Digital interventions are medical devices, and as such, their safety is precisely defined within a regulatory context. An intervention is considered safe if its expected benefits (based on quantifiable evidence gathered to date) outweigh any residual risks, once those risks have been mitigated as far as possible [5]. This highlights the importance of systematic identification and measurement of risk before any safety claims can be made. The safety of an intervention is usually captured by the measurement of unwanted occurrences or so-called *negative effects*. The best way to define and categorize these in the context of psychological therapy (as opposed to medical or pharmaceutical) trials has been the subject of discussion in the literature. Some classification schemes have been proposed [6,7], including one specifically for internet interventions [8]; however, there is still no universally accepted rubric. The most widely recognized distinction is between adverse events (AEs) and serious AEs (SAEs). A recent review found that SAEs in psychotherapy trials were fairly consistently conceptualized [7], which may, in part, be due to the strict regulatory and governance requirements around such events. This obliges researchers to use a prescribed definition that is derived from the pharmaceutical industry and relatively universally accepted. For example, the United Kingdom's Health Research Authority, the United States Food and Drug Administration, and the International Council for Harmonization all consider an SAE to include any event that results in death, disability, incapacity, hospitalization or prolongation of hospitalization, birth defects, or events that might have led to these outcomes were preventative action not taken [9-11]. In contrast, there is considerable heterogeneity in how (nonserious) AEs are conceived, measured, and recorded. In general, AEs include any negative effects or events that occur when a participant is enrolled in a clinical trial. AEs, therefore, encompass a much broader range of less severe possibilities. A systematic review on harm in psychotherapy found that AEs were mentioned significantly more often in pharmacological studies than in psychotherapy studies [7]. In that systematic review, all study protocols that addressed a DMHI (5/115, 4.3%) explicitly considered harm and aimed to assess AEs and SAEs [7]. The review speculated that harm in DMHIs might be more researched compared with face-to-face psychotherapy because of the absence of direct contact with a professional [7].

Previous research on harm concluded that *side effects* are unavoidable in psychotherapy [7,12]. In face-to-face psychotherapy, patients experience approximately 12 AEs per person [7]. The most widely researched risk of mental health interventions is the *deterioration effect* [2]. The deterioration rates reported in DMHIs (12%) are similar to those reported in patients receiving face-to-face therapies [2]. Another risk associated with mental health interventions is novel symptoms—new symptoms not previously experienced by the patient before treatment [13]. Developing dependence on intervention or therapy is a potential risk, which is why therapists usually dedicate sessions toward the end of therapy to prepare patients for termination [13]. Nonresponse (no improvement) is considered a potential negative effect because participation in an ineffective therapy may have prevented access to better alternatives, prolonged the journey to recovery, or hindered recovery [6]. A systematic review found that almost half of patients who receive face-to-face psychotherapy do not experience a significantly positive change in their symptoms posttherapy [7]. AEs, such as deterioration, novel symptoms, and nonresponse are experienced by 5% to 20% of patients [7].

A recent review of study protocols identified problematic heterogeneity in the proposed definitions and assessments of safety, as articulated in the study design stage [7]. Moreover, a recent narrative scoping review explored how AEs are reported in RCTs of DMHIs that are registered in the International Standard Randomized Controlled Trial Number Registry [14]. In this study, we conducted a systematic review of the published safety outcome data to establish the current state of practice, highlight gaps in the literature, and guide future research. Thus, this review aims to answer the following questions.

1. How are the risks and safety of DMHIs currently being assessed? (assessing risk)
2. What are the main reported risks and negative effects of DMHIs? (reporting risk)
3. How do DMHIs mitigate risk in research studies and in the postmarket phase? (mitigating risk)
4. What recommendations can be drawn for future practice based on the current findings?

Methods

This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15].

Eligibility Criteria

Inclusion Criteria

We included all study types that addressed the safety, risks, negative effects, or harm of a DMHI, not necessarily as a main outcome, in an adult sample (aged >16 years). This review includes gray literature.

Exclusion Criteria

Studies on nonclinical digital interventions that target well-being, such as stress management, were excluded from this review. Studies published in languages other than English and those with research protocols were also excluded.

Information Sources

A total of 3 scientific databases were searched in June 2022: PsycINFO, Embase, and MEDLINE. The search did not include a limit on the date or language of the publication. A secondary search was then conducted on the web to retrieve any further safety-related information on the interventions identified in this review.

Search Strategy

The search strategy was run on the 3 selected databases as shown in Textbox 1.

Textbox 1. Search strategy.

1. appab,ti.
2. apps.ab,ti.
3. digital.ab,ti.
4. wearable device.ab,ti.
5. virtual reality.ab,ti.
6. e-mental health.ab,ti.
7. e-health.ab,ti.
8. internet based.ab,ti.
9. mobile health.ab,ti.
10. telehealth.ab,ti.
11. “risk*”.ab,ti.
12. “safe*”.ab,ti.
13. “negative effect*”.ab,ti.
14. “adverse event*”.ab,ti.
15. harm.ab,ti.
16. mental health.ab,ti.
17. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
18. 11 or 12 or 13 or 14 or 15
19. 16 and 17 and 18

Selection Process

One author (RT) conducted the search, removed duplicates, and exported the results to Rayyan, software used to collaborate between reviewers [16]. The retrieved abstracts were independently screened by 4 reviewers (RT, CF, CWH, and C Hampshire) against the inclusion and exclusion criteria. Cohen κ was calculated to assess the interrater reliability between the 4 reviewers and was equal to 0.42, indicating moderate agreement [17]. When the screening was complete, the reviewers met to discuss and resolve any disagreements, which resulted in a unanimous decision in all cases. The full texts of the final agreed list of the included studies were retrieved and screened by RT against the inclusion and exclusion criteria. This was verified by the other reviewers (CF, CWH, and C Hampshire).

Data Collection Process

One author (RT) extracted relevant data from the studies included in this review.

Data Items

The following characteristics of each included study were extracted: title; study design; country and year of publication; sample size; group size (if the study included more than one); study aim; and the relevant Participants, Intervention, Comparison, and Outcome (PICO) details.

The following safety-related data reported in the included studies were extracted: the method of assessment used to measure safety, main findings related to the safety of the DMHI, and measures used to minimize risk. Missing data were reported. Publications were also searched to collect any information about AEs that occurred during the study but were not explicitly reported as AEs by the authors.

Publicly available data on the safety of the interventions identified in this review (intended purpose, contraindications, warnings, or any other safety measures) were extracted from a secondary web-based search.

Study Risk of Bias Assessment

The risk of bias was assessed by 1 reviewer (RT) and verified by another (CWH). Version 2 of the Cochrane risk-of-bias tool for randomized trials (RoB 2) was used to assess the risk of bias in RCTs [18]. RoB 2 estimates the risk of bias that arises because of the randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and the selection of the reported result [18]. The Critical Appraisal Skills Programme (CASP) Cohort Study appraisal tool was used to assess the risk of bias in pre-post studies [19]. The tool assesses the risk of bias arising from a study's validity, bias in recruitment, exposure, outcome measurement, confounding factors, and reporting of results [19].

Effect Measures

The outcomes of this review were mostly qualitative (information about safety and risks); thus, no effect measure was used in the synthesis of the results.

Synthesis Methods

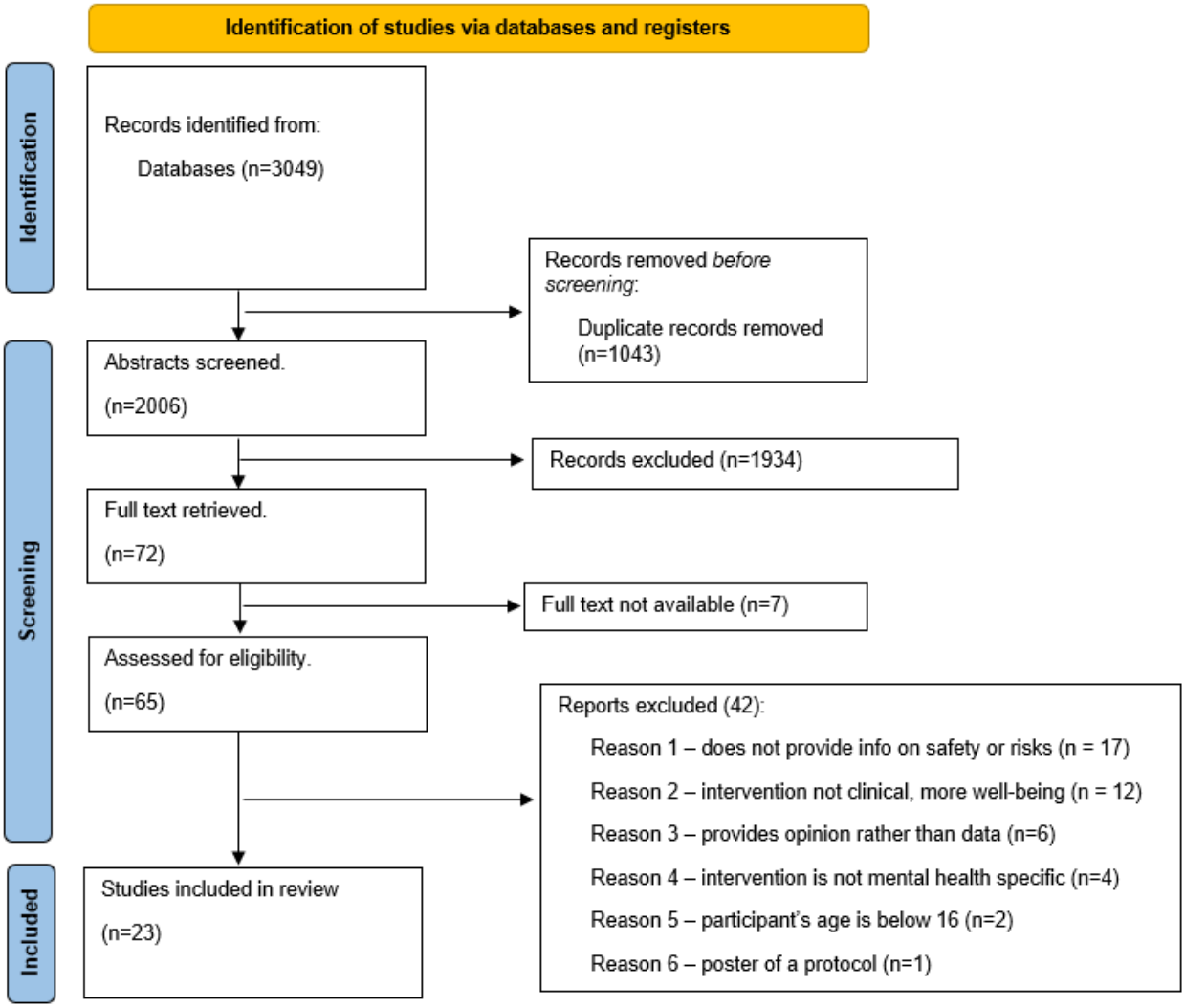
A statistical synthesis method such as meta-analysis was not possible because the outcomes were qualitative (information about safety and risks). The results were synthesized and presented in the form of tables. Where appropriate, descriptive statistics such as percentages were used. In addition, we calculated odds ratios and their CIs, for all studies where sufficient data were available, to ascertain whether the probability of experiencing an AE significantly differed between study arms.

Results

Study Selection

The initial search yielded 3049 results of which 23 were included in this review. In all, 1043 duplicates were removed. At the abstract screening stage of 1934, further papers were excluded for the following reasons: do not address digital mental health, do not address mental health, and do not address safety. Of the remaining 72 papers, 65 full-text articles were available and assessed for eligibility. See Figure 1 for the PRISMA flowchart with further details, including reasons for exclusion at this stage.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



Study Characteristics

Of the 23 included studies, 17 (74%) were RCTs and the rest were pre-post studies. The publication dates ranged between 2008 and 2022, with more than 80% (19/23) of the studies published starting from 2018. The sample sizes ranged from 7 to 3755. One-third (8/23, 35%) of the included studies involved collaborations between 2 or more countries. In total, 37 countries were included in the included studies. The most prevalent country was the United Kingdom (10/23, 43%). The countries involved were mainly high income (22/23, 96%) and Western (20/23, 87%). See Table 1 for more details of the study's characteristics.

Data from the included studies were extracted using PICO criteria:

1. Participants (P): Most (15/23, 65% studies) of the samples in the included studies were clinical: psychosis [4,19-24], depression [25-27], anxiety [28-30], body dysmorphia [31], and eating disorders [32]. Some of the samples represented specific groups, such as first responders who worked during the COVID-19 pandemic [33], veterans with traumatic experiences [34], and physically healthy patients [35,36]. Four of the included studies recruited healthy individuals and screened them for target mental health issues such as suicide [37,38], trauma [39], and insomnia [40].
2. Intervention (I): All interventions in the included studies were digitally delivered. Eleven of them were internet-based programs [25-27,30-35,38,39], 8 were mobile apps [4,19-21,28,29,37,40], 3 were VR-based [22-24], and 1 was social media-based [36].
3. Comparison (C): The 17 RCTs included in this review had different comparison groups. The majority (8/17, 47%) of the comparison groups received treatment as usual [4,20-22,26,34,36,40]. Four comparison groups were waitlist groups [23,29,38,39], 2 were supportive therapy [31,35], 1 was psychoeducation [27], 1 was healthy [28], 1 was a similar neutral intervention [37], and 1 was a similar active intervention [19].
4. Outcome (O): Interventions in the included studies aimed to improve a specific mental health disorder or symptom cluster: depression [25-27,33,35,36], anxiety [22,29,30,33], paranoia [4,23,24], psychosis [19-21,40], suicidal ideation [37,38], posttraumatic stress disorder [33,34,39], eating disorders [32], body dysmorphia [31], and loneliness [28].

Table 1 and Multimedia Appendix 1 [4,20-41] provides more detailed information on the characteristics of the studies and PICO data for each included study.

Table 1. Studies' characteristics.

Study, year	Study design	Total number of participants, N	Study aim	Participants	Intervention	Intervention desired outcome
Arjadi et al [20], 2018	RCT[a]	313	Investigate the efficacy of internet-based behavioral activation with lay counselor support compared with web-based minimal psychoeducation without support for depression in Indonesia.	Those aged 16 y or older or who met the criteria for major depressive disorder or persistent depressive disorder based on the Structured Clinical Interview for DSM-5[b]	Guided Act and Feel Indonesia—the program consists of a series of 8 weekly structured modules that can be completed in 30-45 min per module, including psychoeducation about depression and the basic background of behavioral activation, monitoring mood and behavior or activities, expansion of potential mood-independent pleasurable activities, and building a strategy for relapse prevention	Alleviate depression
Pot-Kolder et al [21], 2018	RCT	116	Investigate the effects of VR-CBT[c] on paranoid thoughts and social participation	Meets the DSM-4[d] diagnosis of a psychotic disorder or paranoid ideation in the past month or aged 18-65 y	VR-CBT—consisting of 16 individual therapy sessions	Improve social participation in people with paranoia
Enander et al [22], 2016	RCT	94	Evaluate the efficacy of therapist guided internet based CBT[e] program for body dysmorphic disorder (BDD-NET) compared with web-based supportive therapy	Aged 18 y or over or a principal diagnosis of body dysmorphic disorder according to the DSM-5	Therapist-guided, internet-based CBT program for body dysmorphic disorder (BDD-NET) - 12 wk long	Decrease the severity of body dysmorphic disorder symptoms
Nissling et al [23], 2020	Single arm (pre- and postdesign)	9	Assess patient experiences; the feasibility, safety, and acceptability; and preliminary effectiveness on anxiety and depression, empowerment, and adherence to treatment in an 8-wk peer-supported iCBT[f] program for patients with anxiety disorders treated in primary care	Aged 18+ y or meeting the diagnostic criteria for an anxiety disorder	iCBT+peer support−includes 13 different tools, and the treatment consists of 8 modules meant to be completed within 8 wk	Alleviate anxiety
Hamatani et al [24], 2019	Single arm (pre- and postdesign)	7	Evaluate the feasibility of iCBT via videoconference for patients with bulimia nervosa or binge-eating disorder	A primary diagnosis of bulimia nervosa or binge-eating disorder according to the DSM-5 criteria or female or aged 16-65 y	iCBT via videoconference—16 weekly sessions via videoconference with real-time therapist support	Reduce eating disorder symptoms
van Luenen et al, [25], 2018	RCT	188	Investigate the effectiveness of the intervention on depressive symptoms in people living with HIV	HIV positive for at least 6 mo or aged 17+ y or presence of mild to moderate depressive symptoms (PHQ-9[g] score>4 and <20)	8-wk-long internet-based intervention (available in Dutch and English) consisted of cognitive behavioral therapy, with minimal telephone coaching	Alleviate depression
Freeman et al [26], 2017	RCT	3755	Assess whether treating insomnia leads to a reduction in paranoia and hallucinations	Attending university or positive screen for insomnia, as indicated by a score of 16 or lower on the SCI[h] or aged 18+ y	Sleepio—digital CBT for insomnia—6 weekly sessions lasting an average of 20 min each	Decrease paranoia and hallucinations

Study, year	Study design	Total number of participants, N	Study aim	Participants	Intervention	Intervention desired outcome
Görges et al [27], 2018	Single arm or pre- and postdesign	81	Investigate the level of satisfaction with a positive psychology web-based training among patients with mild and moderate depression or dysthymia	Aged 18+ y or diagnosis of unipolar depression or dysthymia past or present according to the MIII[i]	"Glück kommt selten allein"—the program comprises 7 modules; each module comprises 2-3 exercises and is meant to be completed within 1 wk	Reduce depression symptoms
Krupnick et al [28], 2017	RCT	34	Assess feasibility, acceptability, and safety of the intervention under study	Aged 18+ y or veteran or had an intake session through the Trauma Services Program	WIRED[j]—10 sessions of a writing intervention	Alleviate PTSD[k] symptoms
Bragesjö et al [29], 2023	RCT	102	Assess the efficacy of CIPE[l] as compared with the waiting list in a larger sample and with a longer controlled follow-up period	Adult or exposed to a traumatic event within 2 mo or at least some symptoms of posttraumatic stress	CIPE - 3 wk long	Alleviate PTSD symptoms
Trottier et al [30], 2022	Single arm or pre- and postdesign	21	Assess the feasibility, acceptability, and initial efficacy of RESTORE[m] in health care workers on the frontline of the COVID-19 pandemic	Canadian health care worker, first responder, or military member who experienced a traumatic or extremely stressful event related to COVID-19 in the course of their work or moderate or more severe symptoms of anxiety, depression, or PTSD symptoms	RESTORE is a web-based guided transdiagnostic intervention, including cognitive-behavioral interventions	Alleviate anxiety, depression, and PTSD
Gumley et al [31], 2022	RCT	74	Establish the feasibility of undertaking a definitive RCT to determine the effectiveness of a blended digital intervention for relapse prevention in schizophrenia	Aged 16+ y or had a schizophrenia or related diagnosis confirmed via case records or experienced a relapse within the previous 2 y	EMPOWER[n]—designed to enable participants to monitor changes in their well-being daily using a mobile phone. Participants could use their own mobile phone.	Prevent relapse
Torok et al [32], 2022	RCT	455	Investigate the efficacy of the life-buoy smartphone app in reducing the severity of suicidal thoughts when compared with an attention-matched smartphone app (LifeBuoy-C)	Between 18 and 25 y of age or in the community (nonclinical sample) or responded in the positive to the question "have you experienced suicidal thoughts in the past 12 months?"	LifeBuoy—a 6-wk self-guided smartphone app based on DBT[o] to improve emotion regulation and distress tolerance	Decrease the severity of suicidal ideation symptoms
Bucci et al [33], 2018	RCT	36	Assesses the feasibility and acceptability of Actissist, a digital health intervention grounded in the cognitive model of psychosis that targets key early psychosis domains	In current contact with Early Intervention Services or at least 4-wk stabilization of positive symptoms (score <3 on the PANSS[p] items) or aged 16+ y	Actissist—a 12-wk digital health intervention grounded in the cognitive model of psychosis that targets key early psychosis domains	Alleviate psychotic symptoms
Steare et al [34], 2020	RCT	40	Test the feasibility and acceptability of a RCT to evaluate a smartphone-based self-management tool in Early Intervention in Psychosis services	Aged 16+ y or had experienced at least one episode of psychosis	My journey 3—designed to help Early Intervention in Psychosis service users recognize early warning signs of illness, recognize and monitor symptoms, and create plans for their recovery	Promote self-management posttreatment (psychosis)

Study, year	Study design	Total number of participants, N	Study aim	Participants	Intervention	Intervention desired outcome
Guo et al [35], 2020	RCT	300	Assess the efficacy of a WeChat-based intervention, Run4Love, with a RCT among 300 PLWHD[q] in China	Aged 18+ y or HIV seropositive or elevated depressive symptoms (measured by the CES-D[r] ≥16)	Run4Love—the program comprises of 2 major components: the adapted cognitive behavioral stress management course and physical activity promotion	Alleviate depression
Carl et al [36], 2020	RCT	256	Investigate the efficacy of a novel digital CBT program in those with GAD[s] for outcomes of anxiety, worry, depressive symptoms, sleep difficulty, well-being, and participant - specific quality of life	Aged 18+ y or a diagnosis of general anxiety disorder	Daylight—digital cognitive behavioral therapy for anxiety	Alleviate anxiety
Lim et al [37], 2019	2 groups (pre-post)	20	Evaluate the acceptability, feasibility, and safety of the program in lonely young people with or without a mental-health diagnosis of social anxiety disorder	Aged 18-25 y or engaged with a current mental health service, general practitioner (or was engaged at time of assessment) or current DSM-5 of social anxiety disorder assessed by the SCID[t] or UCLA[u] Loneliness Scale score >38 d	Connect+—the 6-wk program delivers positive psychology content designed to improve relationship quality	Decrease loneliness
Mühlmann et al [38], 2021	RCT	402	Investigate the effectiveness of a guided internet-based self-help program compared with a waiting list control group in reducing sinical ideation	Aged 18+ y with suicidal ideation	Online Living under control self-help program—6 modules of a self-help program for suicidal ideation based on CBT	Decrease suicidal ideation
Yeung et al [39], 2018	RCT	75	Evaluate the acceptability, feasibility, and effectiveness of using the Chinese translated version of MoodGYM (MoodGYM [C]) as an adjunctive intervention for the treatment of depressive symptoms in patients at outpatient clinics in a hospital in Beijing, China	Aged 18+ y or significant depressive symptoms as judged by the patients' treating clinicians	MoodGYM—4 wk of a web-based computerized CBT	Alleviate depression
Fornells-Ambrojo et al [40], 2008	Single arm or pre- and postdesign	20	Investigate the acceptability and safety of using VR[v] with individuals with current persecutory delusions	Diagnosis of nonaffective psychosis or a score of at least moderate severity on the Suspiciousness item (P6) of the PANSS or current persecutory delusion	VR—a VR underground train containing neutral characters	Use of VR with individuals with persecutory delusions
Freeman et al [41], 2022	RCT	346	Evaluate the efficacy of an automated VR cognitive therapy (gameChange) to treat avoidance and distress in patients with psychosis, and to analyze how and in whom it might work	Aged 16+ y or with a clinical diagnosis of a schizophrenia spectrum disorder or an affective diagnosis with psychotic symptoms or self-reported difficulties going outside due to anxiety	The gameChange VR therapy (6 wk) aims for participants to relearn safety by testing their fear expectations around other people	Decrease agoraphobia symptoms

Study, year	Study design	Total number of participants, N	Study aim	Participants	Intervention	Intervention desired outcome
Garety et al [4], 2021	RCT	362	Examine the effectiveness of SlowMo therapy in reducing paranoia and in improving reasoning, quality of life and well-being, and to examine its mechanisms of action, moderators of effects and acceptability	Aged 18+ y or persistent (≥3 mo) distressing paranoia or a diagnosis of schizophrenia-spectrum psychosis	SlowMo—consists of 8 individual face-to-face sessions, with each module addressing a specific topic and typically lasting 60-90 min. The therapy was delivered by trained therapists within a 12-wk time frame and was assisted by a web-based application, delivered using the SlowMo web app.	Decrease paranoia

[a]RCT: randomized controlled trial.

[b]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[c]VR-CBT: virtual reality–based cognitive behavioral therapy.

[d]DSM-4: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.

[e]CBT: cognitive behavioral therapy.

[f]iCBT: internet-based cognitive behavioral therapy.

[g]PHQ-9: Patient Health Questionnaire-9.

[h]SCI: sleep condition indicator.

[i]MIII: Mini International Neuropsychiatric Interview.

[j]WIRED: warriors internet Recovery and Education.

[k]PTSD: posttraumatic stress disorder.

[l]CIPE: condensed internet-delivered prolonged exposure.

[m]RESTORE: Recovering from Extreme Stressors Through Online Resources and E-health.

[n]EMPOWER: Early Signs Monitoring to Prevent Relapse in Psychosis and Promote Well-Being, Engagement, and Recovery.

[o]DBT: dialectical behavior therapy.

[p]PANSS: Positive and Negative Syndrome Scale.

[q]PLWHD: people living with HIV and depression.

[r]CES-D: Centre for Epidemiologic Studies-Depression Scale.

[s]GAD: General Anxiety Disorder.

[t]SCID: Structured Clinical Interview for DSM Disorders.

[u]UCLA: University of California, Los Angeles.

[v]VR: virtual reality.

Risk of Bias in Studies

The RoB 2 tool was used to assess risk of bias in the 17 included RCTs. More than half (9/17, 53%) of the RCTs has some risk of bias. A total of 4 (24%) RCTs had an overall low risk of bias, and another 4 (24%) had an overall high risk of bias. Randomization, missing outcome data, and outcome measurements were the main sources of bias. See Multimedia Appendix 2 [4,18,20-22,25,26,28,29,31-36,38,39,41] for more details on the RoB 2 tool findings. The CASP tool was used to assess the risk of bias in the 6 pre-post studies. The main risk of bias among these studies was identifying and accounting for the confounding variables. See Multimedia Appendix 3 [19,23,24,27,30,37,40] for the results of the CASP tool.

Results Synthesis

Assessing Safety

The main method used to assess safety in the included studies was to collect and report AEs data during clinical trials or other research studies. Most studies (10/23, 57%) actively collected AE data, that is, proactively and systematically asked participants about the occurrence of an AE [4,20,22,24,27,29,31-33,36-38,40]; 2 (9%) studies passively collected AE data, that is, only recorded AE data that were spontaneously reported by participants [26,41]; and more than one-third (8/23, 35%) of the studies did not collect any safety data to assess risk [21,23,25,28,30,34,35,39].

Studies have varied widely in terms of the collection and monitoring of safety data. Some collected only SAE data [33] or only current suicide-related AE data [33,38], whereas others collected self-report measures (no further details were provided) [22,29] or used standardized measures, such as the Fear of Recurrence Scale [31] and the Symptom Checklist [36], while others specifically collected information on symptom deterioration [20]. A study using a VR-based intervention asked participants whether they experienced anxiety, nausea, or disorientation (known side effects of VR) after the intervention [40]. In another study, therapists asked participants about their general physical and mental health and encouraged them to report AEs via email [24]. Lim et al [37] *measured the frequency of AE data* (without specifying how) throughout the study (33 days) and assessed seriousness but found none. Garety et al [4] actively collected AE data over the period of the trial (24-week follow-up) and categorized AEs based on their severity (mild,

moderate, severe), relatedness to the intervention, and seriousness. Görges et al [27] actively collected AE data during the 4 main assessment points of the trial and categorized them based on relatedness to the intervention; they also conducted the Patient Health Questionnaire-9 weekly to identify any deterioration in mood or suicidal ideation and conducted the Inventory for the Assessment of Negative Effects of Psychotherapy.

Studies that actively collected risk or safety data varied according to how often they collected the data. Some risk data were collected after every session [24,40], during the main study time points [27,32], weekly [29], mid- and posttreatment [22], postintervention [36], when triggered by a self-report measure [20,38], and throughout the trial plus the follow-up period [4]. Others were unclear about how often they collected the AE data [31,33,37].

Reporting Risk

Only one study (EMPOWER [Early Signs Monitoring to Prevent Relapse in Psychosis and Promote Well-Being, Engagement, and Recovery], an app to help prevent relapse in psychosis) reported an SAE related to the DMHI. In this instance, a participant was admitted to the hospital for a physical health complaint that they considered related to feeling overwhelmed by installing the app [31]. Across studies, the main reported AE related to the DMHI was symptom deterioration [27,31,36,38]. Other reported risks include increased anxiety, distress, or depression [22,28,31], triggering traumatic memory [31], increased sleep disturbances [22], and frequent self-monitoring resulting in distress and unhelpful rumination [31]. Technical difficulties were reported in 2 studies [31,36]. In 1 study, 22.6% of the participants felt that they or their problems were not taken seriously by the intervention and 8.1% felt dependent on the program [27].

A total of 7 studies reported that they found no AEs [20,21,24-26,35,37]. However, 2 of these studies reported that participants experienced a deterioration in symptoms [25] and novel symptoms in which sleep treatment led to a sustained increase in mania [26]. Similarly, in another 5 studies, AE data were reported but AEs were not considered by the authors (not reported as such), such as symptom deterioration [23,32,36], changes in medication or treatment [28], and technical problems [34]. See Tables 2 and 3 for more details on the main AE-related findings per study included.

Table 2. Studies' risk-related qualitative findings.

Study, year	AE[a] or risk or harm or negative effects main findings
Arjadi et al [20], 2018	No AEs were reported in either group.
Pot-Kolder et al [21], 2018	No AEs were reported in either group.
Enander et al [22], 2016	No SAEs[b] were reported. 15 (32%) participants in the BDD-NET group and 6 (13%) in the supportive therapy group reported mild AEs (increased levels of anxiety and general negative well-being) at the beginning of the trial, which had subsided for everyone at 3 mo, except for 4 participants in the BDD-NET group. Of these, 2 participants reported increased sleep disturbances because of heightened anxiety levels attributed to the exposure exercises, 1 reported depressive mood, and 1 reported that the insight gained throughout the treatment regarding time spent on concerns about appearance was emotionally painful but also enhanced motivation to make changes. After the start of treatment, 1 participant in the BDD-NET group had been prescribed an antidepressant. At follow-up, 3 participants in the BDD-NET group and 2 participants in the control received a new additional treatment. The occurrence of adverse events during treatment was not related to responder status at follow-up (χ^2_1=0.9; *P*=.34).
Nissling et al [23], 2020	No SAEs were reported during the treatment period or in the interviews with the participants.
Hamatani et al [24], 2019	No mental or physical AEs were reported.
van Luenen et al [25], 2018	No AEs were reported.
Freeman et al [26], 2017	No AEs were reported.
Görges et al [27], 2018	At follow-up, 14.8% of the participants showed worse depression. Subjective deterioration attributed to the program was at 4.8%. A total of 4 AEs (deteriorations in mood) were *possibly related* to the intervention. According to the INEP[c], 8.4% of the participants experienced a negative effect at posttreatment, 2.2% reported negative effects due to the program and 6.2% due to other reasons. At follow-up, negative effects rose to 12.2%. However, negative effects attributed to the program decreased to 1.7%. At posttreatment, 22.6% of the participants felt that they or their problems were not taken seriously by the program, and 8.1% of the participants stated that they felt dependent on the content of the program at posttreatment. However, at follow-up, none of the participants felt this way anymore.
Krupnick et al [28], 2017	Two people chose to get in touch with the therapist during the course of the study because of distress following an early session, but others may have simply chosen to stop the intervention in the face of discomfort.
Bragesjö et al [29], 2023	No SAEs were found. In the treatment group, 16 participants (31%) reported a total of 63 AEs (mild to moderate). In the waiting list, 11 participants (21%) reported a total of 35 AEs. Note that in the 27 participants that reported AEs, the average number of reported events were 4 in the treatment group and 3 in the waiting list group. AEs reported: increase in number of intrusive memories of the index event and a previous event, initial symptom exacerbation, increase in depressive symptoms or anxiety, sleep problems, panic attacks, increase in stress, anger or irritability, severe distress during exposure, tiredness, memory impairment, migraine, increase in impulsivity and pain.
Trottier et al [30], 2022	There were no participants withdrawn for safety-related reasons.
Gumley et al [31], 2022	There was a total of 54 AEs, affecting 29 people. Around half of all events across arms were classified as SAEs, and the vast majority of these were anticipated. There was one death during the study. Six events were related to a study procedure, 1 of which was serious (threat made to a member of research). There were 13 app-related AEs, affecting 11 people, 1 of which was serious (brief hospital admission for a physical health complaint, which the service user described as being in part related to feeling overwhelmed by the recent installation of the app—withdrew from the study). Nonserious app-related AEs: 4 instances in which the app caused unhelpful rumination. In 1 of these instances, where the self-monitoring approach was described as counter to the service user's usual coping strategy of "burying things," the participant withdrew from the study. Other participants described feeling forced to think about being unwell because of questions in the app, with 1 person suggesting less frequent monitoring in future iterations. Unhelpful rumination of this type was identified by 1 participant as an issue when they were well, whereas a second person was affected when they felt more depressed. Two participants specifically cited increased paranoia because of the app. A further participant identified that personalized question content unhelpfully triggered traumatic memories of psychosis. One participant reported experiencing increased anxiety. In 1 case a participant reported increased worry because of losing their provided mobile phone. One participant experienced distress because of a technical fault. The study also reported on intensity, relatedness, and whether it was anticipated or not.
Torok et al [32], 2022	No SAEs (suicide attempts requiring medical care) were reported.
Bucci et al [33], 2018	No SAEs were reported.
Steare et al [34], 2020	No SAEs were reported.
Guo et al [35], 2020	No AEs were reported.

Study, year	AE[a] or risk or harm or negative effects main findings
Carl et al [36], 2020	One participant only, in the treatment group, reported experiencing an AE related to difficulty accessing the intervention, which the participant deemed as distressing and as contributing to increased anxiety. This event was reported to the Ethics Committee. Participants in the treatment group reported significantly fewer occurrences of unwanted symptoms during the study period, including low mood, fatigue or exhaustion, extreme sleepiness, feeling agitated, difficulty remembering things, headache or migraine, difficulty concentrating and focusing on things, reduced motivation or energy, blurred vision, dizziness, and feeling irritable.
Lim et al [37], 2019	No AEs were reported.
Mühlmann et al [38], 2021	In all, 28 (16.8%) participants reported negative effects: 1 unclassifiable, 19 negative emotions or felt worse, 5 increase in suicidal ideation, 3 felt stressed or guilty for not having worked more on the program. Half of the participants that reported negative effects experienced clinically significant improvement in suicidal ideation postintervention. In all, 27 (6.7%) participants—12 in the control group versus 15 in the intervention group—had attempted suicide within the first 6 wk. A total of 44 participants (10.9%) had attempted suicide during the entire period of the study, 22 in each group. Four deaths were reported between postintervention and follow-up, 2 in each group (2 by suicide—1 in the intervention group and 1 in the control group).
Yeung et al [39], 2018	No SAEs were reported.
Fornells-Ambrojo et al [40], 2008	The VR[d] experience did not raise levels of anxiety or symptoms of simulator sickness. No side effects were reported at the follow-up.
Freeman et al [41], 2022	There were 25 AEs (in 21 patients) in the VR therapy group and 29 AEs (in 19 patients) in the usual care alone group (P=.66). There were 12 SAEs (in 9 patients) in the VR therapy group and 8 SAEs (in 7 patients) in the usual care alone group (P=.37).
Garety et al [4], 2021	In all, 19 participants in the treatment group and 21 participants in the control group reported 54 AEs (51 serious events, no deaths). More than half of the SAEs were mental health hospital admission or crisis referrals (SlowMo, n=13; TAU[e], n=16) or physical health crises (SlowMo, n=8; TAU, n=2), none of which was rated as being related to participation in the trial. One SAE in the TAU group was rated as “definitely related” to trial involvement: it involved a complaint made when the research team shared information with the clinical team under a duty of care. None of the AEs were related to the treatment. The types of AEs reported included physical, self-harm, serious violent incident (survivor or accused), referrals to crisis care, admission to psychiatric hospital, and other, along with intensity and relatedness to the intervention.

[a]AE: adverse event.

[b]SAE: serious adverse event.

[c]INEP: Inventory for the Assessment of Negative Effects of Psychotherapy.

[d]VR: virtual reality.

[e]TAU: treatment as usual.

Table 3. Studies' risk-related quantitative findings.

Study, year	Number of participants, N	Intervention group sample size, n (%)	Control group sample size, n	Total number of SAEs[a], n	Total number of AEs[b], n	Number of participants who experienced an AE in the intervention group, n (%)	Number of participants who experienced an AE in the control group, n (%)	Values, OR[c] (95% CI)	*P* value
Arjadi et al [20], 2018	313	159 (50.8)	154 (49.2)	0	0	0 (0)	0 (0)	—[d]	—
Pot-Kolder et al [21], 2018	116	58 (50)	58 (50)	0	0	0 (0)	0 (0)	—	—
Enander et al [22], 2016	94	47 (50)	47 (50)	0	21	15 (31.9)	6 (12.8)	3.2 (1.12-9.19)	.03[e]
Nissling et al [23], 2020	9	N/A[f]	N/A	0	Not reported	N/A	N/A	—	—
Hamatani et al [24], 2019	7	N/A	N/A	0	0	N/A	N/A	—	—
van Luenen et al [25], 2018	188	97 (51.6)	91 (48.4)	0	0	0 (0)	0 (0)	—	—
Freeman et al [26], 2017	3755	1891 (50.4)	1864 (49.6)	0	0	0 (0)	0 (0)	—	—
Görges et al [27], 2018	81	N/A	N/A	0	4	N/A	N/A	—	—
Krupnick et al [28], 2017	34	18 (52.9)	16 (47.1)	0	Not reported	N/A	N/A	—	—
Bragesjö et al [29], 2023	102	51 (50)	51 (50)	0	98	16 (31.4)	11 (21.6)	1.66 (0.68-4.05)	.26
Trottier et al [30], 2022	21	N/A	N/A	0	Not reported	N/A	N/A	—	—
Gumley et al [31], 2022	74	42 (56.8)	31 (41.9)	27	54	29 (69)	25 (80.6)	0.54 (0.18-1.62)	.27
Torok et al [32], 2022	455	228 (50.1)	227 (49.9)	0	Not reported	N/A	N/A	—	—
Bucci et al [33], 2018	36	24 (66.7)	12 (33.3)	0	Not collected	N/A	N/A	—	—
Steare et al [34], 2020	40	20 (50)	20 (50)	0	Not reported	N/A	N/A	—	—
Guo et al [35], 2020	300	150 (50)	150 (50)	0	0	0 (0)	0 (0)	—	—
Carl et al [36], 2020	256	128 (50)	128 (50)	0	1	1 (0)	0 (0)	3.02 (0.12-74.92)	.45
Lim et al [37], 2019	20	9 (45)	11 (55)	0	0	0 (0)	0 (0)	—	—
Mühlmann et al [38], 2021	402	196 (48.8)	206 (51.2)	44	72	22 (11.2; SAE only)	22 (10.7; SAE only)	—	—
Yeung et al [39], 2018	75	37 (49.3)	38 (50.7)	0	Not reported	N/A	N/A	—	—
Fornells-Ambrojo et al [40], 2008	20	N/A	N/A	0	0	N/A	N/A	—	—

Study, year	Number of participants, N	Intervention group sample size, n (%)	Control group sample size, n	Total number of SAEs[a], n	Total number of AEs[b], n	Number of participants who experienced an AE in the intervention group, n (%)	Number of participants who experienced an AE in the control group, n (%)	Values, OR[c] (95% CI)	*P* value
Freeman et al [41], 2022	346	174 (50.3)	172 (49.7)	20	54	30 (17.2)	26 (15.1)	1.17 (0.66-2.08)	.59
Garety et al [4], 2021	362	181 (50)	181 (50)	51	54	25 (13.8)	26 (14.4)	0.96 (0.53-1.73)	.88

[a]SAE: serious adverse event.

[b]AE: adverse event.

[c]OR: odds ratio.

[d]Not available.

[e]Significant at $P<.05$.

[f]N/A: not applicable.

One study analyzed whether negative effects (feeling worse, increased suicidal ideation, feeling stressed, or guilty for using the program) affected the effectiveness of the intervention. They found that out of the 14% (n=28) of people who reported a negative effect, 50% (n=14) experienced clinically significant improvement [38]. However, without a suitable comparator condition (eg, the proportion experiencing improvement in the absence of negative effects), this finding is difficult to interpret.

Odds Ratio Secondary Data Analysis

The data (number of participants who experienced an AE per group) in Table 2 shows that 6 (35%) out of the 17 RCTs collected and reported sufficient data to calculate odds ratios and CIs. Two of these studies conducted similar analyses by statistically comparing the occurrence of AEs between the treatment and control groups, and both found no significant differences [22,41], consistent with our odds ratio findings. In the remaining 4 studies, our odds ratio analysis revealed that one study showed a significantly elevated risk of harm. In this case, those receiving the intervention were over 3 times more likely to experience an AE during the study, compared with those in the control arm. In contrast, the authors reported that the number of AEs experienced during the study was not related to the responder status at follow-up (P=.34) [22]. The authors concluded that their DMHI was safe, with no occurrence of SAEs.

Mitigating Risk—in the Research Stage

Researchers and authors differ in the actions taken to minimize risk in their studies. One study explicitly described providing participants with support resources [20]. One study excluded participants who could not provide contact details for an emergency contact person [38]. Most of the studies minimized risk by excluding specific clinical groups from their samples such as individuals with a high risk of suicide [20,22-24,28-30,32,37-39,41], individuals with bipolar or manic disorder [20,22,24,29,32,36,39], individuals experiencing psychosis or diagnosed with a psychotic disorder [20,22,24,28,29,32,36,37,39], individuals with a personality disorder [4,22,37,41], those at high risk of self-harm [37,39], and individuals with severe depression [29]. Overall, 9 (39%) of the 23 included studies did not explicitly report any other safety precautions [21,25-27,31,33-35,40].

Mitigating Risk—Postmarket

A web-based secondary search was conducted to identify any further available safety information for the interventions under review. Six (26%) of the 23 interventions in this review were accessible to users or patients [22,26,30,33,36,39]. The search identified the safety information for all 6 interventions. This information was available on app stores [33], interventions' websites [30,39], in the instructions for use [26,36], privacy policy [26], and a publicly available analysis that was performed by the Improving Access to Psychological Therapies program in the United Kingdom for one of the products [22].

Different safety strategies have been implemented for different interventions. All 6 interventions identified the indications and contraindications for use. Only 2 studies specified the minimum age required for users to use the intervention [26,39]. One intervention (BDD-NET, an internet-based cognitive behavioral therapy program for body dysmorphic disorder) highlighted the need for therapists to be trained before administering the program [22]. BDD-NET is also the only intervention that sends trigger alerts to therapists regarding potential risks [22]. The MoodGYM program suggested that its users contacted a health professional if they scored above 2 to 3 on their depression quiz [39]. Actissist (an intervention for psychosis) specified that its users should only use the intervention under the supervision of a qualified health care professional [33]. Some interventions highlighted that they are not substitutes for therapy or medication [26,30,33,36]. Three interventions stated that they were not intended for emergency use and provided their users with resources in the case of an emergency [26,36,39]. These interventions also encouraged users to consult their physician if their symptoms worsened [26,36,39]. Finally, only 2 interventions made their AE data publicly available to users through their instructions for use (not in an academic publication) [26,36]. There was no information available on whether any of these 6 interventions continued to monitor risk and update their safety procedures postmarket phase.

Discussion

This is the first review to systematically evaluate literature on the safety of DMHIs. It aimed to better understand how DMHIs assess, report, and mitigate risk and identify any emerging recommendations, building on previous work [6,7].

Principal Findings

Assessing Risk

Only two-thirds (15/23, 65%) of the studies included in this review assessed risk, although this is an improvement compared with the findings of a systematic review conducted in 2017, where none of the 9 included studies assessed safety or the occurrence of AEs [42,43]. We recommend that safety assessment should be systematically assessed in every DMHI study (recommendation 1). However, this is not a new recommendation. In 2014, colleagues were urging researchers to "...systematically probe for negative effects whenever conducting clinical trials involving Internet interventions, as well as to share their findings in scientific journals" [7]. Studies varied in how often they collected risk data, whether it was weekly, after every session, until follow-up, or just postintervention. The minimal approach was the passive collection of spontaneously reported AEs. By contrast, the most proactive actively collected data on any AE (irrespective of relatedness to the intervention) using a regularly repeated measure or set of questions. To effectively assess safety, research using DMHIs should actively and regularly collect safety data both throughout intervention delivery and after the intervention until follow-up (recommendation 2). Others have suggested minimum midtreatment, at the end of treatment, and at one follow-up time point [8]. The frequency of safety data collection should be as high as possible while balancing the burden on the research team and participants. In the digital context, there is potential for automating negative effect reporting within the technology itself [14]. The collection of sufficient safety data observations is also a prerequisite for conducting formal statistical analyses of these data, a point we return to below.

The methods used to assess safety varied widely between studies and included standardized instruments, bespoke questionnaires, or unspecified "self-report." This review highlights the need for a minimum agreed upon standard for assessing risk data in DMHIs. A recent narrative scoping review on identifying and categorizing AEs in DMHIs suggested using the digital functionality of DMHIs to streamline the process of detecting harm and AEs in these interventions [44]. Regardless, the methods and instruments used to measure safety should be reported in sufficient detail to permit replication (recommendation 3).

In the analyses of collected risk data, only 2 studies conducted statistical analyses to compare whether the occurrence of AEs significantly varied between the control and treatment groups. Neither found statistical support for significant differences. This provides evidence consistent with the treatments being safe [22,41]. The odds ratio analysis presented in this review illustrates another possible comparison that could add value but has not yet been used. For example, according to this metric, participants in the intervention group of one study were 3 times more likely to experience AEs compared with those in the control group; however, unaware of this, the study authors deemed the DMHI safe [22]. Thus, rates of harm should be statistically compared between study arms and across different studies using standardized quantitative metrics (recommendation 4). Irrespective of the specific analytical approach used, formal statistical analysis of AE data, interpreted in the wider context of a study, is a key requirement to allow stakeholders to make fully informed, evidence-based judgments about safety. Our suggestions include reporting the mean (and median) SAEs per participant for each study, listing the number of SAEs for each participant, and reporting odds ratios, as illustrated here.

Reporting Risk

Symptom deterioration was one of the main risks reported in this review, similar to face-to-face therapies [13,45]. Some studies have argued that short-term and transient deterioration of symptoms during psychological therapy can be a normal and integral part of treatment [46,47]. Others argue that AEs are negatively correlated with positive therapy outcomes and are not expected to be a lasting consequence of effective therapy [48,49]. If deterioration is an expected part of the treatment, this should be acknowledged a priori, and a quantitative threshold is identified to define this as an AE (recommendation 5). The Australian National Safety and Quality Digital Mental-Health Standards state "Recognizing and responding to acute deterioration" as one of their standards [50]. Moreover, a review on the identification of AEs in DMHIs suggested the development of a digitally delivered symptom checklist and setting predefined cutoffs to detect symptom deterioration in DMHIs [44].

This review also highlighted the lack of agreement and conceptual clarity regarding which events are considered AEs. All 7 studies in this review that reported "no AEs" in their publication reported events that indicated an AE, such as deterioration in symptoms and novel symptoms, but did not categorize them as AEs. This finding is in line with a recent review on the topic, which found that authors presented events that indicated an AE or SAE but did not categorize it as such [44]. The authors of that review speculated that the difficulties faced in classifying AEs in DMHIs were due to the lack of guidelines [44]. Researchers have previously highlighted the need to improve the classification of AEs in DMHIs [7,14]. A consensus classification framework for AEs would allow the field to more clearly establish whether the benefits of a new DMHI outweigh the risks, and how it compares to alternative treatment options.

Mitigating Risk

The main method used to mitigate risk during DMHI trials and research studies was the exclusion of at-risk groups, such as those at high risk of suicide or those diagnosed with a specific mental health condition or severity of the condition. The literature shows that individuals with suicidal thoughts and behaviors are routinely excluded from mental health research because of concerns about safety or lack of resources to implement effective risk management measures [1].

Exclusive practices are unintentionally reinforced by regulatory requirements that prioritize safety over inclusivity, resulting in narrow and specific indications for use. This is justifiable for new, untested treatments. However, once basic safety parameters are established, the ongoing exclusion of more vulnerable groups is a serious limitation that the research community must address. Exclusive practices do not reflect the real world, limit generalizability, and deny the possible benefits of DMHIs for more *risky* segments of the target population [1]. The research community is urged to seek future opportunities to assess the safety of their interventions in specific groups that they have previously excluded [51]. Safety and efficacy should be assessed in high-risk groups with appropriate safeguards (recommendation 6).

The included studies mainly used their risk-related findings to conclude the safety of the DMHI under study rather than to inform future practice, product iterations, or safety measures. Only one study used risk data to make suggestions for possible safety measures needed in similar DMHIs in the future. The authors suspected that one of the modules that required engaging with people was responsible for the deterioration experienced by their users [27]. As a result, they suggested excluding similar exercises from similar interventions or introducing them toward the end to reduce possible negative effects [27]. In addition, most of the safety information provided publicly to users, although necessary, was not provided by the risk data collected in their respective studies. For example, although research studies excluded at-risk groups when assessing safety, these groups were often not excluded when the DMHI was implemented. Safety data collected during the research phase should be used to inform risk-mitigation postmarket (recommendation 7).

It seems that *assessing risk* in the research phase is an independent process from *mitigating risk* in the postmarket phase. We speculate that this was due to the quality of the safety data collected during the research phase. To use risk data from the research phase of the DMHI in the mitigation process, studies need to enhance the assessment and analysis of safety in the ways recommended above. In addition, research is needed to assess the effectiveness of common methods used to mitigate risks in psychotherapy in a digital context. Ideally, such findings can result in a list of actions to mitigate specific identified risks that have been proven to be effective.

Limitations

This study has some limitations. Most RCTs included in this review had some concerns about the risk of bias. More than one-quarter (6/23, 26%) of the included studies were not RCTs, and thus had concerns around "confounding variables." Finally, most of the included studies were conducted in the high-income Western world; thus, their generalizability is limited.

Only studies published in English were included in this review. Only studies including adult populations were included in this review. In addition, our search results were dictated by the words included in our search strategy (safety, risk, negative effects, AEs, etc), meaning that we only extracted and reviewed studies that assessed risk. This may have led to an inflation of studies that assessed and reported AEs, as studies that did not explicitly state that they assessed the risks of a DMHI (eg, did not label their findings as such) will have been excluded.

Conclusions

This review highlights that the approach to assessing and mitigating risk in DMHIs varies widely and is sometimes inadequate. To formulate a set of 7 recommendations (see Textbox 2 for the full list), we focused on relatively indisputable points that are specific, measurable, and achievable. This review also endorses the widely recognized need for collaboration between key stakeholders, including academics, health professionals, developers, product managers, commissioners, and regulatory bodies, to reach a consensus on how the risks of DMHI should be assessed, reported, and mitigated. Standardized definitions and guidelines are needed to provide professionals with tools to reliably assess the safety of their interventions, manage risk, and protect patients and users from unnecessary harm. Finally, research on patients or users' experiences and their concerns about the safety of DMHIs is needed, as it is currently nonexistent.

Textbox 2. Full list of recommendations.

- Recommendation 1: Safety assessment as standard—risk and safety should be systematically and proactively assessed in every digital mental health intervention (DMHI) study
- Recommendation 2: Frequency of safety assessment—risk and safety assessments should take place at prespecified, regular intervals throughout both the intervention and follow-up phases
- Recommendation 3: Measures of safety assessment—the methods and instruments used to measure safety should be reported in sufficient detail to permit replication
- Recommendation 4: Statistical comparison—rates of harm should be statistically compared both between study arms and across different studies using standardized quantitative metrics
- Recommendation 5: Symptom Deterioration—if deterioration is an expected part of treatment, this should be acknowledged a priori, and a quantitative threshold identified for defining this as an adverse event
- Recommendation 6: Inclusivity—safety and efficacy should be assessed in high-risk groups, with appropriate safeguards
- Recommendation 7: Postmarket mitigation—safety data collected during the research phase should be used to inform risk mitigation postmarket

Acknowledgments

The authors thank Dr Charlotte Hall and Dr Aislinn Gómez Bergin from the National Institute for Health and Care Research (NIHR) MindTech MedTech Co-operative and the University of Nottingham Biomedical Research Centre for their comments on an earlier draft of this manuscript. This work was supported by the Medical Research Council Biomedical Catalyst: Developmental Pathway Funding Scheme (DPFS), MRC Reference: MR/V027484/1. The authors would also like to express their gratitude to the NIHR Biomedical Research Centre hosted in South London and Maudsley National Health Service (NHS) Foundation Trust in partnership with King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, Department of Health and Social Care, the Economic and Social Research Council, or King's College London. Finally, the authors gratefully acknowledge the contributions made to this work by the STOP study Lived Experience Panel, supported by Dr Thomas Kabir and the McPin Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Other studies details.
[DOCX File , 24 KB - mental_v10i1e47433_app1.docx]

Multimedia Appendix 2
Version 2 of the Cochrane risk-of-bias tool for randomized trials (RoB 2) generated an output and legend.
[DOCX File , 138 KB - mental_v10i1e47433_app2.docx]

Multimedia Appendix 3
Critical Appraisal Skills Programme tool results.
[DOCX File , 24 KB - mental_v10i1e47433_app3.docx]

Multimedia Appendix 4
PRISMA Checklist.
[PDF File (Adobe PDF File), 106 KB - mental_v10i1e47433_app4.pdf]

References

1. Meyerhoff J, Kruzan KP, Kim KY, Van Orden K, Mohr DC. Exploring the safety of a general digital mental health intervention to effect symptom reduction among individuals with and without suicidal ideation: a secondary analysis. Arch Suicide Res 2023 Jul 12;27(3):966-983. [doi: 10.1080/13811118.2022.2096520] [Medline: 35822235]
2. Batterham PJ, Christensen H, Calear AL, Werner-Seidler A, Kazan D. Rates and predictors of deterioration in a trial of internet-delivered cognitive behavioral therapy for reducing suicidal thoughts. Arch Suicide Res 2022 Nov 20;26(2):937-947. [doi: 10.1080/13811118.2020.1848671] [Medline: 33215554]
3. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
4. Garety P, Ward T, Emsley R, Greenwood K, Freeman D, Fowler D, et al. Digitally supported CBT to reduce paranoia and improve reasoning for people with schizophrenia-spectrum psychosis: the SlowMo RCT. Efficacy Mech Eval 2021 Aug;8(11):1-90. [doi: 10.3310/eme08110] [Medline: 34398537]
5. Pharmacovigilance – how the MHRA monitors the safety of medicines. Medicines and Healthcare Products Regulatory Agency. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949131/Pharmacovigilance___how_the_MHRA_monitors_the_safety_of_medicines.pdf [accessed 2023-07-04]
6. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of Internet interventions. Internet Interv 2014 Mar;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]
7. Klatte R, Strauss B, Flückiger C, Färber F, Rosendahl J. Defining and assessing adverse events and harmful effects in psychotherapy study protocols: a systematic review. Psychotherapy (Chic) 2023 Mar 20;60(1):130-148. [doi: 10.1037/pst0000359] [Medline: 35049321]
8. Gullickson KM, Hadjistavropoulos HD, Dear BF, Titov N. Negative effects associated with internet-delivered cognitive behaviour therapy: an analysis of client emails. Internet Interv 2019 Dec;18:100278 [FREE Full text] [doi: 10.1016/j.invent.2019.100278] [Medline: 31890627]
9. Safety guidelines. International Council for Harmonization. URL: https://www.ich.org/page/safety-guidelines [accessed 2023-03-19]

10. What is a serious adverse event? U.S. Food and Drug Administration. 2023 May 18. URL: https://www.fda.gov/safety/reporting-serious-problems-fda/what-serious-adverse-event [accessed 2023-03-19]
11. Safety and progress reports (other research) procedural table. National Health Service Health Research Authority. URL: https://www.hra.nhs.uk/approvals-amendments/managing-your-approval/safety-reporting/safety-and-progress-reports-other-research-procedural-table/ [accessed 2023-03-19]
12. Linden M, Schermuly-Haupt ML. Definition, assessment and rate of psychotherapy side effects. World Psychiatry 2014 Oct 01;13(3):306-309 [FREE Full text] [doi: 10.1002/wps.20153] [Medline: 25273304]
13. Rozental A, Kottorp A, Forsström D, Månsson K, Boettcher J, Andersson G, et al. The Negative Effects Questionnaire: psychometric properties of an instrument for assessing negative effects in psychological treatments. Behav Cogn Psychother 2019 Sep 15;47(5):559-572. [doi: 10.1017/S1352465819000018] [Medline: 30871650]
14. Gómez Bergin AD, Valentine AZ, Rennick-Egglestone S, Slade M, Hollis C, Hall CL. Identifying and categorizing adverse events in trials of digital mental health interventions: narrative scoping review of trials in the international standard randomized controlled trial number registry. JMIR Ment Health 2023 Feb 22;10:e42501 [FREE Full text] [doi: 10.2196/42501] [Medline: 36811940]
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
16. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
17. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012 Oct 15;22(3):276-282. [doi: 10.11613/BM.2012.031]
18. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
19. Buccheri RK, Sharifi C. Critical appraisal tools and reporting guidelines for evidence-based practice. Worldviews Evid Based Nurs 2017 Dec 12;14(6):463-472. [doi: 10.1111/wvn.12258] [Medline: 28898556]
20. Arjadi R, Nauta MH, Scholte WF, Hollon SD, Chowdhary N, Suryani AO, et al. Internet-based behavioural activation with lay counsellor support versus online minimal psychoeducation without support for treatment of depression: a randomised controlled trial in Indonesia. Lancet Psychiatry 2018 Sep;5(9):707-716. [doi: 10.1016/S2215-0366(18)30223-2] [Medline: 30006262]
21. Pot-Kolder RM, Geraets CN, Veling W, van Beilen M, Staring AB, Gijsman HJ, et al. Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: a single-blind randomised controlled trial. Lancet Psychiatry 2018 Mar;5(3):217-226. [doi: 10.1016/S2215-0366(18)30053-1] [Medline: 29429948]
22. Enander J, Andersson E, Mataix-Cols D, Lichtenstein L, Alström K, Andersson G, et al. Therapist guided internet based cognitive behavioural therapy for body dysmorphic disorder: single blind randomised controlled trial. BMJ 2016 Feb 02;352:i241 [FREE Full text] [doi: 10.1136/bmj.i241] [Medline: 26837684]
23. Nissling L, Fahlke C, Lilja JL, Skoglund I, Weineland S. Primary care peer-supported internet-mediated psychological treatment for adults with anxiety disorders: mixed methods study. JMIR Form Res 2020 Aug 20;4(8):e19226 [FREE Full text] [doi: 10.2196/19226] [Medline: 32815819]
24. Hamatani S, Numata N, Matsumoto K, Sutoh C, Ibuki H, Oshiro K, et al. Internet-based cognitive behavioral therapy via videoconference for patients with bulimia nervosa and binge-eating disorder: pilot prospective single-arm feasibility trial. JMIR Form Res 2019 Oct 23;3(4):e15738 [FREE Full text] [doi: 10.2196/15738] [Medline: 31647472]
25. van Luenen S, Garnefski N, Spinhoven P, Kraaij V. Guided internet-based intervention for people with HIV and depressive symptoms: a randomised controlled trial in the Netherlands. Lancet HIV 2018 Sep;5(9):e488-e497. [doi: 10.1016/S2352-3018(18)30133-4] [Medline: 30135045]
26. Freeman D, Sheaves B, Goodwin GM, Yu LM, Nickless A, Harrison PJ, et al. The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. Lancet Psychiatry 2017 Oct;4(10):749-758 [FREE Full text] [doi: 10.1016/S2215-0366(17)30328-0] [Medline: 28888927]
27. Görges F, Oehler C, von Hirschhausen E, Hegerl U, Rummel-Kluge C. GET.HAPPY - acceptance of an internet-based self-management positive psychology intervention for adult primary care patients with mild and moderate depression or dysthymia: a pilot study. Internet Interv 2018 Jun;12:26-35 [FREE Full text] [doi: 10.1016/j.invent.2018.03.001] [Medline: 30135766]
28. Krupnick JL, Green BL, Amdur R, Alaoui A, Belouali A, Roberge E, et al. An Internet-based writing intervention for PTSD in veterans: a feasibility and pilot effectiveness trial. Psychol Trauma 2017 Jul;9(4):461-470. [doi: 10.1037/tra0000176] [Medline: 27607767]
29. Bragesjö M, Arnberg FK, Olofsdotter Lauri K, Aspvall K, Särnholm J, Andersson E. Condensed Internet-delivered prolonged exposure provided soon after trauma: a randomised trial. Psychol Med 2023 Apr 14;53(5):1989-1998 [FREE Full text] [doi: 10.1017/S0033291721003706] [Medline: 37310324]

30. Trottier K, Monson CM, Kaysen D, Wagner AC, Liebman RE, Abbey SE. Initial findings on RESTORE for healthcare workers: an internet-delivered intervention for COVID-19-related mental health symptoms. Transl Psychiatry 2022 Jun 01;12(1):222 [FREE Full text] [doi: 10.1038/s41398-022-01965-3] [Medline: 35650179]
31. Gumley AI, Bradstreet S, Ainsworth J, Allan S, Alvarez-Jimenez M, Birchwood M, et al. Digital smartphone intervention to recognise and manage early warning signs in schizophrenia to prevent relapse: the EMPOWER feasibility cluster RCT. Health Technol Assess 2022 May;26(27):1-174 [FREE Full text] [doi: 10.3310/HLZE0479] [Medline: 35639493]
32. Torok M, Han J, McGillivray L, Wong Q, Werner-Seidler A, O'Dea B, et al. The effect of a therapeutic smartphone application on suicidal ideation in young adults: findings from a randomized controlled trial in Australia. PLoS Med 2022 May 31;19(5):e1003978 [FREE Full text] [doi: 10.1371/journal.pmed.1003978] [Medline: 35639672]
33. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080 [FREE Full text] [doi: 10.1093/schbul/sby032] [Medline: 29566206]
34. Steare T, O'Hanlon P, Eskinazi M, Osborn D, Lloyd-Evans B, Jones R, et al. Smartphone-delivered self-management for first-episode psychosis: the ARIES feasibility randomised controlled trial. BMJ Open 2020 Aug 26;10(8):e034927 [FREE Full text] [doi: 10.1136/bmjopen-2019-034927] [Medline: 32847902]
35. Guo Y, Hong YA, Cai W, Li L, Hao Y, Qiao J, et al. Effect of a WeChat-based intervention (Run4Love) on depressive symptoms among people living with HIV in China: a randomized controlled trial. J Med Internet Res 2020 Feb 09;22(2):e16715 [FREE Full text] [doi: 10.2196/16715] [Medline: 32044751]
36. Carl JR, Miller CB, Henry AL, Davis ML, Stott R, Smits JA, et al. Efficacy of digital cognitive behavioral therapy for moderate-to-severe symptoms of generalized anxiety disorder: a randomized controlled trial. Depress Anxiety 2020 Dec 29;37(12):1168-1178. [doi: 10.1002/da.23079] [Medline: 32725848]
37. Lim MH, Rodebaugh TL, Eres R, Long KM, Penn DL, Gleeson JF. A pilot digital intervention targeting loneliness in youth mental health. Front Psychiatry 2019 Aug 23;10:604 [FREE Full text] [doi: 10.3389/fpsyt.2019.00604] [Medline: 31507469]
38. Mühlmann C, Madsen T, HjorthøjC C, Forman JL, Kerkhof AJ, Nordentoft M, et al. Effectiveness of an internet-based self-help therapy program for suicidal ideation with follow-up at 6 months: results of a randomized controlled trial. J Clin Psychiatry 2021 Aug 31;82(5):20m13803. [doi: 10.4088/JCP.20m13803] [Medline: 34464522]
39. Yeung A, Wang F, Feng F, Zhang J, Cooper A, Hong L, et al. Outcomes of an online computerized cognitive behavioral treatment program for treating chinese patients with depression: a pilot study. Asian J Psychiatr 2018 Dec;38:102-107. [doi: 10.1016/j.ajp.2017.11.007] [Medline: 29146042]
40. Fornells-Ambrojo M, Barker C, Swapp D, Slater M, Antley A, Freeman D. Virtual reality and persecutory delusions: safety and feasibility. Schizophr Res 2008 Sep;104(1-3):228-236. [doi: 10.1016/j.schres.2008.05.013] [Medline: 18571899]
41. Freeman D, Lambe S, Kabir T, Petit A, Rosebrock L, Yu LM, gameChange Trial Group. Automated virtual reality therapy to treat agoraphobic avoidance and distress in patients with psychosis (gameChange): a multicentre, parallel-group, single-blind, randomised, controlled trial in England with mediation and moderation analyses. Lancet Psychiatry 2022 May;9(5):375-388 [FREE Full text] [doi: 10.1016/S2215-0366(22)00060-8] [Medline: 35395204]
42. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological momentary assessment and intervention in the treatment of psychotic disorders: a systematic review. Psychiatr Serv 2017 Nov 01;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523] [Medline: 28669284]
43. Digital clinical safety strategy. National Health Service. 2021 Sep 17. URL: https://transform.england.nhs.uk/key-tools-and-info/digital-clinical-safety-strategy/ [accessed 2022-11-04]
44. Gómez Bergin AD, Valentine AZ, Rennick-Egglestone S, Slade M, Hollis C, Hall CL. Identifying and categorizing adverse events in trials of digital mental health interventions: narrative scoping review of trials in the international standard randomized controlled trial number registry. JMIR Ment Health 2023 Feb 22;10:e42501 [FREE Full text] [doi: 10.2196/42501] [Medline: 36811940]
45. Berk M, Parker G. The elephant on the couch: side-effects of psychotherapy. Aust N Z J Psychiatry 2009 Sep 01;43(9):787-794. [doi: 10.1080/00048670903107559] [Medline: 19670051]
46. Nutt DJ, Sharpe M. Uncritical positive regard? Issues in the efficacy and safety of psychotherapy. J Psychopharmacol 2008 Jan 22;22(1):3-6. [doi: 10.1177/0269881107086283] [Medline: 18187527]
47. Curran J, Parry GD, Hardy GE, Darling J, Mason AM, Chambers E. How does therapy harm? A model of adverse process using task analysis in the meta-synthesis of service users' experience. Front Psychol 2019 Mar 13;10:347 [FREE Full text] [doi: 10.3389/fpsyg.2019.00347] [Medline: 30930805]
48. Moritz S, Nestoriuc Y, Rief W, Klein JP, Jelinek L, Peth J. It can't hurt, right? Adverse effects of psychotherapy in patients with depression. Eur Arch Psychiatry Clin Neurosci 2019 Aug 7;269(5):577-586. [doi: 10.1007/s00406-018-0931-1] [Medline: 30088072]
49. Moritz S, Fieker M, Hottenrott B, Seeralan T, Cludius B, Kolbeck K, et al. No pain, no gain? Adverse effects of psychotherapy in obsessive–compulsive disorder and its relationship to treatment gains. J Obsessive Compuls Relat Disord 2015 Apr;5:61-66. [doi: 10.1016/j.jocrd.2015.02.002]
50. Brown P, Prest B, Miles P, Rossi V. The development of national safety and quality digital mental health standards. Australas Psychiatry 2022 Apr 27;30(2):154-157. [doi: 10.1177/10398562211042361] [Medline: 34569321]

51. Kalmbach DA, Cheng P, O'Brien LM, Swanson LM, Sangha R, Sen S, et al. A randomized controlled trial of digital cognitive behavioral therapy for insomnia in pregnant women. Sleep Med 2020 Aug;72:82-92 [FREE Full text] [doi: 10.1016/j.sleep.2020.03.016] [Medline: 32559716]

Abbreviations

AE: adverse event
CASP: Critical Appraisal Skills Programme
DMHI: digital mental health intervention
EMPOWER: Early Signs Monitoring to Prevent Relapse in Psychosis and Promote Well-Being, Engagement, and Recovery
PICO: Participants, Intervention, Comparison, and Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB 2: Version 2 of the Cochrane risk-of-bias tool for randomized trials
SAE: serious adverse event
VR: virtual reality

Edited by J Torous; submitted 20.03.23; peer-reviewed by R Harrison, S Locke; comments to author 22.06.23; revised version received 05.07.23; accepted 06.07.23; published 09.10.23.

Please cite as:
Taher R, Hsu CW, Hampshire C, Fialho C, Heaysman C, Stahl D, Shergill S, Yiend J
The Safety of Digital Mental Health Interventions: Systematic Review and Recommendations
JMIR Ment Health 2023;10:e47433
URL: https://mental.jmir.org/2023/1/e47433
doi:10.2196/47433
PMID:37812471

Review

The Effectiveness of Fully Automated Digital Interventions in Promoting Mental Well-Being in the General Population: Systematic Review and Meta-Analysis

Julia Groot[1], MSc; Alexander MacLellan[1], MSc; Madelaine Butler[1]; Elisa Todor[1], BSc; Mahnoor Zulfiqar[1], BSc; Timothy Thackrah[2], BDS; Christopher Clarke[3], PhD; Mark Brosnan[4], PhD; Ben Ainsworth[5], PhD

[1]Department of Psychology, University of Bath, Bath, United Kingdom

[2]Cyberlimbic Systems Ltd, London, United Kingdom

[3]Department of Computer Science, University of Bath, Bath, United Kingdom

[4]Centre for Applied Autism Research, Department of Psychology, University of Bath, Bath, United Kingdom

[5]School of Psychology, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Julia Groot, MSc
Department of Psychology
University of Bath
Claverton Down
Bath, BA2 7AY
United Kingdom
Phone: 44 01225 383800
Email: jmdg20@bath.ac.uk

Abstract

Background: Recent years have highlighted an increasing need to promote mental well-being in the general population. This has led to a rapidly growing market for fully automated digital mental well-being tools. Although many individuals have started using these tools in their daily lives, evidence on the overall effectiveness of digital mental well-being tools is currently lacking.

Objective: This study aims to review the evidence on the effectiveness of fully automated digital interventions in promoting mental well-being in the general population.

Methods: Following the preregistration of the systematic review protocol on PROSPERO, searches were carried out in MEDLINE, Web of Science, Cochrane, PsycINFO, PsycEXTRA, Scopus, and ACM Digital (initial searches in February 2022; updated in October 2022). Studies were included if they contained a general population sample and a fully automated digital intervention that exclusively used psychological mental well-being promotion activities. Two reviewers, blinded to each other's decisions, conducted data selection, extraction, and quality assessment of the included studies. Narrative synthesis and a random-effects model of per-protocol data were adopted.

Results: We included 19 studies that involved 7243 participants. These studies included 24 fully automated digital mental well-being interventions, of which 15 (63%) were included in the meta-analysis. Compared with no intervention, there was a significant small effect of fully automated digital mental well-being interventions on mental well-being in the general population (standardized mean difference 0.19, 95% CI 0.04-0.33; *P*=.02). Specifically, mindfulness-, acceptance-, commitment-, and compassion-based interventions significantly promoted mental well-being in the general population (*P*=.006); insufficient evidence was available for positive psychology and cognitive behavioral therapy–based interventions; and contraindications were found for integrative approaches. Overall, there was substantial heterogeneity, which could be partially explained by the intervention duration, comparator, and study outcomes. The risk of bias was high, and confidence in the quality of the evidence was very low (Grading of Recommendations, Assessment, Development, and Evaluations), primarily because of the high rates of study dropout (average 37%; range 0%-85%) and suboptimal intervention adherence (average 40%).

Conclusions: This study provides a novel contribution to knowledge regarding the effectiveness, strengths, and weaknesses of fully automated digital mental well-being interventions in the general population. Future research and practice should consider these findings when developing fully automated digital mental well-being tools. In addition, research should aim to investigate positive psychology and cognitive behavioral therapy–based tools as well as develop further strategies to improve adherence and

reduce dropout in fully automated digital mental well-being interventions. Finally, it should aim to understand when and for whom these interventions are particularly beneficial.

Trial Registration: PROSPERO CRD42022310702; https://tinyurl.com/yc7tcwy7

(JMIR Ment Health 2023;10:e44658) doi:10.2196/44658

KEYWORDS

mental well-being; promotion; intervention; digital; web-based; apps; mobile phone

Introduction

General Background

Mental well-being is commonly defined as a complex construct that includes a subjective experience (subjective well-being, which is often referred to as "happiness") [1] and a process of self-realization (psychological well-being) [2,3]. Traditionally, it was thought that mental well-being would arise in the absence of mental illness, as they were considered opposite ends of 1 continuum [4]. However, the absence of mental illness was found to be insufficient to produce good mental well-being [5]. The dual-continuum model has identified that mental well-being and mental illness are 2 distinct but related continua instead [6], both of which could be considered part of mental health [7]. It is important to focus exclusively on the effective promotion of mental well-being [8], as only a small proportion of the general population has optimal levels of mental well-being [7,9].

In addition, mental well-being in the general population is crucial for allowing society and the individuals within it to thrive. Improved mental well-being is connected to increased productivity, personal growth, a higher quality of life, stronger social cohesion, and more fulfilling and lasting relationships, as well as a decreased likelihood of developing diseases and mental illnesses and a longer lifespan [5,7,10,11]. Promoting mental well-being in the general population is therefore considered a fundamental goal by the World Health Organization (WHO), as described in the Mental Health Action Plan 2013-2030 [12]. Mental well-being promotion interventions provide "various activities or practices that aim to promote, build on, increase or foster primarily individuals' strengths, resourcefulness or resiliency" [10].

Evidence suggests that a variety of psychological approaches are effective in promoting mental well-being, including acceptance and commitment therapy (ACT), compassion, cognitive behavioral therapy (CBT), mindfulness, positive psychology, and multitheoretical interventions [7]. These psychological approaches were found to have small to moderate effects on mental well-being in the general population, whereby mindfulness-based interventions (MBIs) and multicomponent positive psychology interventions were particularly efficacious [7,13]. Further meta-analyses focusing on positive psychology interventions, MBIs, and ACT-based interventions separately also found similar effects on mental well-being [14-16].

However, these systematic reviews did not focus on fully automated digital interventions. Fully automated digital interventions are interventions that are delivered entirely by the technology itself, not requiring any form of human support (by clinicians or nonclinicians) [17]. Although fully automated digital interventions might be less effective, as recent research has found that any form of human support enhances the effectiveness of interventions [18], fully automated digital interventions allow for great scalability and are highly cost-effective and accessible [19]. Therefore, fully automated digital interventions provide a particularly pertinent way to promote mental well-being in the general population.

Overall, there is a need to systematically review the evidence of the effectiveness of fully automated digital mental well-being interventions to improve mental well-being (which includes subjective and psychological well-being) in the general population. Furthermore, an understanding of what psychological approaches work when delivered fully automated digitally and for whom (as one approach does not suit all) [20] is needed.

Main Objective

This systematic review aims to understand the effectiveness of fully automated digital interventions in promoting mental well-being in the general population.

Secondary Objectives

Furthermore, the systematic review aims to explore the effectiveness of fully automated digital mental well-being interventions across psychological approaches and population subgroups.

Methods

Study Protocol

The systematic review protocol was registered on PROSPERO (CRD42022310702). The Cochrane handbook was used when designing and conducting the systematic review [21], and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed for reporting of the systematic review [22].

Eligibility Criteria

Studies were included if they used a fully automated digital intervention that aimed to promote mental well-being in the general population.

The study needed to include adults, meaning that the population needed to be aged ≥18 years. General population was further defined as any adult population subgroup that was not a clinical population and was not specifically recruited by the researchers because of (expected) lower mental well-being baseline scores.

Digital interventions were defined according to the National Institute for Health and Care Excellence [17] as interventions that are delivered through hardware and electronic devices (eg,

smartwatches and smartphones), software (eg, computer programs and apps), and websites. The intervention needed to be fully automated, which means it should be delivered by the technology itself entirely, independent from health care professionals, and not containing any other form of social support [17]. For example, a digital web-based intervention in which video content was delivered automatically would have been included, whereas a digital video call intervention in which a health care professional delivered content would have been excluded. Although the content should be delivered entirely by the technology itself, the elements of the study could still have been conducted by the researchers. For example, researchers could have screened, obtained measures, and obtained informed consent (digitally or in person), after which they could have provided the participants with access to the intervention.

Furthermore, the intervention needed to use individual mental well-being promotion, defined by the WHO as "various activities or practices that aim to promote, build on, increase or foster primarily individuals' strengths, resourcefulness or resiliency" [10]. This should be a psychological intervention.

Interventions that included physical activity–related or lifestyle-related interventions were excluded. If an intervention contained elements that did not include mental well-being promotion, they would also be excluded, as the detection of the effectiveness of mental well-being promotion strategies would not be possible. For example, an MBI would have been included; however, an MBI that included a yoga session would have been excluded.

The outcome needed to consider a validated measure of mental well-being, including psychological well-being or subjective well-being.

Finally, studies needed to investigate the effectiveness of this digital intervention on mental well-being. Therefore, quantitative randomized and nonrandomized studies of interventions, such as before-after studies, were considered appropriate, as they can provide insights into the effectiveness of interventions [23]. For further details regarding the inclusion and exclusion criteria, please refer to the protocol [24].

Searches

The initial search was conducted in February 2022 and updated using a title and keyword search in October 2022. The databases searched included MEDLINE, Web of Science, Cochrane, PsycINFO, PsycEXTRA, Scopus, and ACM Digital. Combinations of the following key search terms were used: "mental well being," "mental wellbeing," "psychological well being," "psychological wellbeing," "subjective well being," and "subjective wellbeing," in combination with "digital*," "online," "internet," "web-based," "app," "apps," "smartphone application*," and "mobile application*." No restrictions were applied. Refer to Multimedia Appendix 1 [25-42] for the detailed searches conducted in each database.

Study Selection

Each record was double screened, and the reviewers were blinded to each other's decisions throughout the process. To ensure consistency and quality of the screening process, the lead author (JG) screened all records, and double screening was conducted by MB, ET, and MZ. After screening 10.71% (776/7764) of the records, interreviewer reliability was calculated, which ranged from moderate to substantial agreement (Cohen κ=0.54-0.79) [43]. Inconsistencies in the screening process were discussed, and conflicts were resolved through discussion. If conflicts remained, an additional discussion with a third, senior reviewer (BA) was conducted. Upon completion of the screening, interreviewer reliability was recalculated (Cohen κ=0.42-0.80), and conflicts were again resolved using the same process. This process was then repeated for full-text screening.

Data Extraction

Before data extraction, the Cochrane data collection form was adapted and prepiloted for this review. Data extraction included information regarding the study population, participant demographics, and setting; details of the intervention and control conditions (such as duration, frequency, timing, and activities); study methodology; recruitment and study completion rates; outcomes, outcome measures, and times of measurement; and information for the assessment of the risk of bias (RoB). Two reviewers (JG and AM) independently extracted all relevant data from the included studies and held meetings to discuss any discrepancies in data extraction. When conflicting views on the data extraction occurred, a third, senior reviewer (BA) advised on how to resolve the issue. Missing data were sought by contacting the lead author of the study via email, which was identified through the journal paper.

RoB Assessment

RoB was assessed independently by 2 reviewers (JG and AM) using the Cochrane RoB 2.0 tool for randomized controlled trials (RCTs) [44]. No standardized tools were available for noncontrolled before-after studies; therefore, the National Institutes of Health tool, "Quality Assessment Tool for Before-After (Pre-Post) Studies with No Control Group," was used as a guidance to provide an indication of the RoB in these studies [45]. However, it was considered that these studies would provide a lower quality of evidence. Following the RoB assessments, discussions were held to discuss conflicts, and any remaining disagreements were resolved through verbal discussion with a third reviewer (BA).

Data Synthesis and Meta-Analysis

Mean, SD, and total number of participants were extracted for each postintervention mental well-being outcome in the study arms that met the inclusion criteria of the digital mental well-being intervention and control group. The effect estimates were averaged, where the studies included multiple study outcomes. This method was also adopted for multiarm studies because it was considered meaningful to combine the intervention effects, as all the included intervention arms were digital mental well-being interventions. In addition, this avoided double counting of participants in the control group. Standardized mean differences (SMDs) were used in a random-effects model.

Initially, both the per-protocol (PP) and intention to treat (ITT) data were extracted. However, only PP data were included in

the meta-analysis, as high dropout rates (ranging up to 85%) led to ITT data being less meaningful.

Visual inspection of the forest plot and the chi-square and I^2 tests were used to assess heterogeneity. A value of >50% was considered to represent substantial heterogeneity. Heterogeneity was explored, interpreted, and contextualized.

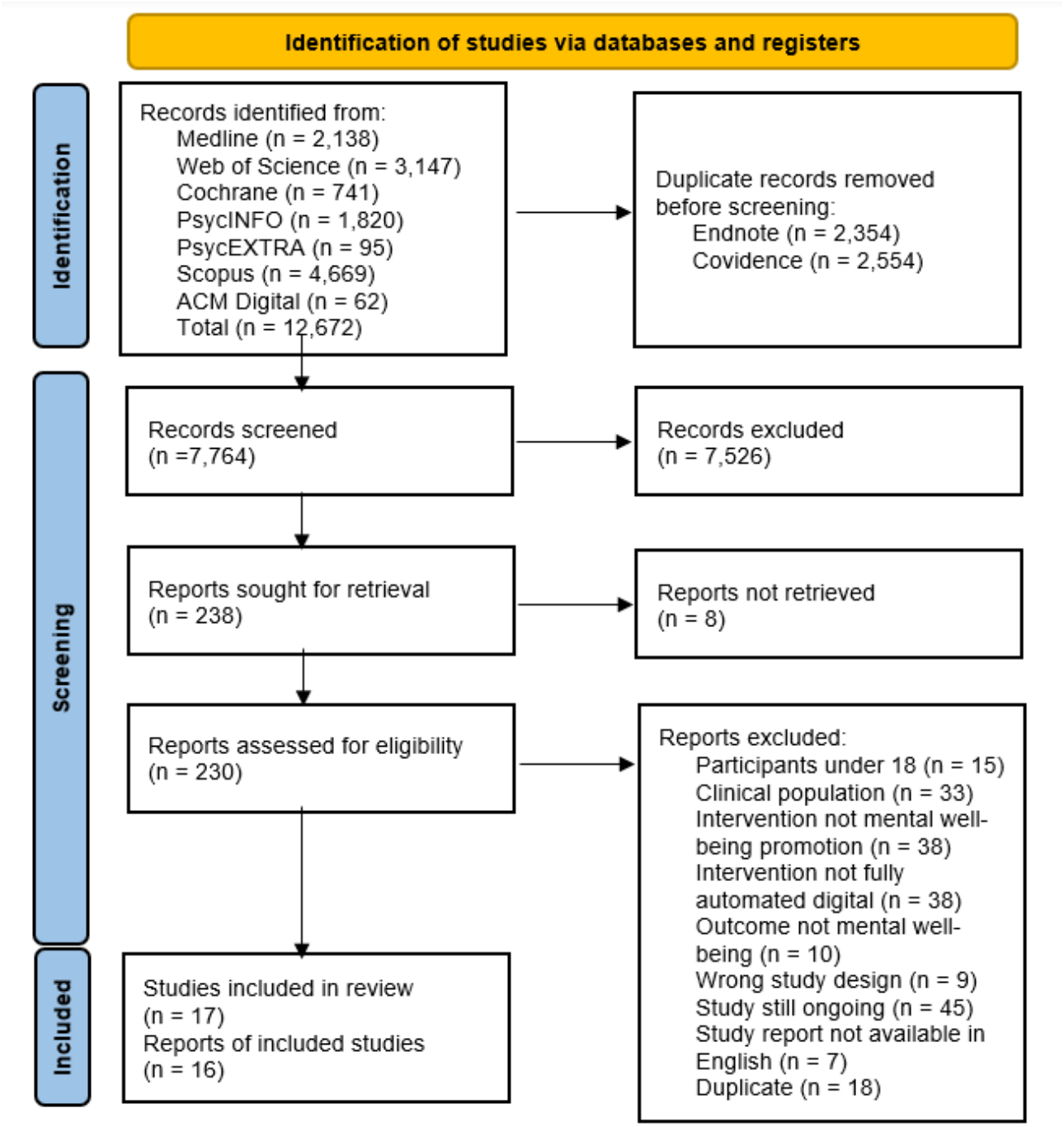
Results

Description of Studies

An initial search yielded 12,672 records. Following deduplication, 7764 records were screened in Covidence (Veritas Health Innovation). A total of 7526 records were excluded following title and abstract screening, and 238 records were sought for retrieval for full-text screening. A total of 230 full-text records were screened, leading to the exclusion of another 213 records. The most common reasons for exclusion were the population being a clinical population, intervention not solely using mental well-being promotion, intervention not being fully automated and digital, or that the study was still ongoing. For full details of the study selection process, refer to Figure 1.

An updated title and keyword search in October 2022 yielded another 525 records. After deduplication, 366 articles were screened in Covidence. A total of 347 articles were excluded, and full texts of 19 articles were obtained. Furthermore, 17 articles were excluded following full-text screening.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the search strategy outcomes.



Narrative Summary

A total of 18 records containing 19 studies were included in this systematic review, including 17 RCTs and 2 non-RCTs before-after trials.

Setting and Participants

Studies mainly occurred in Western countries; the participants were primarily female and highly educated; and the study populations were students, employees, mothers, and other general population samples (Table 1).

Table 1. Characteristics of the included studies.

Study, year	Population	Setting	Comparator	Outcome[a]
Study 3 from Avey et al [25], 2022	Employees	United States and Australia	Unknown	PWB[b]
Bakker et al [26], 2018	General population	Australia	Waitlist control	MWB[c]
Brazier et al [27], 2022	Trainees	United Kingdom	Waitlist control	MWB
Champion et al [28], 2018	Employees	United Kingdom and United States	Waitlist control	SWB[d]
Chung et al [29], 2021	Students	Australia and United Kingdom	Waitlist control	MWB
Study 1 from Di Consiglio et al [30], 2021	Students	Italy	Active control	PWB
Study 2 from Di Consiglio et al [30], 2021	Students	Italy	None	PWB
Eisenstadt et al [31], 2021	Real-world app users	United Kingdom	None	MWB
Gammer et al [32], 2020	Mothers of infants aged <1 y	United Kingdom	Waitlist control	MWB
Liu et al [33], 2021	Students	China	Placebo	SWB
Ly et al [34], 2017	General population	Sweden	Waitlist control	PWB and SWB
Mak et al [35], 2018	General population	China	Active control	MWB
Manthey et al [36], 2016	General population	Germany	Active control	SWB
Mitchell et al [37], 2009	Adults	Australia	Placebo	PWB
Neumeier et al [38], 2017	Employees	Germany and Australia	Waitlist control	SWB
Pheh et al [39], 2020	General population	Malaysia	Active control	MWB
Schulte-Frankenfeld and Trautwein [40], 2021	Students with a part-time job	Germany	Waitlist control	SWB
Shin et al [41], 2020	Students	United States	Placebo	SWB
Walsh et al [42], 2019	Students	Canada	Active control	PWB

[a]Mental well-being outcomes included 5-item mental well-being index (World Health Organization-5) [46] and Warwick-Edinburgh Mental Well-Being Scale (version 1) [47]. Subjective well-being outcomes included Satisfaction With Life Scale [48], Positive And Negative Affect Schedule [49], Satisfaction with Life and happiness [50], Subjective Happiness Scale [51], and single-item life satisfaction and affect measure [38]. Psychological well-being outcomes included psychological well-being [52], Psychological Well-Being Scale [53], Psychological Well-Being Index (adult) scale [54], and Flourishing Scale [34].

[b]PWB: psychological well-being.

[c]MWB: mental well-being.

[d]SWB: subjective well-being.

Psychological Approaches

Several different psychological approaches were used, including the following: (1) mindfulness, ACT, and self-compassion; (2) positive psychology; (3) cognitive behavioral; and (4) integrative (Table 2). The most frequently used psychological approach was mindfulness, ACT, and self-compassion. General intervention activities and behavior change techniques, such as well-being tips and behavior change techniques to form habits, were adopted across psychological approaches and in most interventions (Textbox 1).

The intervention content was primarily developed by the study researchers and clinical psychologists (15/19, 79% of studies), some studies collaborated with companies or digital laboratories to develop the intervention (2/19, 11%), and some studies tested a preexisting intervention developed by a company (2/19, 11%).

Table 2. Description of intervention characteristics[a].

Psychological approach underpinning the intervention	Activities or practices	Studies adopting the approach
Mindfulness, ACT[b], and self-compassion	• Meditation: awareness of inner experiences, present moment, and acceptance • Overcoming obstacles in mindfulness meditation • Body scan • Increasing awareness through biofeedback • Being mindful in daily life • Loving-kindness meditation • Compassionate journaling and breaks • Self-kindness activities	[28,29,32,35,39,40,42]
Positive psychology	• Gratitude (gratitude diary and letter) • Positive future imagination • Best possible self • Counting blessings • Random acts of kindness • Replaying positive experiences • Using strengths • Savoring the moment • Wearing a smile • Brainstorming meaningfulness	[33,36-38,41]
Cognitive behavioral approach	• Mood-related activities (eg, mood tracker, mood diary, and mood improvement activities) • Challenging thoughts and behaviors • Problem-solving • Goal setting (SMART[c] goals and planning) • Committed actions • Journaling	[26,37]
Integrative approach	• A combination of intervention activities or practices of these psychological approaches	[25,27,30,31,34]

[a]For more detailed intervention description, refer to Multimedia Appendix 2.

[b]ACT: acceptance and commitment therapy.

[c]SMART: Specific, Measurable, Achievable, Relevant, and Time-bound.

Textbox 1. General psychological intervention components.

General intervention components adopted across interventions

- Psychoeducation (eg, on emotions, needs, values, and mental illness)
- Support-seeking information
- Well-being tips

Behavior change techniques adopted across interventions [55]

- Habit formation
- Goal setting
- Action planning (eg, implementation intentions)
- Prompts or cues
- Self-monitoring of behavior or outcome of behavior
- Self-assessment of affective consequences
- Feedback on behavior
- Material or nonspecific reward

Intervention Delivery

A total of 24 fully automated digital mental well-being interventions were included. The interventions were app based (n=10), web based (n=11), both app and web based (n=2), and SMS text message (n=1) interventions (Table 3).

Table 3. Intervention characteristics and dropout[a].

Study, year	Participants randomized, N[b]	Intervention	Duration	Frequency	Mode of delivery	Dropout, n (%)[c]
Study 3 from Avey et al [25], 2022	102	Resilience intervention	10 wk	Weekly	App based	3 (2.9)
Bakker et al [26], 2018	226	Moodkit, Moodprism	30 d	Daily	App based	108 (47.8)
Brazier et al [27], 2022	279	Dear Doctor	10 mo	Fortnightly	SMS text message	126 (45.2)
Champion et al [28], 2018	74	Headspace	30 d	Daily	App based	12 (16.2)
Chung et al [29], 2021	427	Brief MBI[d]	6 wk	Weekly	Web based	280 (65.6)
Study 1 from Di Consiglio et al [30], 2021	24	Noibene	3 mo	4 times	Web based	0 (0)
Study 2 from Di Consiglio et al, [30], 2021	178	Noibene	None	None	Web based	119 (66.9)
Eisenstadt et al [31], 2021	115	Paradym	2 wk	Daily	App based	81 (70.4)
Gammer et al [32], 2020	206	Kindness For Mums Online	5 wk	Weekly	Web based	80 (38.8)
Liu et al [33], 2021	1000	Positive psychology intervention	1-3 d	Twice	Web based	132 (13.2)
Ly et al [34], 2017	30	Shim	2 wk	Daily	App based	3 (10)
Mak et al [35], 2018	2282	Mindfulness-based program and self-compassion program	28 d	Daily	App based and web based	1933 (84.7)
Manthey et al [36], 2016	666	Best possible self and gratitude	8 wk	Weekly	Web-based video	112 (16.8)
Mitchell et al [37], 2009	160	Strengths intervention and problem-solving intervention	3 wk	Daily	Web based	111 (77.6)
Neumeier et al [38], 2017	431	PERMA[e] program and gratitude program	7 d	Daily	App based	128 (29.7)
Pheh et al [39], 2020	206	Brief MBI	1 d	Once	Web based	100 (48.5)
Schulte-Frankenfeld and Trautwein [40], 2021	99	Balloon	8 wk	Daily	App based	35 (35.4)
Shin et al [41], 2020	630	Gratitude writing	20 min	Once	Web based	49 (7.8)
Walsh et al [42], 2019	108	Wildflowers	3 wk	Daily	App based	22 (20.4)

[a]This table represents the general characteristics of the studies included in this systematic review. Only interventions of the studies that met the inclusion criteria are presented in this table.

[b]N denotes the number of participants randomized in the study, irrespective of whether people conducted baseline and follow-up assessments.

[c]Dropout rates are calculated from randomization to final assessment.

[d]MBI: mindfulness-based intervention.

[e]PERMA: Positive emotion, Engagement, Relationships, Meaning, Accomplishment.

Intervention Duration, Frequency, and Timing

The participants were expected to use the intervention for substantially varied duration across interventions, ranging from 1 single session to 10 months, and there did not appear to be a clear end strategy across interventions. Most commonly, intervention use was recommended daily for up to 30 days, weekly for up to 8 weeks, and fortnightly for up to 10 months. Participants were often encouraged to use and access the intervention content for 5 to 15 minutes at a time, irrespective of the duration of the intervention.

Level of Automation of Interventions

Access was generally automated with instant, sequential, or weekly access to content (Table 4). Most digital content was delivered in a standard way, and tailoring and dynamic delivery of content occurred in only 2 mental well-being interventions [34,42].

Table 4. Level of automation and engagement of intervention.

Study, year	Intervention	Frequency of content release	How access to intervention content was provided	Tailoring of content to improve or maintain engagement	Other digital intervention strategies to improve or maintain engagement	Actual engagement with intervention content[a] (%)
Study 3 from Avey et al [25], 2022	Resilience intervention	Unknown	Unknown	None	None	Unknown
Bakker et al [26], 2018	Moodkit	Instant access	N/A[b]	N/A	None	Unknown
Bakker et al [26], 2018	Moodprism	Instant access	N/A	Feedback on mental well-being	None	Unknown
Brazier et al [27], 2022	Dear Doctor	Fortnightly	Automated text message	None	None	Unknown
Champion et al [28], 2018	Headspace	Sequential access	Automated access upon completion of step in the app	None	None	20.7
Chung et al [29], 2021	Brief MBI[c]	Fortnightly or weekly	Unknown	None	Notifying of new content	Unknown
Study 1 from Di Consiglio et al [30], 2021	Noibene	Instant access	N/A	None	None	100
Study 2 from Di Consiglio et al [30], 2021	Noibene	Instant access	N/A	None	None	Unknown
Eisenstadt et al [31], 2021	Paradym	Unknown	Unknown	None	Push notification	32.1
Gammer et al [32], 2020	Kindness for Mums Online	Weekly	Unknown	None	None	Unknown
Liu et al [33], 2021	Positive psychology intervention	Sequential access	Unknown	None	None	Unknown
Ly et al [34], 2017	Shim	Upon opening of app	Automated by digital conversational agent	On the basis of individual and external factors (eg, time of day)	None	126.5
Mak et al [35], 2018	Mindfulness-based Program	Weekly	Unknown	None	Sticker earning and alarm feature	29.5
Mak et al [35], 2018	Compassion-based program	Weekly	Unknown	None	Sticker earning and alarm feature	32.2
Manthey et al [36], 2016	Best possible self	Weekly	Automated email	None	None	Unknown
Manthey et al [36], 2016	Gratitude	Weekly	Automated email	None	None	Unknown
Mitchell et al [37], 2009	Strengths intervention	Instant access	N/A	None	Interactive features and automated email reminders	Unknown
Mitchell et al [37], 2009	Problem-solving intervention	Instant access	N/A	None	Interactive features and automated email reminders	Unknown
Neumeier et al [38], 2017	PERMA[d] program	Sequential access	Automated access upon completion of step in program	None	None	Unknown
Neumeier et al [38], 2017	Gratitude program	Sequential access	Automated access upon completion of step in program	None	None	Unknown
Pheh et al [39], 2020	Brief MBI	Instant access	N/A	None	None	Unknown
Schulte-Frankenfeld and Trautwein [40], 2021	Balloon	Sequential access	Automated access upon completion of step in the app	None	A reminder was sent if a session was missed.	40.2

Study, year	Intervention	Frequency of content release	How access to intervention content was provided	Tailoring of content to improve or maintain engagement	Other digital intervention strategies to improve or maintain engagement	Actual engagement with intervention content[a] (%)
Shin et al [41], 2020	Gratitude writing	Instant access	N/A	None	None	100
Walsh et al [42], 2019	Wildflowers	Sequential access	Automated access upon completion of step in the app	On the basis of mood and stress levels recommendations were made for meditations	None	77.7

[a]Actual engagement with content is based on the requested frequency of engagement with the intervention (eg, daily for 2 wk=14 d=100%) compared with the actual frequency of engagement in the intervention (eg, on average, participants engaged with the intervention on 5 d=35.7%).

[b]N/A: not applicable.

[c]MBI: mindfulness-based intervention.

[d]PERMA: Positive emotion, Engagement, Relationships, Meaning, Accomplishment.

Intervention Engagement

Overall, intervention engagement was suboptimal, below the required or recommended intervention engagement levels (Table 4). On average, participants engaged in 40.2% (median) of the recommended intervention sessions or days. Only few studies (3/19, 16%) contained optimal levels of engagement, engaging in the recommended intervention sessions or days or more [30,34,41].

Studies attempted to improve intervention engagement in a variety of different ways (Tables 2 and 4), including (1) sending automated email reminders or notifications to use the intervention, (2) increasing participant motivation (eg, increasing awareness of potential benefits and using in-app reward earning features), (3) increasing habit formation, and (4) tailoring intervention content based on external factors (such as time of day) or internal factors (such as suggestion of a specific activity based on someone's mood).

Although caution should be used when interpreting the impact of these strategies on the engagement with the intervention because of the variety and inconsistency in reporting, preliminary results imply that tailored content improves engagement more than interventions that use reminders (habit formation and prompts) or sticker earning features (nonspecific rewards). Furthermore, it seems that interventions that require little engagement—engaging once or 4 times in the intervention in total [30,41]—also allow for more optimal intervention engagement. This is in line with studies showing that engagement was generally highest at the start of the intervention and decreased with time.

Study Dropout and Attrition

Dropout occurred at any point throughout the study period when a participant failed to complete the research protocol associated with the digital intervention [56].

On average, there was a 37% dropout rate (mean), which ranged from 0% to 85% in the studies (Table 3). Strategies used to reduce study dropout included monetary incentives, the intervention being a mandatory element of university courses, and follow-up of participants by sending email reminders.

There were a range of findings across studies on the association between participants' demographic characteristics and dropout. One study found that male participants were more likely to drop out [36], whereas others (2/19, 11%) found no difference [27,31]. Some studies (2/19, 11%) found that participants who remained in the study were older [35,38], although other studies (2/19, 11%) did not find this effect [31,36]. One study found that educational level was higher among participants who dropped out [35], whereas another study did not find this effect [38].

Several studies have compared whether baseline mental well-being was associated with dropout. Most studies (5/19, 26%) did not find any differences in baseline mental well-being levels between participants who did and did not drop out [27,29,32,35,36]. However, 1 study found that participants with lower mental well-being and higher levels of anxiety, depression, and distress were more likely to drop out [30], whereas another study found that participants with higher mental well-being and lower levels of anxiety, depression, and distress were more likely to drop out [31].

Few studies (2/19, 11%) excluded participants from their analysis (considered them to have dropped out) if they did not adhere with the intervention content at a minimum required level [37,42]; most studies (17/19, 89%) included participants with any level of intervention engagement.

Outcomes

A variety of validated standardized questionnaires were used to measure mental well-being across studies, including the WHO 5 item mental well-being index and Warwick-Edinburgh Mental Well-Being Scale for mental well-being, Psychological Well-Being scale and Flourishing Scale for psychological well-being, and Satisfaction with Life Scale and Positive And Negative Affect Schedule for subjective well-being (Table 1). Nevertheless, the authors of 1 study created and validated their own mental well-being questionnaires, which included a combination of different measures. Although not included in this systematic review (as it is not considered the primary aim of mental well-being promotion), most studies (17/19, 89%) included additional outcome measures such as distress, depression, anxiety, and stress.

RoB Assessments

Generally, the RoB of the included studies was considered to be high (Table 5). High levels of dropout and nonadherence led to a high RoB in domain 2 of Cochrane's RoB-2.0 tool. This domain assesses RoB because of deviations from the intended interventions (effect of adhering to the intervention) and leads to high RoB, as the included studies did not appropriately account for intervention nonadherence in their analysis. For example, the Cochrane RoB-2.0 tool recommends using an instrumental variable analysis or inverse probability weighting to appropriately account for nonadherence; however, none of the included studies conducted these analyses.

Table 5. Bias assessment using Cochrane's risk of bias (RoB) 2.0 tool[a].

Study, year	Randomization process	Deviation from intended intervention	Missing outcome data	Measurement of outcome	Selection of the reported results	Overall RoB[b]
Study 3 from Avey et al [25], 2022	Some concerns[c]	High[d]	Low[e]	Some concerns	High	High
Bakker et al [26], 2018	Some concerns	High	Low	Some concerns	Some concerns	High
Brazier et al [27], 2022	Low	High	Low	Some concerns	Some concerns	High
Champion et al [28], 2018	Some concerns	High	Low	Some concerns	Low	High
Chung et al [29], 2021	High	High	High	High	Some concerns	High
Study 1 from Di Consiglio et al [30], 2021	Some concerns	High	Low	Some concerns	Some concerns	High
Gammer et al [32], 2020	Low	High	Low	Some concerns	Low	High
Liu et al [33], 2021	Some concerns	High	High	High	High	High
Ly et al [34], 2017	Low	High	Low	Some concerns	Some concerns	High
Mak et al [35], 2018	Low	High	Low	Some concerns	Some concerns	High
Manthey et al [36], 2016	Low	High	Some concerns	Low	Some concerns	High
Mitchell et al [37], 2009	Low	High	High	Low	Some concerns	High
Neumeier et al [38], 2017	Some concerns	High	High	High	Some concerns	High
Pheh et al [39], 2020	Some concerns	High	Some concerns	High	Some concerns	High
Schulte-Frankenfeld and Trautwein [40], 2021	Low	High	High	Some concerns	Some concerns	High
Shin et al [41], 2020	Low	Low	Low	High	Some concerns	High
Walsh et al [42], 2019	Low	High	Some concerns	High	Some concerns	High

[a]The National Institutes of Health bias assessment tool: before-after studies with no control group was used for study 2 from Di Consiglio et al [30], 2021 (overall RoB: high) and Eisenstadt et al [31], 2021 (overall RoB: high).

[b]The overall RoB judgement for that specific study.

[c]Some concerns: indicates that the authors considered there to be some concerns with the RoB for that study on that specific domain of the Cochrane RoB-2.0 tool.

[d]High: indicates that the authors considered there to be a high RoB for that study on that specific domain of the Cochrane RoB-2.0 tool.

[e]Low: indicates that the authors considered there to be a low RoB for that study on that specific domain of the Cochrane RoB-2.0 tool.

Furthermore, domain 4 in the RoB-2.0 tool, assessing RoB in measuring the outcome, led to a high RoB because of the nature of the research being fully automated and digital. Self-report measures were used to digitally assess mental well-being; however, participants were aware of the intervention they received when self-reporting their mental well-being scores, as most studies (11/19, 58%) included a waitlist control group. Although active controls account for this issue, these control interventions also contained high levels of dropout and therefore might not be appropriate as a control group [35].

A high RoB was also detected in studies because of the lack of general high-quality research practice. For example, several studies (7/19, 37%) did not provide any information regarding the randomization process, most studies did not preregister (12/19, 63%), and studies that did preregister (2/19, 11%) sometimes did not indicate their preintended analysis plan.

Intervention Effects

All studies included fully automated digital mental well-being interventions in the general population and were therefore considered sufficiently homogeneous for a meta-analysis. Methodological homogeneity was also considered, which led to a comparison across RCTs only, as these were considered sufficiently homogeneous for a meta-analysis. Considering the incredibly high range of missing values, a meta-analysis based on ITT data was considered inappropriate; therefore, we conducted a meta-analysis based on PP data instead. Nevertheless, this increases the risk of underestimating or overestimating the real effect, which should be considered when

interpreting the meta-result. Full PP data were available for a subset of 12 studies. A random-effect model was applied, as different measures were used to measure the same multidimensional construct *mental well-being*. Average effect estimates were computed for each study, with *negative affect* scores reversed to ensure that a higher score in each study indicated elevated levels of mental well-being. SMDs, 95% CIs, and 2-sided *P* values were calculated.

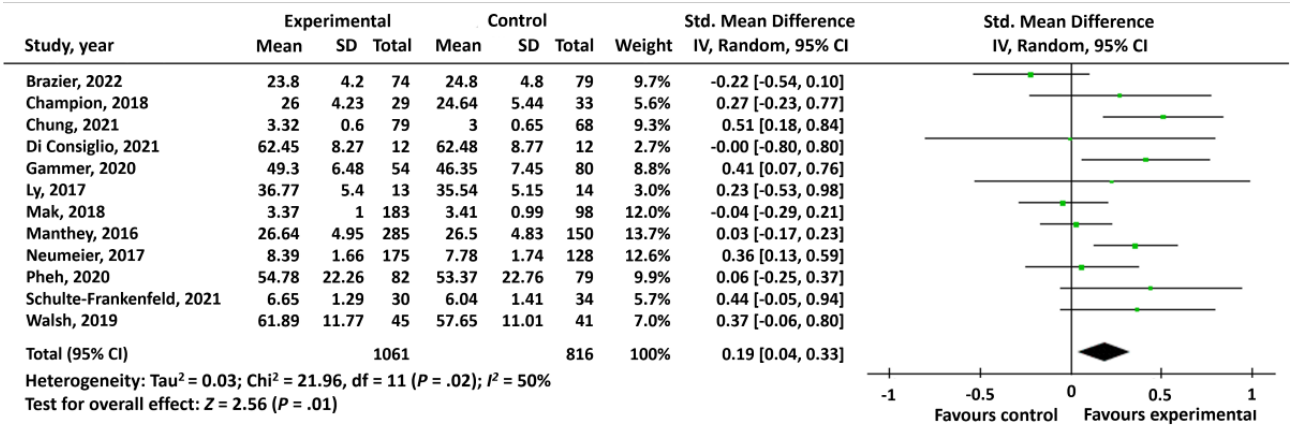
Outlier

During data extraction, the negative affect score in the intervention group of 1 study [33] was flagged by both reviewers as unexpectedly high, and further information was sought to identify what could potentially explain this unusually large result. Normative data for negative affect was mean 14.8 (SD 5.4) [57]; however, the negative affect score in the waitlist control group in this study was mean 26.98 (SD 5.19). When exploring this data further, no methodological or clinical differences could reliably explain this result in our opinion. In addition, when included in the meta-analysis, CIs were entirely outside the range of any other study, and heterogeneity was incredibly high (92%; Multimedia Appendix 3). Removing this study from the meta-analysis reduced the overall heterogeneity from 92% to 50%. Therefore, the study was considered an outlier and was excluded from the meta-analysis.

Main Effect

The pooled SMD, for the 12 trials, calculated using a random-effects model was 0.19 (95% CI 0.04-0.33; *P*=.01), indicating a small clinical effect in favor of digital mental well-being interventions (Figure 2). There was substantial heterogeneity (I^2=50%).

Figure 2. Per-protocol meta-analysis of fully automated digital interventions compared with control groups on mental well-being in the general population [27-30,32,34-36,38-40,42].



Sensitivity Analyses

As there was substantial heterogeneity (I^2=50%), sensitivity analyses were performed to explore, interpret, and contextualize heterogeneity. First, intervention duration was explored using subgroups of interventions lasting up to 2 weeks (short), 2 to 6 weeks (medium), and >6 weeks (long).

A small significant effect was found for short interventions (SMD 0.24, 95% CI 0.04-0.45; *P*=.02) and medium interventions (SMD 0.29, 95% CI 0.05-0.52; *P*=.02); however, no effect was found for long interventions (SMD 0.02, 95% CI −0.22 to 0.26; Figure S1 in Multimedia Appendix 4). No significant levels of heterogeneity were found in any of the subgroups (all *P*>.05), and the subgroups substantially reduced the overall level of heterogeneity (I^2=28.6%).

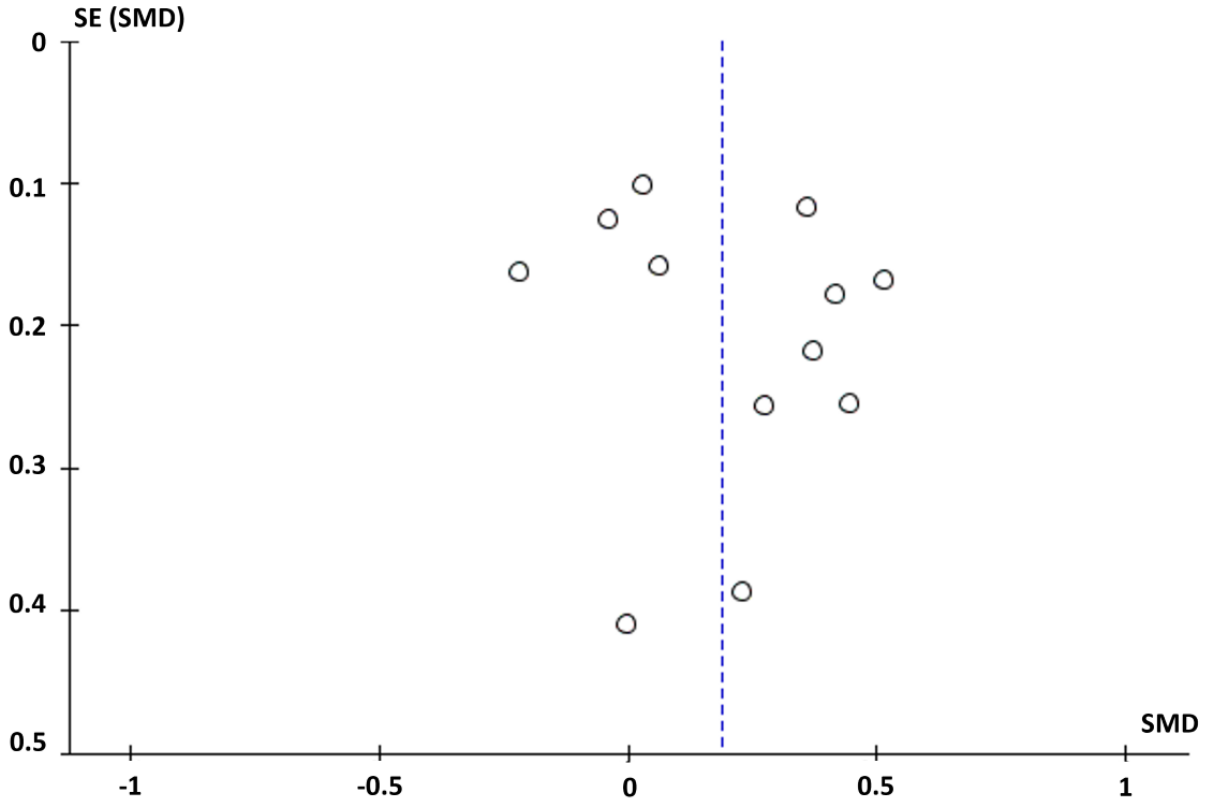
Another sensitivity analysis was performed to explore methodological heterogeneity across studies based on the comparator. We argue that placebo controls are not feasible in psychological interventions, considering the difficulty in isolating intervention components in psychological interventions [58]. Therefore, we grouped placebo controls under active controls in this review. A small significant effect was found in studies using a waitlist control as a comparator (SMD 0.28, 95% CI 0.07-0.50; *P*=.008), but no significant effect was found in studies using a placebo or active control as a comparator (SMD 0.05, 95% CI −0.08 to 0.18; *P*=.49; Figure S2 in Multimedia Appendix 4). No significant levels of heterogeneity were present in either of the 2 subgroups (all *P*>.05), although substantial heterogeneity remained in studies that used a waitlist control comparator (I^2=53%).

Finally, a sensitivity analysis was performed based on the outcomes of mental well-being, psychological well-being, and subjective well-being. A small significant effect was found on subjective well-being (SMD 0.23, 95% CI 0.04-0.42; *P*=.02). However, no significant effect was found on mental well-being (SMD 0.14, 95% CI −0.12 to 0.40; *P*=.31) or psychological well-being (SMD 0.26, 95% CI −0.08 to 0.59; *P*=.14; Figure S3 in Multimedia Appendix 4). Despite reducing heterogeneity in subjective well-being and psychological well-being, substantial heterogeneity was found in mental well-being (I^2=72%).

Reporting Bias

Visual inspection of the funnel plot, which appeared asymmetrical, indicated evidence of reporting bias (Figure 3). Few smaller studies were found, and larger random variation would be expected within smaller studies; this is potentially because of a publication bias, although other aspects such as heterogeneity can also cause asymmetrical funnel plots.

Figure 3. Funnel plot. Asymmetrical plot due to the presence of publication bias or low methodological quality studies. The funnel plot only represents studies that were included in the main per-protocol meta-analysis. SMD: standardized mean difference.



Certainty of Body of Evidence (Grading of Recommendations, Assessment, Development, and Evaluations)

The certainty of the body of evidence was assessed using Grading of Recommendations, Assessment, Development, and Evaluations [59]. The evidence was downgraded because of high RoB (effect of adhering to the intervention; Table 5), inconsistency (heterogeneity was considered substantial; Figure 2), imprecision (wide CIs and insufficiently small sample sizes were observed; Figure 2), and publication bias (visual asymmetry in the funnel plot; Figure 3). Thus, we consider a very low confidence in the quality of evidence of the main PP meta-effect (Figure 2), meaning that we are very uncertain about the estimate of the effect.

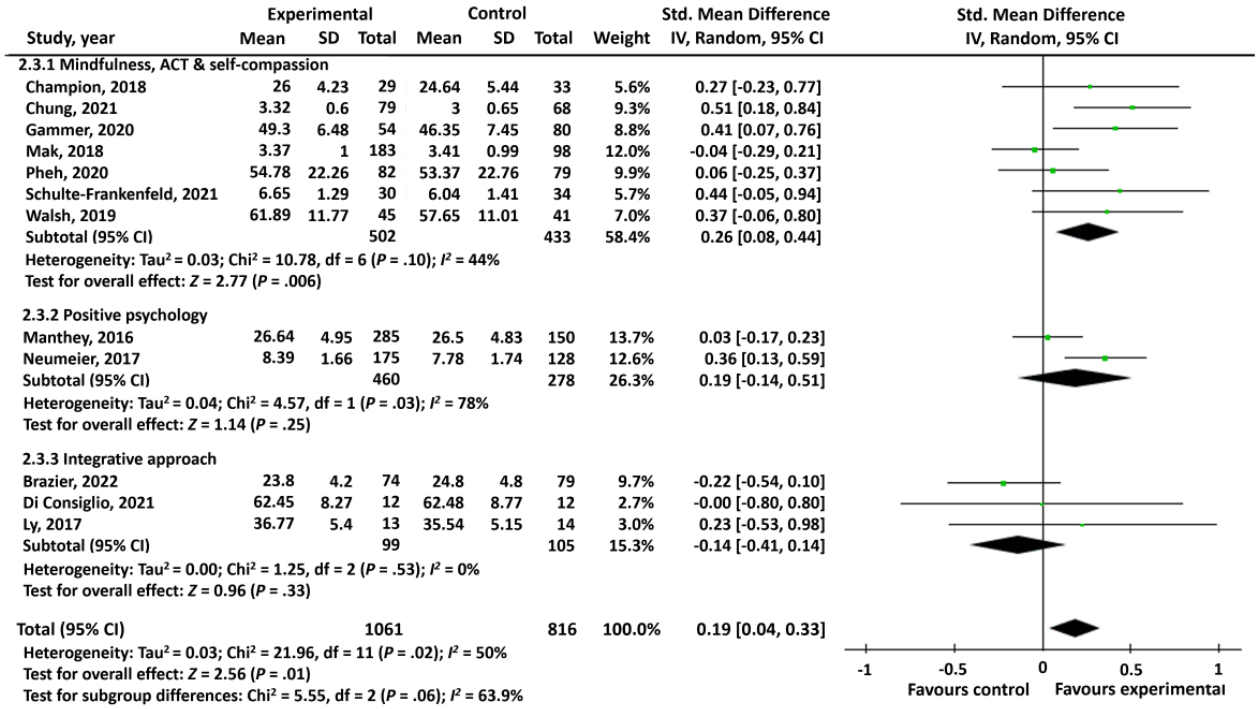
Subgroup Analysis

An a priori subgroup analysis was planned to detect the effects of digital mental well-being interventions across individual differences (eg, age, sex, and educational level). Nevertheless, insufficient data were available for a meaningful comparison to be made.

Another a priori subgroup analysis was planned to identify the effectiveness across psychological approaches. Mindfulness, ACT, and self-compassion interventions were the most common. A total of 7 studies were included in this subgroup. A small significant effect was found for fully automated digital mindfulness, ACT, and self-compassion interventions to promote mental well-being in the general population (SMD 0.26, 95% CI 0.08-0.44; P=.006), with moderate levels of heterogeneity (I^2=44%; Figure 4). The positive psychology intervention subgroup only included 2 studies, and there were significant levels of heterogeneity (P=.03; I^2=78%). Studies investigating CBT-based interventions did not contain any PP data and could therefore not be included as a subgroup in the analysis. The final subgroup included an integrative approach; 3 studies contained sufficient PP data to be included. There was no significant level of heterogeneity in this subgroup (P=.53; I^2=0%); however, integrative approaches did not have a significant effect on mental well-being in the general population (P=.33).

Overall, no significant subgroup difference was found when comparing the effects of mindfulness, ACT, self-compassion, positive psychology, and integrative interventions on mental well-being (P=.06).

Figure 4. Subgroup analysis of different psychological approaches to promote mental well-being [27-30,32,34-36,38-40,42].



Discussion

Main Effect

The aim of this systematic review and meta-analysis was to understand the effectiveness of fully automated digital interventions in promoting mental well-being in the general population. We evaluated 24 fully automated digital mental well-being interventions lasting from a single session to 10 months, with daily, weekly, and biweekly delivery. After the intervention, we found a small significant effect of fully automated digital mental well-being interventions compared with control groups on mental well-being in the general population.

The effect found in this meta-analysis of fully automated digital interventions (SMD 0.19) was smaller than the effect found in previous meta-analyses of nonautomated mental well-being interventions (effect sizes ranging between 0.26 and 0.42) [7,15,16]. This could highlight the importance of nonspecific psychological factors, such as the therapeutic relationship and social support, in the effectiveness of these psychological interventions. In contrast, this could also indicate the importance of social support in the adherence to mental well-being interventions. Previous research found that improved adherence was linked to better mental well-being outcomes and that adherence tended to be higher in nonautomated interventions [18,56]. As suboptimal intervention adherence was observed in this review, with average engagement in 40% of the intervention content, it is likely that the reported effectiveness in this review is an underestimation of the potential effectiveness of fully automated digital interventions that could be achieved when reaching optimal levels of engagement (the level of engagement recommended by the researchers). Nevertheless, the recommended engagement levels differed tremendously between studies, and studies lacked a clear end strategy.

Exploratory Findings

We found that short (<2 wk) and medium (<6 wk) interventions were effective in promoting mental well-being in the general population but long (>6 wk) interventions were not. This could be further related to intervention adherence, as (in line with previous research findings) intervention adherence reduced with time [56]. It does appear that the optimal intervention duration may also depend on the outcome that is being targeted. Research has found that short interventions led to a greater effect on subjective well-being, whereas long interventions had a greater effect on psychological well-being [60]. As most studies (9/19, 47%) in this review included a subjective well-being outcome, this might explain why shorter interventions were found to be effective in this review.

In contrast to prior research, our exploratory analysis showed no significant effect on general mental well-being outcomes (eg, Warwick-Edinburgh Mental Well-Being Scale) [7,15]. Measures of general mental well-being might lack the sensitivity to detect subtle changes occurring within the general population. This could be attributed to the concise nature of mental well-being measures, which encompass both subjective and psychological aspects [47]. Previous research includes a clinical population alongside a general population and nonautomated interventions alongside fully automated digital mental well-being interventions [7,14,15]. Both these factors increase the effectiveness of mental well-being interventions, which could lead to a sufficiently large effect to detect using a general mental well-being measure.

Furthermore, we found a small significant effect when comparing a fully automated digital mental well-being intervention with a waitlist control group, although no significant

effect was found when comparing it with an active or placebo control group. The effect when compared with an active and placebo control is expected to be smaller than the effect when compared with a passive control [61]. This indicates that the effects of mental well-being interventions and other psychological interventions (eg, active control) on mental well-being do not currently differ.

Subgroup Effects

It was not possible to analyze the effects of digital mental well-being promotion across population subgroups (based on age, sex, socioeconomic status, and educational level) because of a lack of studies reporting these results separately.

Nevertheless, studies did provide exploratory findings on the relationship between individual differences and dropout in fully automated digital mental well-being interventions. These exploratory findings indicated largely conflicting evidence on whether and how individual differences were related to dropout, which is in line with previous research findings [56].

A subgroup analysis comparing psychological approaches adopted in fully automated digital mental well-being interventions indicated a small significant effect of fully automated digital mindfulness-, ACT-, and compassion-based interventions on mental well-being in the general population, with most studies (7/19, 37%) adopting this psychological approach. The effectiveness of fully automated digital positive psychology and CBT-based approaches remains largely unknown. A potential explanation for this is the large focus of CBT-based interventions on symptom reduction rather than on mental well-being improvement [62]. Furthermore, positive psychology interventions have been criticized recently because of the limited ability of studies to replicate positive psychology results [63], potentially leading to fewer studies investigating positive psychology interventions.

Finally, although several studies (6/19, 32%) have adopted an integrative approach, we did not find an effect of fully automated digital integrative approaches on mental well-being in the general population. This contradicts previous meta-analytic findings that found a significant effect of multitheoretical interventions on mental well-being in the general population [7]. Nevertheless, previous meta-analysis also found a smaller effect for multitheoretical interventions compared with MBIs [7], indicating that these interventions might generally be less effective. This might explain why no effect of integrative approaches was found in fully automated digital interventions.

Limitations

Several methodological limitations should be recognized; however, as they could have impacted the findings of this systematic review. First, the specific search terms adopted in this systematic review limit the findings. Although searches should aim to be as comprehensive as possible, it is necessary to balance sensitivity and specificity when conducting searches [64]. The specificity adopted in this systematic review may not have allowed the searches to be comprehensive, as the literature uses many different terms to describe fully automated digital mental well-being interventions. Second, the inclusion criteria in this systematic review are ambiguous and require judgement [64]. This subjectivity could lead to lower reproducibility of the findings and random errors and biases [65]. Finally, the review adopts an exclusive focus on mental well-being (which includes both subjective and psychological well-being). Although improving mental well-being could be considered the primary aim of digital mental well-being promotion [10], the exclusive focus on mental well-being does not allow the review to provide insights into indirect positive or negative intervention effects.

In addition to methodological limitations, we observed several limitations of the included studies that lowered confidence in the quality of evidence (Grading of Recommendations, Assessment, Development, and Evaluations). We saw a high RoB in the included studies because of the following reasons: (1) missing outcome data—although it is unknown what impact the dropout has on the overall effect (eg, underestimation or overestimation) as reasons for dropout remain largely unknown; (2) the effect of adherence—suboptimal adherence might lead to an underestimation of the effectiveness; and (3) measurement of the outcome—because of the use of self-report measures while participants are aware of their allocated intervention, potentially leading to overestimation of the effectiveness. In addition, we also found a lack of general high-quality research practice in studies. Several studies were underpowered, did not provide sufficient information regarding randomization, and did not preregister or contain a prespecified analysis plan.

Furthermore, we detected a publication bias of the studies included in the meta-analysis. This publication bias indicated that smaller studies with a larger random variation were largely missing, perhaps because they were less likely to be published.

Finally, the fully automated digital mental well-being interventions were primarily delivered in a Western context and typically included a sample of participants who were highly educated and female, which might limit the generalizability of the findings. In particular, there is evidence that females and highly educated individuals might engage with and therefore benefit from these interventions differently.

Recommendations for Future Research

The systematic review findings lead to several implications for future research. First, future research should aim to focus in more detail on supporting engagement and reducing dropout in fully automated digital mental well-being interventions—by understanding the impact of behavioral strategies, such as habit formation and nonspecific rewards [55], and also by examining what is considered *effective engagement*—the target level of intervention engagement needed for change [66]. This will allow for evidence-based recommendations of the level of intervention engagement in future research and practice and for studies to adopt effective end strategies.

Second, future research should look to understand how automated digital interventions can be *tailored* to deliver relevant content according to the preferences of the user and whether tailoring is necessary to ensure intervention effectiveness and whether acceptability can be ensured across different populations (eg, Western vs non-Western) and

intervention types (eg, positive psychology vs mindfulness and ACT).

Finally, we recommend that future research strictly follows high-quality research recommendations, such as the CONSORT (Consolidated Standards of Reporting Trials) statement [67], when investigating fully automated digital mental well-being interventions to allow for higher confidence in the quality of the evidence.

Conclusions

Overall, this review provides a novel insight into the effectiveness of fully automated digital mental well-being interventions in the general population. It shows that fully automated digital mental well-being interventions can effectively promote mental well-being in the general population (particularly when adopting a mindfulness-, ACT-, and self-compassion–based approach), despite low levels of intervention adherence and high study dropout.

Acknowledgments

The authors are grateful for the support of the librarians at the University of Bath and advice from Emma Fisher in conducting the meta-analysis. This work was supported by the Economic and Social Research Council (grant 2572559) and Cyberlimbic Systems Ltd.

Data Availability

For the purpose of open access, the author has applied a Creative Commons Attribution (CC-BY) license. Data supporting this study are openly available from UK Data Service.

Authors' Contributions

JG, BA, MB, CC, and TT designed the study, created the study protocol, and developed the study methodology. Then study selection took place which was conducted by JG, MB, ET, MZ, and BA. This was followed by data extraction and quality assessment which JG, AM and BA conducted. The data was then analyzed by JG, BA, MB, and CC. After which it was written-up and edited by JG, BA, MB, CC, AM, TT, MB, ET, and MZ.

Conflicts of Interest

JG received partial funding for this research project from Cyberlimbic Systems Ltd. TT is the chief executive officer and cofounder of Cyberlimbic Systems Ltd. BA received funding from the National Institute of Health Research and the UK Research and Innovation on the topic of digital health interventions. BA also sits on the scientific advisory board of the Medito Foundation and earGym. The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
Search strategy per database.
[DOCX File , 21 KB - mental_v10i1e44658_app1.docx]

Multimedia Appendix 2
Table with detailed intervention description.
[DOCX File , 28 KB - mental_v10i1e44658_app2.docx]

Multimedia Appendix 3
Main per-protocol analysis including outlier [27-30,32-36,38-40,42].
[DOCX File , 35 KB - mental_v10i1e44658_app3.docx]

Multimedia Appendix 4
Exploring heterogeneity [27-30,32,34-36,38-40,42].
[DOCX File , 114 KB - mental_v10i1e44658_app4.docx]

Multimedia Appendix 5
PRISMA Checklist.
[PDF File (Adobe PDF File), 69 KB - mental_v10i1e44658_app5.pdf]

References

1. Diener E. Subjective well-being. Psychol Bull 1984;95(3):542-575 [FREE Full text] [doi: 10.1037/0033-2909.95.3.542]

2. Ryff CD. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. J Pers Soc Psychol 1989;57(6):1069-1081 [FREE Full text] [doi: 10.1037/0022-3514.57.6.1069]
3. Ryan RM, Deci EL. On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annu Rev Psychol 2001;52:141-166 [FREE Full text] [doi: 10.1146/annurev.psych.52.1.141] [Medline: 11148302]
4. Westerhof GJ, Keyes CL. Mental illness and mental health: the two continua model across the lifespan. J Adult Dev 2010 Jun;17(2):110-119 [FREE Full text] [doi: 10.1007/s10804-009-9082-y] [Medline: 20502508]
5. Promotion of mental well-being: pursuit of happiness. World Health Organization. 2013 Nov 30. URL: https://www.who.int/publications/i/item/9789290224228 [accessed 2022-09-15]
6. Tudor K. Mental Health Promotion: Paradigms and Practice. Milton Park, UK: Routledge; 1996.
7. van Agteren J, Iasiello M, Lo L, Bartholomaeus J, Kopsaftis Z, Carey M, et al. A systematic review and meta-analysis of psychological interventions to improve mental wellbeing. Nat Hum Behav 2021 May;5(5):631-652 [FREE Full text] [doi: 10.1038/s41562-021-01093-w] [Medline: 33875837]
8. Bohlmeijer ET, Westerhof GJ. A new model for sustainable mental health. In: Making an Impact on Mental Health. Milton Park, UK: Routledge; 2020.
9. Keyes CL. Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. Am Psychol 2007;62(2):95-108 [FREE Full text] [doi: 10.1037/0003-066x.62.2.95]
10. Promoting mental health: concepts, emerging evidence, practice. World Health Organization. 2004. URL: https://apps.who.int/iris/bitstream/handle/10665/42940/9241591595.pdf [accessed 2022-09-15]
11. Bendtsen M, Müssener U, Linderoth C, Thomas K. A mobile health intervention for mental health promotion among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e17208 [FREE Full text] [doi: 10.2196/17208] [Medline: 32196462]
12. Comprehensive mental health action plan 2013-2030. World Health Organization. 2021 Sep 21. URL: https://www.who.int/publications/i/item/9789240031029 [accessed 2022-09-15]
13. Weiss LA, Westerhof GJ, Bohlmeijer ET. Can we increase psychological well-being? The effects of interventions on psychological well-being: a meta-analysis of randomized controlled trials. PLoS One 2016;11(6):e0158092 [FREE Full text] [doi: 10.1371/journal.pone.0158092] [Medline: 27328124]
14. Carr A, Cullen K, Keeney C, Canning C, Mooney O, Chinseallaigh E, et al. Effectiveness of positive psychology interventions: a systematic review and meta-analysis. J Posit Psychol 2020 Sep 10;16(6):749-769 [FREE Full text] [doi: 10.1080/17439760.2020.1818807]
15. Galante J, Friedrich C, Dawson AF, Modrego-Alarcón M, Gebbing P, Delgado-Suárez I, et al. Mindfulness-based programmes for mental health promotion in adults in nonclinical settings: a systematic review and meta-analysis of randomised controlled trials. PLoS Med 2021 Jan 11;18(1):e1003481 [FREE Full text] [doi: 10.1371/journal.pmed.1003481] [Medline: 33428616]
16. Howell AJ, Passmore HA. Acceptance and Commitment Training (ACT) as a positive psychological intervention: a systematic review and initial meta-analysis regarding ACT's role in well-being promotion among university students. J Happiness Stud 2018 Sep 5;20(6):1995-2010 [FREE Full text] [doi: 10.1007/s10902-018-0027-7]
17. Behaviour change: digital and mobile health interventions. National Institute for Health and Care Excellence. 2020 Oct 7. URL: https://www.nice.org.uk/guidance/ng183/resources/behaviour-change-digital-and-mobile-health-interventions-pdf-66142020002245#:~:text=This%20guideline%20covers%20interventions%20that,wearable%20devices%20or%20the%20internet [accessed 2022-09-15]
18. Leung C, Pei J, Hudec K, Shams F, Munthali R, Vigo D. The effects of nonclinician guidance on effectiveness and process outcomes in digital mental health interventions: systematic review and meta-analysis. J Med Internet Res 2022 Jun 15;24(6):e36004 [FREE Full text] [doi: 10.2196/36004] [Medline: 35511463]
19. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
20. Ainsworth B, Atkinson MJ, AlBedah E, Duncan S, Groot J, Jacobsen P, et al. Current tensions and challenges in mindfulness research and practice. J Contemp Psychother 2023 May 20:1-6 [FREE Full text] [doi: 10.1007/s10879-023-09584-9]
21. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.3. London, UK: The Cochrane Collaboration; 2019.
22. Page M, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 2021 Apr;88:105906 [FREE Full text] [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
23. Reeves BC, Shea BJ, Wells GA, Sharma Waddington H. Updated guidance from the agency for healthcare research and quality effective healthcare program on including nonrandomized studies of interventions in systematic reviews: a work in progress. J Clin Epidemiol 2022 Dec;152:309-310 [FREE Full text] [doi: 10.1016/j.jclinepi.2022.11.001] [Medline: 36371046]
24. Groot J, Ainsworth B, Brosnan M, Clarke C. The effectiveness of digital interventions to enhance mental well-being in the general population: a systematic review. PROSPERO. 2022. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022310702 [accessed 2023-09-11]

25. Avey J, Newman A, Herbert K. Fostering employees' resilience and psychological well-being through an app-based resilience intervention. Pers Rev (forthcoming) 2022 Aug 18 [FREE Full text] [doi: 10.1108/pr-08-2021-0612]
26. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83 [FREE Full text] [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
27. Brazier A, Larson E, Xu Y, Judah G, Egan M, Burd H, et al. 'Dear Doctor': a randomised controlled trial of a text message intervention to reduce burnout in trainee anaesthetists. Anaesthesia 2022 Apr 13;77(4):405-415 [FREE Full text] [doi: 10.1111/anae.15643] [Medline: 35026055]
28. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. PLoS One 2018;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]
29. Chung J, Mundy ME, Hunt I, Coxon A, Dyer KR, McKenzie S. An evaluation of an online brief mindfulness-based intervention in higher education: a pilot conducted at an Australian university and a British university. Front Psychol 2021 Oct 28;12:752060 [FREE Full text] [doi: 10.3389/fpsyg.2021.752060] [Medline: 34777149]
30. Di Consiglio M, Fabrizi G, Conversi D, La Torre G, Pascucci T, Lombardo C, et al. Effectiveness of NoiBene: a web-based programme to promote psychological well-being and prevent psychological distress in university students. Appl Psychol Health Well Being 2021 May 17;13(2):317-340 [FREE Full text] [doi: 10.1111/aphw.12251] [Medline: 33595198]
31. Eisenstadt A, Liverpool S, Metaxa AM, Ciuvat RM, Carlsson C. Acceptability, engagement, and exploratory outcomes of an emotional well-being app: mixed methods preliminary evaluation and descriptive analysis. JMIR Form Res 2021 Nov 01;5(11):e31064 [FREE Full text] [doi: 10.2196/31064] [Medline: 34569466]
32. Gammer I, Hartley-Jones C, Jones FW. A randomized controlled trial of an online, compassion-based intervention for maternal psychological well-being in the first year postpartum. Mindfulness 2020 Jan 17;11(4):928-939 [FREE Full text] [doi: 10.1007/s12671-020-01306-9]
33. Liu K, Duan Y, Wang Y. The effectiveness of a web-based positive psychology intervention in enhancing college students' mental well-being. Soc Behav Pers Int J 2021 Aug 04;49(8):1-13 [FREE Full text] [doi: 10.2224/sbp.10459]
34. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
35. Mak WW, Tong AC, Yip SY, Lui WW, Chio FH, Chan AT, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60 [FREE Full text] [doi: 10.2196/mental.8597] [Medline: 30309837]
36. Manthey L, Vehreschild V, Renner KH. Effectiveness of two cognitive interventions promoting happiness with video-based online instructions. J Happiness Stud 2016 Feb;17(1):319-339 [FREE Full text] [doi: 10.1007/s10902-014-9596-2]
37. Mitchell J, Stanimirovic R, Klein B, Vella-Brodrick D. A randomised controlled trial of a self-guided internet intervention promoting well-being. Comput Hum Behav 2009 May;25(3):749-760 [FREE Full text] [doi: 10.1016/j.chb.2009.02.003]
38. Neumeier LM, Brook L, Ditchburn G, Sckopke P. Delivering your daily dose of well-being to the workplace: a randomized controlled trial of an online well-being programme for employees. Eur J Work Organ Psychol 2017 May 04;26(4):555-573 [FREE Full text] [doi: 10.1080/1359432x.2017.1320281]
39. Pheh K, Tan H, Tan C. Ultra-brief online mindfulness-based intervention effects on mental health during the coronavirus disease outbreak in Malaysia: a randomized controlled trial. Makara Hum Behav Stud Asia 2020 Dec 31;24(2):118-128 [FREE Full text] [doi: 10.7454/hubs.asia.2140920]
40. Schulte-Frankenfeld PM, Trautwein FM. App-based mindfulness meditation reduces perceived stress and improves self-regulation in working university students: a randomised controlled trial. Appl Psychol Health Well Being 2022 Nov 27;14(4):1151-1171 [FREE Full text] [doi: 10.1111/aphw.12328] [Medline: 34962055]
41. Shin M, Wong YJ, Yancura L, Hsu K. Thanks, mom and dad! An experimental study of gratitude letter writing for Asian and white American emerging adults. Couns Psychol Q 2018 Nov 19;33(3):267-286 [FREE Full text] [doi: 10.1080/09515070.2018.1542519]
42. Walsh KM, Saab BJ, Farb NA. Effects of a mindfulness meditation app on subjective well-being: active randomized controlled trial and experience sampling study. JMIR Ment Health 2019 Jan 08;6(1):e10844 [FREE Full text] [doi: 10.2196/10844] [Medline: 30622094]
43. Watson PF, Petrie A. Method agreement analysis: a review of correct methodology. Theriogenology 2010 Jun;73(9):1167-1179 [FREE Full text] [doi: 10.1016/j.theriogenology.2010.01.003] [Medline: 20138353]
44. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
45. Study quality assessment tools. National Institutes of Health National Heart Lung and Blood Institute. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2022-09-15]
46. World Health Organization. Regional Office for Europe. Wellbeing measures in primary health care/the DepCare Project: report on a WHO meeting: Stockholm, Sweden, 12–13 February 1998. World Health Organization. 1998. URL: https://apps.who.int/iris/handle/10665/349766 [accessed 2022-09-15]

47. Stewart-Brown S, Janmohamed K. Warwick-Edinburgh Mental Well-being Scale (WEMWBS) User Guide Version 1. National Health Service. 2008 Jun. URL: http://www.mentalhealthpromotion.net/resources/user-guide.pdf [accessed 2022-09-15]
48. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75 [FREE Full text] [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
49. Mackinnon A, Jorm AF, Christensen H, Korten AE, Jacomb PA, Rodgers B. A short form of the Positive and Negative Affect Schedule: evaluation of factorial validity and invariance across demographic variables in a community sample. Pers Individ Differ 1999 Sep;27(3):405-416 [FREE Full text] [doi: 10.1016/s0191-8869(98)00251-7]
50. Ciccarello L, Reinhard MA. LGS. Lebensglückskala [Verfahrensdokumentation und Fragebogen mit Auswertung]. Open Test Archive. 2014. URL: https://www.testarchiv.eu/de/test/9006602 [accessed 2022-09-15]
51. Lyubomirsky S, Lepper HS. A measure of subjective happiness: preliminary reliability and construct validation. Soc Indic Res 1999;46:137-155. [doi: 10.1023/A:1006824100041]
52. Berkman PL. Life stress and psychological well-being: a replication of Langner's analysis in the midtown Manhattan study. J Health Soc Behav 1971 Mar;12(1):35-45 [FREE Full text] [doi: 10.2307/2948452]
53. Ryff CD, Keyes CL. The structure of psychological well-being revisited. J Pers Soc Psychol 1995;69(4):719-727. [doi: 10.1037/0022-3514.69.4.719]
54. International Wellbeing Group. Personal Wellbeing Index: 5th Edition. Australian Centre on Quality of Life, Deakin University. 2013. URL: https://www.acqol.com.au/uploads/pwi-a/pwi-a-english.pdf [accessed 2022-09-15]
55. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
56. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13 [FREE Full text] [doi: 10.1037/ccp0000459] [Medline: 31697093]
57. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070 [FREE Full text] [doi: 10.1037/0022-3514.54.6.1063]
58. Ainsworth B, Hardman D, Thomas M. The importance of differentiating behavioural and psychological treatment effects from placebo in respiratory interventions. Eur Respir J 2019 Apr;53(4):1900156 [FREE Full text] [doi: 10.1183/13993003.00156-2019] [Medline: 31023864]
59. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE Handbook. London, UK: The Cochrane Collaboration; 2013.
60. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013 Feb 08;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
61. Au J, Gibson BC, Bunarjo K, Buschkuehl M, Jaeggi SM. Quantifying the difference between active and passive control groups in cognitive interventions using two meta-analytical approaches. J Cogn Enhanc 2020 Jun;4(2):192-210 [FREE Full text] [doi: 10.1007/s41465-020-00164-6] [Medline: 34337311]
62. Bannink FP. Positive CBT: from reducing distress to building success. J Contemp Psychother 2013 Jun 26;44(1):1-8 [FREE Full text] [doi: 10.1007/s10879-013-9239-7]
63. White CA, Uttl B, Holder MD. Meta-analyses of positive psychology interventions: the effects are much smaller than previously reported. PLoS One 2019;14(5):e0216588 [FREE Full text] [doi: 10.1371/journal.pone.0216588] [Medline: 31141537]
64. Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf MI, et al. Searching for and selecting studies. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. London, UK: The Cochrane Collaboration; 2019.
65. Herner M. Perfect top of the evidence hierarchy pyramid, maybe not so perfect: lessons learned by a novice researcher engaging in a meta-analysis project. BMJ Evid Based Med 2019 Aug;24(4):130-132 [FREE Full text] [doi: 10.1136/bmjebm-2018-111141] [Medline: 30842066]
66. Ainsworth B, Steele M, Stuart B, Joseph J, Miller S, Morrison L, et al. Using an analysis of behavior change to inform effective digital intervention design: how did the PRIMIT website change hand hygiene behavior across 8993 users? Ann Behav Med 2017 Jun;51(3):423-431 [FREE Full text] [doi: 10.1007/s12160-016-9866-9] [Medline: 27909944]
67. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1-9 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]

Abbreviations

ACT: acceptance and commitment therapy
CBT: cognitive behavioral therapy

CONSORT: Consolidated Standards of Reporting Trials
ITT: intention to treat
MBI: mindfulness-based intervention
PP: per-protocol
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB: risk of bias
SMD: standardized mean difference
WHO: World Health Organization

Edited by J Torous; submitted 28.11.22; peer-reviewed by S Toh, S Hermsen; comments to author 15.02.23; revised version received 30.06.23; accepted 07.07.23; published 19.10.23.

Please cite as:
Groot J, MacLellan A, Butler M, Todor E, Zulfiqar M, Thackrah T, Clarke C, Brosnan M, Ainsworth B
The Effectiveness of Fully Automated Digital Interventions in Promoting Mental Well-Being in the General Population: Systematic Review and Meta-Analysis
JMIR Ment Health 2023;10:e44658
URL: https://mental.jmir.org/2023/1/e44658
doi:10.2196/44658
PMID:37856172

Review

Freely Available Training Videos for Suicide Prevention: Scoping Review

Katherine Wislocki[1], BA; Shari Jager-Hyman[2], PhD; Megan Brady[2], BS; Michal Weiss[2], BA; Temma Schaechter[2], BA; Gabriela Khazanov[2,3], PhD; Sophia Young[2]; Emily Becker-Haimes[2,4], PhD

[1]Department of Psychological Science, University of California, Irvine, Irvine, CA, United States

[2]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, United States

[3]Corporal Michael J Crescenz VA Medical Center, Philadelphia, PA, United States

[4]Hall Mercer Community Mental Health, University of Pennsylvania Health System, Philadelphia, PA, United States

Corresponding Author:
Katherine Wislocki, BA
Department of Psychological Science
University of California, Irvine
214 Pereira Drive
Irvine, CA, 92617
United States
Phone: 1 9498246803
Email: kwislock@uci.edu

Abstract

Background: Freely available and asynchronous implementation supports can reduce the resource burden of evidence-based practice training to facilitate uptake. Freely available web-based training videos have proliferated, yet there have been no efforts to quantify their breadth, depth, and content for suicide prevention.

Objective: This study presents results from a scoping review of freely available training videos for suicide prevention and describes a methodological framework for reviewing such videos.

Methods: A scoping review of freely available training videos (≥2 minutes) for suicide prevention practices was conducted using 4 large video-sharing platforms: YouTube, Vimeo, Bing Video, and Google Video. Identified suicide prevention training videos (N=506) were reviewed and coded.

Results: Most content was targeted toward gatekeepers or other lay providers (n=370) versus clinical providers (n=136). Videos most commonly provided content related to suicidal thoughts or behaviors (n=420). Many videos (n=274, 54.2%) included content designed for certain communities or organizations. Less than half (n=232, 45.8%) of training videos included formal clinical content pertaining to assessment or intervention for suicide prevention.

Conclusions: Results suggested an abundance of videos providing broad informational content (eg, “signs and symptoms of someone at risk for suicide”) and a limited portion of videos with instructional content aimed at clinical providers delivering formal evidence-based assessments or interventions for suicide prevention. Development of resources to address identified gaps may be needed. Future work may leverage machine learning techniques to expedite the review process.

(JMIR Ment Health 2023;10:e48404) doi:10.2196/48404

KEYWORDS

freely available videos; asynchronous training; suicide prevention; evidence-based practice; dissemination; implementation

Introduction

Suicide continues to be a pervasive public health crisis, with over 700,000 individuals dying by suicide annually worldwide [1]. Globally, research has indicated that the lifetime prevalence of suicidal ideation is nearly 10%, while the lifetime prevalence of suicide attempts is around 3% [2]. Health care providers play an essential role in delivering evidence-based practices (EBPs) to prevent or reduce suicidal thoughts or behaviors (STB) [3]. Many suicide prevention programs and initiatives also aim to leverage the influence of the general public, lay providers, or gatekeepers (ie, nonmental health professionals who have regular contact with the target population or community; hereafter referred to collectively as “gatekeepers”) in preventing

suicide. For example, previous work has demonstrated the role of gatekeepers in identifying those at heightened risk for suicide and connecting them to services [4,5]. Decades of research has emphasized the need for increased training for providers (both gatekeepers and clinical providers) in suicide prevention to maximize the reach of evidence-based suicide prevention techniques to reduce the number of people experiencing suicidal thoughts and behaviors, and ultimately, the number of lives lost to suicide [3,6]. The need to deliver and rapidly scale training opportunities is critical for increasing the number of individuals who can effectively deliver evidence-based suicide prevention techniques [3,7,8].

Effective suicide prevention efforts comprise a mix of strategies, including community education, assessment, and intervention (see [9,10] for reviews of leading evidence-based suicide prevention strategies). Several modalities, including workshops, exist to train gatekeepers and clinical providers in these strategies. In-person, workshop-based training is a prominent method for suicide prevention training. While workshop-based training alone is not sufficient for successful implementation [11], it is often considered a necessary step toward increasing the uptake and use of EBPs. Importantly, previous work has illustrated the importance of leveraging multiple modalities to deliver EBP training to support behavior change, such as the use of video-based skill demonstrations and in-person experiential education or one-on-one hybrid consultation following in-person workshops [11,12]. Unfortunately, the cost and time associated with workshop-based training can be prohibitive for organizations and providers. Training in a new EBP can cost providers (ie, clinicians or organizations) thousands of dollars, and costs escalate when including the consultation and support required for behavior change [13]. These costs are often infeasible, particularly for those in the public mental health system with limited funds [14,15]. An additional barrier to accessing training in suicide prevention EBPs is the limited number of programs that provide training on this topic [16].

It is imperative that the barriers to traditional training approaches are addressed to increase the accessibility of suicide prevention training. Previous work has identified digital approaches, including e-learning, as promising alternatives to traditional in-person methods of training [11,17,18]. Web-based training holds promise as a way for providers to access evidence-based training as either standalone training or as part of a broader training effort (ie, in support of training workshops and graduate training). Further, research assessing web-based training has indicated that it can be potentially comparable to in-person methods of training [11,18]. For suicide prevention specifically, results from a randomized controlled trial of Collaborative Assessment and Management of Suicidality training formats demonstrated comparable outcomes between asynchronous e-learning (leveraging video content) and traditional, in-person learning [18]. This intervention was only freely available to mental health providers within the US Veterans Administration [18]. However, outside of research on specific digital training interventions, there is a limited understanding of other digital training resources for suicide prevention, particularly those that are freely available to the public.

Freely available web-based training can assist in scaling access to training in suicide prevention EBPs. Free digital training content across health care topics has been associated with several benefits, including user satisfaction [19,20], usefulness [19,21], knowledge [22], and self-efficacy [23]. Notably, however, research assessing freely available web-based content is challenging due to the inherent decentralization of producing, distributing, and hosting web-based training content. Previous work on freely available mental health-related content available on the internet related to assessment and treatment has demonstrated a wide range in the quality of content [24-26]. Further complicating this issue, existing platforms that host or distribute web-based training content largely do not review or examine the content beyond aspects related to terms of services, such as copyright infringement or community safety violations (ie, encouraging harm to self or others) [27,28].

Prior research has focused on web-based training interventions described in academic publications [23,29]. However, this approach is limited, as only a small portion of web-based content is likely to be disseminated or examined through academic journals (ie, such as those developed, disseminated, and assessed in intervention studies that are subsequently published in academic journals). Additionally, the included content in reviews targeting academic sources is a potentially biased sampling of extant web-based training. These reviews may not address or include training videos disseminated by clinicians or nonacademic stakeholders. More work aggregating and reviewing training content directly from sources and end users is necessary to better understand the current landscape.

Identifying the landscape of existing freely available video content for suicide prevention is a critical first step toward understanding how to effectively leverage and disseminate training resources to gatekeepers and clinical providers. In addition, understanding existing resources may indicate gaps that would point to areas for future development. For example, it may be that certain interventions are covered in depth, whereas others are not, highlighting where additional development of freely available training resources is needed. Further, reviewing freely available training videos may provide insight into how to improve freely available training resources. While asynchronous training opportunities make it challenging to include experiential components that increase learning [11], some may include skill demonstrations of interventions in action that may facilitate deeper learning, compared to talking about skills conceptually [12]. Identifying the extent to which freely available training videos include skill demonstrations and other components to enhance learning can inform future development.

We conducted a scoping review of freely available web-based training videos for suicide prevention using 4 large, publicly available platforms: YouTube, Vimeo, Bing Video, and Google Video. Given that this is the first scholarly work on this topic, this scoping review focuses broadly on reviewing training videos related to the prevention and management of suicidal thoughts and behaviors, as well as related concepts like nonsuicidal self-injury (NSSI). Our primary aim was to identify and examine the landscape of freely available web-based training for suicide prevention strategies—across screening, assessment, and intervention. Our secondary aim was to demonstrate a novel

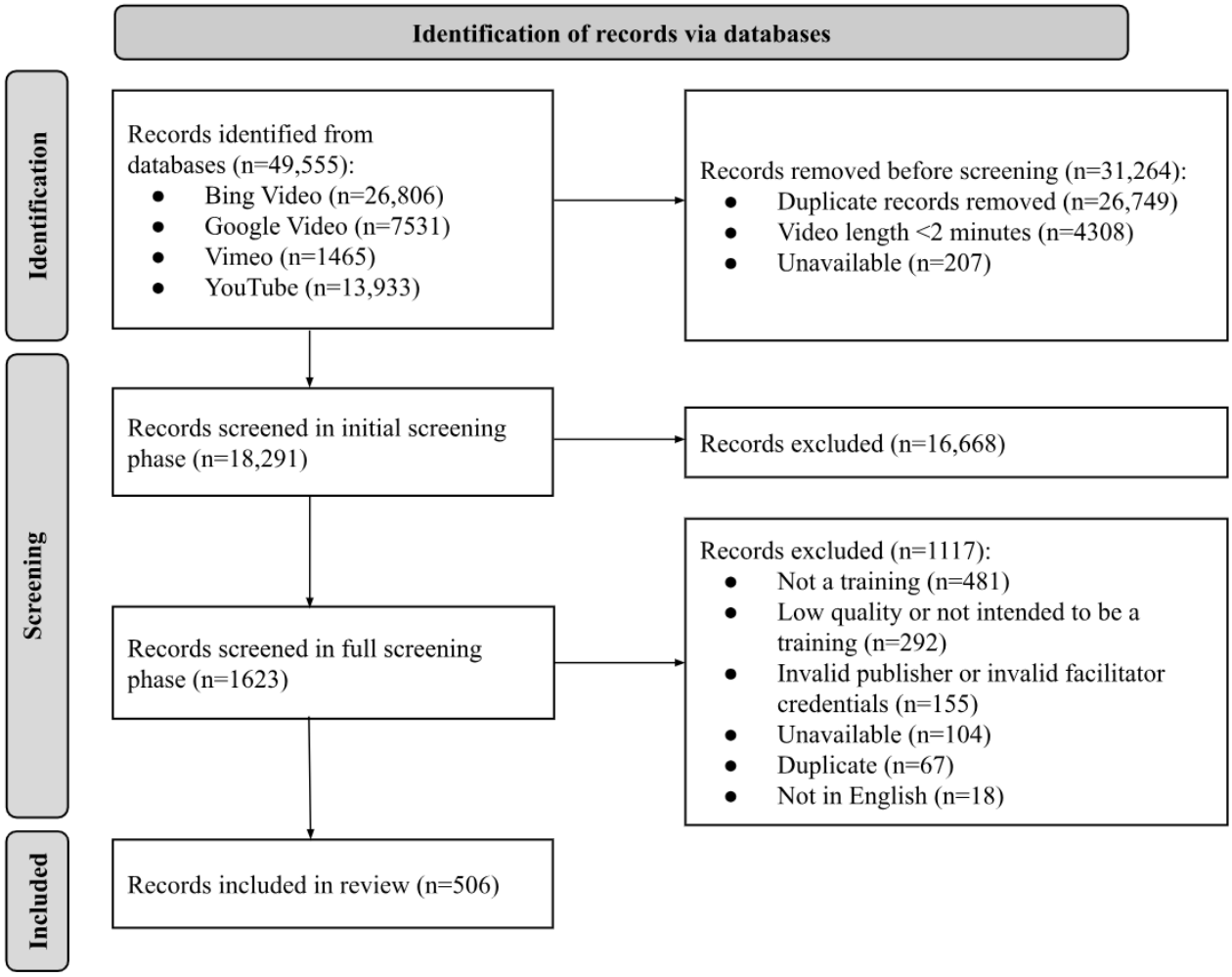
methodology for reviewing freely available video training content, which can be leveraged for other content areas. Ultimately, this work will provide critical knowledge to support translation efforts in suicide prevention training by illustrating the landscape of freely available training content and providing support for future research aimed at further understanding the quality of available content and areas where additional development is needed.

Methods

The 4 web-based sources selected for this search were YouTube, Vimeo, Bing Video, and Google Video. These sources were chosen for several reasons: they collectively host (YouTube, Vimeo, Google Video, and Bing Video) and index (Google Video and Bing Video) billions of freely available videos that are published on the internet, billions of users access these platforms on a daily basis, and there are application programming interfaces (APIs) available for each source that can be used to query and collect data [30-32]. Thus, there is a high likelihood that many freely available web-based training videos would be aggregated through these sources. The initial search took place in March and April 2021.

Search terms, search strategy, inclusion criteria, exclusion criteria, and content codebook were developed collaboratively by experts in suicide prevention, clinical psychology, and computational social science. Permutations of topic (ie, "suicide" and "self-harm") and medium (ie, "webinar" and "training") were made to create each search term. Each search term was queried using Python (Python Software Foundation) and each relevant platform search-specific API [30-32]. Note that a proprietary API search service was used for Google Video. An additional API was used to collect metadata from YouTube results [33] related to our primary inclusion criteria. Guidance for using these resources can be found in the documentation for each specific API [30-33], and the code used for querying each API can be obtained upon request. Our final search terms included: "suicide workshop," "suicide training," "suicide education," "suicide in-service," "suicide webinar," "suicide learning," "suicide online course," "suicide certification," "self-harm workshop," "self-harm training," "self-harm education," "self-harm in-service," "self-harm webinar," "self-harm learning," "self-harm online course," "self-harm certification," "self-injury workshop," "self-injury training," "self-injury education," "self-injury in-service," "self-injury webinar," "self-injury learning," "self-injury online course," and "self-injury certification." Total results (N=49,555) were returned and aggregated across Bing Video (n=26,806), YouTube (n=13,933), Google Video (n=7351), and Vimeo (n=1465). Information including hyperlinks, video identification number, account, length, title, and description was collected. Figure 1 depicts a modified PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram illustrating the filtering of information through each phase of this review.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



The identified videos were required to meet the following criteria to be eligible for inclusion in the final sample: (1) address detection, intervention, or postvention related to suicidal thoughts, suicidal behavior, or NSSI; (2) be free and publicly available; (3) be at least 2 minutes in length (a 2-minute minimum was set to account for existing microtraining sequences; videos less than 2 minutes were deemed unlikely to contain meaningful content); (4) have a working hyperlink; (5) present content in English; (6) have been posted by a government, community organization, academic-affiliated, or otherwise platform-verified source (this was chosen to exclude videos created by lay users such as that for school projects); and (7) have been designed to be delivered through a web-based format (eg, videos that were simply recordings of in-person training were excluded). We also excluded videos that (1) were not predominantly focused on detection, intervention, or postvention related to suicidal thoughts, suicidal behavior, or NSSI (defined as less than 50% of the video content); (2) were interviews with experts; (3) were primarily patient-facing (eg, how to seek help for yourself); (4) included promotional content (eg, advertisements for training with associated costs); (5) did not have distinguishable audio; (6) were news reports or were posted by news organizations; and (7) only included suicide-related content that was limited to describing risk factors without any guidance on how to screen, assess, or intervene with someone at risk.

A total of 3 rounds of initial screening coding were performed by the screening team (EB-H, SJ-H, GK, KW, and MN) prior to the final review and abstraction phase, with overall reliability equal to 80%. Results were deduplicated based on video links, and videos (N=18,291) were split equally between 5 screeners for the primary screening. In the first round of screening (akin to “abstract screening” in a traditional review), the screening team reviewed video titles, brief video descriptions, and author information. Screeners additionally identified clearly ineligible videos.

The second round took place in August 2021, where “maybes” or videos (n=899) in which an inclusion or exclusion decision was not made by a single author were uploaded to Rayyan (Rayyan), a reference management software [34]. In this round, each video was rescreened by at least 2 different authors using Rayyan [34]. Videos in this subset that did not receive either an inclusion or exclusion decision were discussed by authors (EB-H, SJ-H, GK, and KW) through weekly consensus discussions to make final determinations and identify videos that required a full review before eligibility could be determined.

In September 2021, the final phase included a full screening and abstraction of videos (N=1623) performed by MB, TS, and MW. The codebook for abstraction was developed through collaboration between authors. Codes were proposed based on relevance to suicide prevention and feasibility associated with abstracting information for each proposed code. Codes were subsequently refined through weekly team meetings after application to a test subset of videos. Disagreements during the development of the codebook, or other review procedures, were solved through discussion among authors in weekly team meetings. Coding reliability was established prior to screening and abstracting the final sample. Through this process, reviewers achieved good consensus (80% reliability) with a subset of videos (n=66) prior to beginning independent coding. The final sample consisted of 506 videos (1.02% of the initial sample) that met inclusion for review. Coders abstracted relevant information from each video using a structured code sheet (Multimedia Appendix 1). Briefly, we abstracted information related to video content target (STB, NSSI, or a combination of STB and NSSI content), target audience (eg, gatekeeper and clinical provider), content (information about recognizing risk factors and information about treatment, screening, or assessment), whether specialized populations were addressed, video series information (if applicable), and whether the training included any skill demonstrations. Content targets were determined by the focus of the video, such that videos coded as STB include information directed at STB only, videos coded as NSSI offer information directed at NSSI only, and videos coded as a combination of STB and NSSI include information directed at both STB and NSSI. Bimonthly meetings were held among coders, and coders consulted with the senior author (EB-H) to maintain reliability and prevent drift. The full process, from initial screening to abstraction, took place from April 2021 to September 2022. A full list of included videos can be found in Multimedia Appendix 2. This scoping review adhered to the guidelines set by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 3).

Results

Overview of Identified Videos

Included videos (N=506) had a mean length of 44.6 (SD 33.03; range 2-175) minutes. As mentioned previously, at least 50% of each included video was dedicated to detection, intervention, or postvention related to suicidal thoughts, suicidal behavior, or NSSI. Most videos were hosted by YouTube (n=365, 72.1%) and intended for gatekeepers or nonprofessional providers (n=370, 73.1%) rather than clinical providers (n=136, 26.9%). Most videos were published after 2018 (n=296, 58.5%), and nearly a third of the videos were published in 2020 (n=155, 30.6%). Training content within the videos was produced by academic institutions or health care facilities (n=179, 31.08%), community organizations (n=211, 36.63%), government entities (n=166, 28.82%), and others (n=20, 3.47%). Of note, videos can be produced by more than one organization.

Roughly half of included videos included content targeted specifically to certain communities or organizations (eg, schools, military, religious communities, and tribal communities; n=274, 54.2%). Similarly, most included videos offered general content, such as warning signs or how to help individuals at risk for suicide more broadly (n=274, 54.2%; hereafter referred to as “broad content”). For example, some of these videos focused on recognizing suicide risk signs and providing information on how to connect at-risk individuals with appropriate care. In contrast, other videos included formal intervention or assessment content (ie, instructional content about clinical assessment and intervention; n=232, 45.8%; hereafter referred to as “formal clinical content”). For example, some of these videos included content focused on delivering formal clinical assessments in response to STB and responding appropriately to STB in various clinical settings. Notably, videos across both categories can be directed at gatekeepers or clinical providers. Several interventions were presented within the latter category, including safety planning (n=63; note this category included interventions referencing an “action plan” or “crisis plan” in addition to the formal Safety Planning Intervention given their similar nature; [35]), postvention (n=29), limiting access to lethal means (n=26), Question, Persuade, Refer (QPR; n=13; [36]), and general coping strategies (n=7). In general, content in these domains tended to consist more of broad overviews of information (eg, the types of information included in a safety plan), rather than in-depth content or illustrations of how to deliver formal clinical content.

Within the videos that included formal clinical content, main content targets included STB (n=204, 87.9%; Table 1), NSSI (n=19, 8.2%; Table 1), or a combination of both STB and NSSI (n=9, 3.9%; Table 1).

Table 1. Target audience, video content target, and focus of videos offering formal clinical content[a].

Target audience and video content target	Total videos, n	Targeted toward organization or system, n	Targeted toward specific lay population, n	Discusses screening or assessment, n	Discusses intervention strategies, n	Skill demonstration, n	Framework for video content, n
Gatekeepers—STB[b]	123	58	55	28	103	8	1
Gatekeepers—NSSI[c]	15	8	9	6	9	0	0
Gatekeepers—combination of STB and NSSI	5	2	2	2	3	0	1
Clinical providers—STB	81	55	N/A[d]	38	58	14	1
Clinical providers—NSSI	4	4	N/A	2	2	1	1
Clinical providers—combination of STB and NSSI	4	3	N/A	4	2	1	0
Total (N)	232	130	66	80	177	24	4

[a]Categories are not mutually exclusive. Videos may receive a code for multiple categories based on video content.

[b]STB: suicidal thoughts or behaviors.

[c]NSSI: nonsuicidal self-injury.

[d]N/A: not available.

Gatekeeper Videos

Of videos directed at gatekeepers (n=370), most videos offered broad content (n=227, 61.4%) compared to a minority that offered formal clinical content (n=143, 38.6%; Table 1). Some videos were directed at specific populations of gatekeepers, such as educators or school personnel (eg, school staff, teachers, and college or university staff; n=65), caregivers (n=27), military personnel (n=16), students or peers (n=11), and employers (n=11). Videos for gatekeepers provided content mostly focused on STB (n=306, 82.5%), compared to NSSI (n=47, 12.7%) or a combination of STB/NSSI (n=17, 4.6%).

Among the subset of gatekeeper videos that included formal clinical content (n=143), 46.2% (n=66) identified and targeted a specific type or group of gatekeepers, and 47.6% (n=68) were directed at a specific organization or system (Table 1). Of that formal clinical content, a vast majority of videos included intervention content (n=115, 80.4%; Table 1) and only around a quarter of videos included screening or assessment content (n=36, 25.2%; Table 1). Many of these videos varied in the degree to which clinical content was covered, with fewer videos offering in-depth instruction. Notably, only 8 (5.6%) videos included skill demonstrations (Table 1), and only 2 (1.4%) videos included a framework to support the presented content (eg, Cornell University Mental Health Framework, Public Health Action for the Prevention of Suicide; Table 1). A portion of videos designed for gatekeepers had content geared toward providing services for specialized populations (Table 2), including children or youth (n=55, 38.5%), veterans or military (n=14, 9.8%), native or indigenous populations (n=3, 2.1%), LGBTQIA+ (lesbian, gay, bisexual, transgender, queer, intersex, and asexual) individuals (n=1, 0.7%), and other groups (n=15, 10.5%). Other groups included specific racial or ethnic groups; individuals experiencing substance use disorder, bipolar disorder, or autism spectrum disorder; individuals who experienced traumatic events; older adult populations; religious groups; individuals with disabilities; sex workers; and families of military service members.

Table 2. Target audience, video content target, and specialized populations for videos offering formal clinical content[a].

Target audience and video content target	Specialized populations				
	Child or teen, n	LGBTQIA+[b], n	Native or indigenous individuals, n	Veterans or military, n	Other[c], n
Gatekeepers—STB[d]	47	1	3	8	12
Gatekeepers—NSSI[e]	5	0	0	6	2
Gatekeepers—combination of STB and NSSI	3	0	0	0	1
Clinical providers—STB	19	0	0	5	16
Clinical providers—NSSI	3	0	0	0	0
Clinical providers—combination of STB and NSSI	4	0	0	0	0
Total, N	81	1	3	19	31

[a]Categories are not mutually exclusive. Videos may receive a code for multiple categories based on video content.

[b]LGBTQIA+: lesbian, gay, bisexual, transgender, queer, intersex, and asexual.

[c]Other specialized populations included (from most to least frequently mentioned): specific racial or ethnic groups; individuals experiencing substance use disorder, bipolar disorder, or autism spectrum disorder; individuals who experienced traumatic events; older adult populations; religious groups; individuals with disabilities; sex workers; and families of military service members.

[d]STB: suicidal thoughts or behaviors.

[e]NSSI: nonsuicidal self-injury.

Clinical Provider Videos

In contrast to gatekeepers, most videos directed at clinical providers (n=136) offered formal clinical content (n=89, 65.4%), and a minority of videos offered broad content (n=47, 34.6%). Videos for clinical providers mostly targeted STB (n=114, 83.8%) compared to NSSI (n=11, 8.1%) or a combination of STB and NSSI (n=11, 8.1%).

Of videos including formal clinical content (n=89), most were targeted to a specific organization or system, such as academic or hospital settings (eg, universities, medical schools, and K-12 schools), government (eg, state health departments and federal organizations), health care settings (eg, emergency departments and primary care offices), community-based organizations (eg, nonprofits and foundations), or clinical settings (n=62, 70%). Most of these videos focused on intervention strategies (n=62, 70%), whereas fewer focused on screening or assessment (n=44, 49.4%). The depth of formal clinical content varied across videos, with few providing in-depth instruction and skill demonstrations (Table 1). A portion of videos for clinical providers dedicated a majority of content (ie, ≥50% of video content) to specialized populations, including children or youth (n=26, 29.2%), veterans or soldiers (n=5, 5.6%), and other groups (n=16, 18%; Table 2). A small subset of videos (n=16, 18%) contained skill demonstrations, and only a few videos contained a framework for the video content (eg, Comprehensive Approach to Suicide Prevention, Polyvagal Model of NSSI; n=2, 2.2%; Table 1).

Discussion

This scoping review aimed to examine the landscape of freely available suicide prevention training videos and to demonstrate a novel methodology for reviewing free online video training content. To our knowledge, this is the first effort to empirically quantify the landscape of free video trainings related to mental health. Our methods for identifying existing resources proved to be feasible, albeit time-intensive. We focused on suicide prevention in this first effort given the significant public health burden of suicide and the critical need to advance training efforts related to suicide prevention.

Overall, our findings suggest that freely available suicide prevention videos largely focus on STB (as compared to NSSI), general information about suicide risk rather than specific suicide prevention strategies, and are primarily designed for gatekeeper audiences. A vast majority of videos did not provide a framework underlying the presented information. Most videos were published within the last 6 years, with a disproportionate number published in 2020. This trend likely reflects a global shift to digital and hybrid training opportunities during the COVID-19 pandemic.

Our results point to major gaps in the current landscape of freely available suicide prevention training videos. In particular, there is a relative dearth of formal clinical training content (ie, instructional content about delivering clinical assessments or interventions) for clinical providers working with individuals at risk for suicide. Although some videos provided general overviews of EBPs, few presented in-depth implementation guidance or demonstrations. Few videos presented content relating to the use of formal treatments for STB or NSSI, like dialectical behavior therapy, for example. In addition, few videos, across target audiences, focused on NSSI in isolation or in combination with STB. There was also less content directed at certain specialized populations at higher risk of STB or NSSI (eg, children or youth, veterans or military, native or indigenous populations, LGBTQIA+ individuals, and individuals with disabilities). Although a few videos included skill

demonstrations to illustrate how to deliver presented interventions, most content consisted of passive, didactic content inconsistent with recommended practices for training.

While the identified gaps may be a function of the search method (ie, focusing on suicide-focused content instead of specifically searching for interventions that can be applied to STB), this is also likely indicative of notable gaps in freely available resources related to suicide prevention. There may be several reasons for the lack of freely available training videos on certain topics, such as the proprietary nature of certain clinical intervention training content (ie, clinical protocols) and training formats (ie, workshops). There are likely other systemic factors that contribute to these gaps as well, such as the lack of incentives for the development and dissemination of freely available training videos for suicide prevention. Nonetheless, addressing these existing gaps is likely critical for improving and scaling up freely available resources for suicide prevention.

This review also demonstrated the feasibility of leveraging APIs to search, aggregate, and facilitate digital content review to bolster dissemination and implementation research. Partnerships between platforms and researchers can provide greater access to content for research purposes [37]. That said, the number of returned videos in this scoping review exceeded expectations, resulting in greater resources being devoted to screening and coding these videos than originally anticipated. Reviews are typically resource-intensive [38]; however, this review entailed sifting through thousands of videos and then reviewing and abstracting hundreds of hours of video footage. To address this, future work aiming specifically to characterize freely available training resources would benefit from leveraging machine learning techniques to expedite the review process. This can be accomplished at numerous steps in the review process. However, prior work suggests it may be most feasible and accurate within certain phases, such as the screening phase [39]. For example, the use of classification algorithms based on video features (ie, title, description, and publisher information) can potentially accelerate the screening process and is an interesting area for future development. Importantly, this has not been previously assessed with reviews of video content, and it should be validated prior to implementation.

Several limitations of this work should be noted. First, the methodology we used may not capture all freely available web-based training content due to the decentralized nature of freely available web-based resources. While we used some of the most robust platforms for web-based content, freely available resources could be hosted or distributed through different platforms (ie, individual blogs and websites) that were not included in this review. However, given the scope of the reviewed content and the number of duplicate videos identified, we believe that this is a generally representative sample of extant freely available videos for suicide prevention. Second, our coding scheme was constrained to factors that could be reliably assessed across videos, which precluded the inclusion of certain factors of interest, such as sponsorship characteristics or cost of video production, that are not routinely included in training videos. Third, this scoping review was limited to videos with English content. Fourth, other exclusion criteria (ie, video length of at least 2 minutes in length and publisher requirements) may have prohibited the inclusion of content focused on suicide prevention techniques. Further work should explore freely available videos for suicide prevention in different languages, from a range of source types, and of varying lengths. Finally, since this was the first review of its kind and followed established methodological guidelines for scoping reviews [40], no coding was done related to the quality of these training videos.

A critical next step in this line of research will be to conduct an appraisal of the quality of existing videos. As a quality appraisal is not typically included in a scoping review [40], more work is needed to thoroughly evaluate the quality of the content presented within freely available training videos for suicide prevention. While training videos that explicitly encouraged harm or harmful practices were excluded as part of this review during the screening phase, we did not limit our search to evidence-based practices. This is a potential limitation for disseminating this content and an important area for future research. Previous work has used different criteria for assessing the quality of video content [19,25], and this may be relevant for future research.

More research is needed to establish a taxonomy for matching specific resources to end users. Importantly, individuals and organizations should screen training videos to evaluate (1) the quality of the specified resource and (2) how this resource may align with their needs. For example, gatekeepers may benefit from using training videos that correspond to their needs in working with specific populations. Given the lack of skill demonstrations we found in our review, the development of new resources would benefit from prioritizing the provision of detailed instruction on EBPs. It is also important to create evidence-based resources for individuals working with populations that are under-addressed in existing videos. These gaps should ideally be filled by individuals and organizations who are qualified to deliver training related to EBPs for suicide prevention. Video developers should strive to collaborate with researchers and end users to create evidence-based resources that fulfill the needs of consumers.

In conclusion, this is the first review of its kind focused on suicide prevention training, and to our knowledge, the first of its kind in the mental health field. Given calls for increased training for both gatekeepers and professional providers, the findings within this review are critical for advancing existing and future research and training efforts related to suicide prevention. The sample for this scoping review was drawn from large publicly available video-hosting platforms that reach a vast audience, including lay and professional health care providers. Providers often use platforms such as these to search and access information [20]. Increasingly, producers are leveraging these platforms to develop and disseminate training content to health care providers [19,20]. Previous work in other contexts has focused on aggregating and reviewing video content from content developers through academic papers or research studies, rather than searching these platforms directly [23,25,29]. By searching these platforms directly, this review targeted content that is consumer-focused and likely not represented in academic literature. Further, several video-hosting platforms were used to obtain a large sample of representative content for

this review. Future research may opt to assess video resources by surveying consumers or producers directly, rather than video-hosting platforms. Results from this scoping review provide information on the landscape of freely available video content for suicide prevention. Future work may leverage and adapt this methodology provided to explore other topics.

Acknowledgments

This work was supported in part by funds from the National Institute of Mental Health (P50 MH127511; MPIs: Brown, Oquendo). The authors gratefully acknowledge Ms Meredith Nally for her contribution to this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook for full review phase.
[DOCX File , 25 KB - mental_v10i1e48404_app1.docx]

Multimedia Appendix 2
Included videos.
[XLSX File (Microsoft Excel File), 77 KB - mental_v10i1e48404_app2.xlsx]

Multimedia Appendix 3
Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.
[PDF File (Adobe PDF File), 594 KB - mental_v10i1e48404_app3.pdf]

References

1. Suicide worldwide in 2019: global health estimates. World Health Organization. 2021. URL: https://apps.who.int/iris/handle/10665/341728 [accessed 2023-01-10]
2. Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, et al. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. Br J Psychiatry 2008;192(2):98-105 [FREE Full text] [doi: 10.1192/bjp.bp.107.040113] [Medline: 18245022]
3. Schmitz WM, Allen MH, Feldman BN, Gutin NJ, Jahn DR, Kleespies PM, et al. Preventing suicide through improved training in suicide risk assessment and care: an American Association of Suicidology Task Force report addressing serious gaps in U.S. mental health training. Suicide Life Threat Behav 2012;42(3):292-304. [doi: 10.1111/j.1943-278X.2012.00090.x] [Medline: 22494118]
4. Burnette C, Ramchand R, Ayer L. Gatekeeper training for suicide prevention: a theoretical model and review of the empirical literature. Rand Health Q 2015;5(1):16 [FREE Full text] [Medline: 28083369]
5. Isaac M, Elias B, Katz LY, Belik SL, Deane FP, Enns MW, et al. Gatekeeper training as a preventative intervention for suicide: a systematic review. Can J Psychiatry 2009;54(4):260-268. [doi: 10.1177/070674370905400407] [Medline: 19321032]
6. Dexter-Mazza ET, Freeman KA. Graduate training and the treatment of suicidal clients: the students' perspective. Suicide Life Threat Behav 2003;33(2):211-218. [doi: 10.1521/suli.33.2.211.22769] [Medline: 12882422]
7. Hogan MF, Grumet JG. Suicide prevention: an emerging priority for health care. Health Aff (Millwood) 2016;35(6):1084-1090 [FREE Full text] [doi: 10.1377/hlthaff.2015.1672] [Medline: 27269026]
8. Spottswood M, Lim CT, Davydow D, Huang H. Improving suicide prevention in primary care for differing levels of behavioral health integration: a review. Front Med (Lausanne) 2022;9:892205 [FREE Full text] [doi: 10.3389/fmed.2022.892205] [Medline: 35712115]
9. Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E, Sarchiapone M, et al. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry 2016;3(7):646-659. [doi: 10.1016/S2215-0366(16)30030-X] [Medline: 27289303]
10. Mann JJ, Michel CA, Auerbach RP. Improving suicide prevention through evidence-based strategies: a systematic review. Am J Psychiatry 2021;178(7):611-624 [FREE Full text] [doi: 10.1176/appi.ajp.2020.20060864] [Medline: 33596680]
11. Frank HE, Becker-Haimes EM, Kendall PC. Therapist training in evidence-based interventions for mental health: a systematic review of training approaches and outcomes. Clin Psychol (New York) 2020;27(3):e12330 [FREE Full text] [doi: 10.1111/cpsp.12330] [Medline: 34092941]
12. Baum BE, Gray JJ. Expert modeling, self-observation using videotape, and acquisition of basic therapy skills. Prof Psychol Res Pract 1992;23(3):220-225. [doi: 10.1037/0735-7028.23.3.220]

13. Okamura KH, Wolk CLB, Kang-Yi CD, Stewart R, Rubin RM, Weaver S, et al. The price per prospective consumer of providing therapist training and consultation in seven evidence-based treatments within a large public behavioral health system: an example cost-analysis metric. Front Public Health 2017;5:356 [FREE Full text] [doi: 10.3389/fpubh.2017.00356] [Medline: 29359126]
14. Adams DR, Williams NJ, Becker-Haimes EM, Skriner L, Shaffer L, DeWitt K, et al. Therapist financial strain and turnover: interactions with system-level implementation of evidence-based practices. Adm Policy Ment Health 2019;46(6):713-723 [FREE Full text] [doi: 10.1007/s10488-019-00949-8] [Medline: 31203492]
15. Pegg SL, Walsh LM, Becker-Haimes EM, Ramirez V, Jensen-Doss A. Money makes the world go 'round: a qualitative examination of the role funding plays in large-scale implementation and sustainment of youth evidence-based practice. Psychol Serv 2021;18(2):265-274. [doi: 10.1037/ser0000399] [Medline: 31599621]
16. Brodsky BS, Spruch-Feiner A, Stanley B. The zero suicide model: applying evidence-based suicide prevention practices to clinical care. Front Psychiatry 2018;9:33 [FREE Full text] [doi: 10.3389/fpsyt.2018.00033] [Medline: 29527178]
17. German RE, Adler A, Frankel SA, Stirman SW, Pinedo P, Evans AC, et al. Testing a web-based, trained-peer model to build capacity for evidence-based practices in community mental health systems. Psychiatr Serv 2018;69(3):286-292 [FREE Full text] [doi: 10.1176/appi.ps.201700029] [Medline: 29137558]
18. Magruder KM, York JA, Knapp RG, Yeager DE, Marshall E, DeSantis M. RCT evaluating provider outcomes by suicide prevention training modality: in-person vs. e-learning. J Ment Health Train Educ Pract 2015;10(4):207-217. [doi: 10.1108/jmhtep-09-2014-0028]
19. Curran V, Simmons K, Matthews L, Fleet L, Gustafson DL, Fairbridge NA, et al. YouTube as an educational resource in medical education: a scoping review. Med Sci Educ 2020;30(4):1775-1782 [FREE Full text] [doi: 10.1007/s40670-020-01016-w] [Medline: 34457845]
20. Topps D, Helmer J, Ellaway R. YouTube as a platform for publishing clinical skills training videos. Acad Med 2013;88(2):192-197 [FREE Full text] [doi: 10.1097/ACM.0b013e31827c5352] [Medline: 23269305]
21. Desai T, Sanghani V, Fang X, Christiano C, Ferris M. Assessing a nephrology-focused YouTube channel's potential to educate health care providers. J Nephrol 2013;26(1):81-85. [doi: 10.5301/jn.5000115] [Medline: 22460179]
22. Katz M, Nandi N. Social media and medical education in the context of the COVID-19 pandemic: scoping review. JMIR Med Educ 2021;7(2):e25892 [FREE Full text] [doi: 10.2196/25892] [Medline: 33755578]
23. Sterling M, Leung P, Wright D, Bishop TF. The use of social media in graduate medical education: a systematic review. Acad Med 2017;92(7):1043-1056 [FREE Full text] [doi: 10.1097/ACM.0000000000001617] [Medline: 28225466]
24. Becker-Haimes EM, Tabachnick AR, Last BS, Stewart RE, Hasan-Granier A, Beidas RS. Evidence base update for brief, free, and accessible youth mental health measures. J Clin Child Adolesc Psychol 2020;49(1):1-17 [FREE Full text] [doi: 10.1080/15374416.2019.1689824] [Medline: 31825683]
25. Helming AG, Adler DS, Keltner C, Igelman AD, Woodworth GE. The content quality of YouTube videos for professional medical education: a systematic review. Acad Med 2021;96(10):1484-1493 [FREE Full text] [doi: 10.1097/ACM.0000000000004121] [Medline: 33856363]
26. King S, Ritchie KC, McGonnell M, Doe C, Corkum PV, Côté E, et al. Paging Dr. Google: availability and reliability of online evidence-based treatment information about ADHD. Evid Based Pract Child Adolesc Ment Health 2021;6(2):277-289 [FREE Full text] [doi: 10.1080/23794925.2021.1901632]
27. Google. YouTube community guidelines enforcement. Google Transparency Report. 2022. URL: https://transparencyreport.google.com/youtube-policy/removals?hl=en [accessed 2023-01-10]
28. Vimeo acceptable use community guidelines. Vimeo. URL: https://vimeo.com/help/guidelines [accessed 2023-01-10]
29. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ 2022;22(1):382 [FREE Full text] [doi: 10.1186/s12909-022-03446-z] [Medline: 35590410]
30. Search: list. YouTube Data API. URL: https://developers.google.com/youtube/v3/docs/search/list [accessed 2023-03-31]
31. Microsoft. Bing Video search API. Microsoft. URL: https://www.microsoft.com/en-us/bing/apis/bing-video-search-api [accessed 2023-01-10]
32. The basics. Vimeo. URL: https://developer.vimeo.com/api/guides/start [accessed 2023-01-10]
33. Videos. YouTube Data API. URL: https://developers.google.com/youtube/v3/docs/videos [accessed 2023-03-31]
34. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
35. Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. Cogn Behav Pract 2012;19(2):256-264. [doi: 10.1016/j.cbpra.2011.01.001]
36. Quinnett P. QPR: Certified QPR Gatekeeper Instructors Training Manual. Spokane, WA: The QPR Institute; 1995.
37. YouTube research program. YouTube. URL: https://research.youtube/ [accessed 2023-01-10]
38. Tsertsvadze A, Chen YF, Moher D, Sutcliffe P, McCarthy N. How to conduct systematic reviews more expeditiously? Syst Rev 2015;4:160 [FREE Full text] [doi: 10.1186/s13643-015-0147-7] [Medline: 26563648]
39. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Syst Rev 2019;8(1):163 [FREE Full text] [doi: 10.1186/s13643-019-1074-9] [Medline: 31296265]

40. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015;13(3):141-146 [FREE Full text] [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]

Abbreviations

API: application programming interface
EBP: evidence-based practice
LGBTQIA+: lesbian, gay, bisexual, transgender, queer, intersex, and asexual
NSSI: nonsuicidal self-injury
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QPR: Question, Persuade, Refer
STB: suicidal thoughts or behaviors

Edited by J Torous; submitted 22.04.23; peer-reviewed by H Szlyk, P Batterham, E Chan; comments to author 28.06.23; revised version received 23.08.23; accepted 31.08.23; published 03.11.23.

Please cite as:
Wislocki K, Jager-Hyman S, Brady M, Weiss M, Schaechter T, Khazanov G, Young S, Becker-Haimes E
Freely Available Training Videos for Suicide Prevention: Scoping Review
JMIR Ment Health 2023;10:e48404
URL: https://mental.jmir.org/2023/1/e48404
doi:10.2196/48404
PMID:37921847

Review

The Efficacy and Therapeutic Alliance of Augmented Reality Exposure Therapy in Treating Adults With Phobic Disorders: Systematic Review

Safa Hasan[1*], MBBS, MSc; Hamid Alhaj[2*], MD, PhD, FRCPsych; Athanasios Hassoulas[1], PhD, SFHEA, CPsychol

[1]Cardiff University, Cardiff, United Kingdom

[2]University of Sharjah, Sharjah, United Arab Emirates

[*]these authors contributed equally

Corresponding Author:
Hamid Alhaj, MD, PhD, FRCPsych
University of Sharjah
College of Medicine, University of Shajah
Sharjah
United Arab Emirates
Phone: 971 65057239
Email: halhaj@sharjah.ac.ae

Abstract

Background: Phobic disorders are characterized by excessive fear of a stimulus that can affect the quality of a patient's life. The lifetime prevalence in adults is 7.7% to 12.5%. The current literature provides evidence-based inferences about the effectiveness of in-vivo exposure therapy (IVET) in treating phobia. However, this method can put the therapist and the client in danger, with high drop out and refusal rates. A newer approach for exposure therapy using augmented reality technology is under assessment.

Objective: This systematic review investigated the novel technology's efficacy, cost-efficacy, and therapeutic alliance in treating adults with phobia.

Methods: An extensive search was conducted using 4 major databases (MEDLINE, PsycINFO, Embase, and Scopus) using a comprehensive list of synonyms for augmented reality exposure therapy (ARET) and phobic disorders. The search targeted any randomized control trial testing ARET in adults with phobic disorders up to August 8, 2022.

Results: A total of 6 studies were included, with 208 participants providing results. Studies investigating the efficacy of ARET compared to no intervention showed significant results (P<.05) in the ARET group improvement. Head-to-head comparative studies comparing ARET to IVET showed no significant difference (P>.05) in the effectiveness and therapeutic alliance between both therapies. Further, the results demonstrated that the ARET group had a better long-term effect than IVET, with the ability to put the patients in more situations to face the feared object.

Conclusions: The current data suggest clinically significant efficacy and a promising therapeutic alliance of ARET. However, no data are available investigating the cost-effectiveness of ARET. Further research is warranted to ascertain ARET's cost-effectiveness and examine its efficacy in other populations and anxiety conditions.

(JMIR Ment Health 2023;10:e51318) doi:10.2196/51318

KEYWORDS

augmented reality; virtual reality; anxiety disorders; phobic disorders; exposure therapy; augmented reality exposure; phobia; excessive fear; prevalence; technology; cost-effectiveness; fear; phobia; phobic

Introduction

Background

Phobic disorders are classified under the umbrella of anxiety and fear-related disorders and are characterized by excessive and abnormal fear or anxiety toward specific things or situations [1]. Fear is a psychological response to current stimuli, while anxiety is more concerned with future perceived anticipated stimuli or threats [1,2]. In the United States, adults' lifetime prevalence of specific phobias is estimated to range between 7.7% and 12.5% [3].

Notably, the first line of management in phobic disorders is psychotherapy; however, it can be augmented with medications [4,5]. Different modalities of cognitive behavioral therapy (CBT) can be used, although CBT with exposure is believed to be the most effective treatment [6,7]. CBT targets the patients' way of cognition and thinking to help them change their behaviors toward a specific issue and help them develop new techniques (deconditioning and counterconditioning) for dealing with their problems [7,8]. To avoid the limitations of in-vivo exposure therapy (IVET), like high dropout rate and low treatment acceptance [6], newer methods are being developed, including the virtual reality exposure therapy (VRET) [9]. Virtual reality (VR) technology was noted to assist in the exposure therapy with the aim of encountering the object or the case in a safer environment, and it will help create exposures that are hard to test in real life [9,10].

VR technology has been extensively studied, with evidence suggesting the superiority of VR over traditional psychotherapies in multiple areas of psychiatric disorders [11]. Riva et al [12] discussed the different uses of VR in psychiatric illnesses, and based on 27 systematic reviews and meta-analysis, they concluded the effectiveness of VR in the treatment of anxiety disorders, posttraumatic stress disorder, eating disorders, obesity, and pain management.

In the past decade, augmented reality (AR) has been routinely used in various fields such as medicine, entertainment, maintenance, design and architecture, teaching, and cognitive and motor rehabilitation [13]. AR has been shown to be effective in some regions of medicine, including improving physical activity and learning in autism [14,15]. Given that AR enhances the real-world environment with computer-generated sensory input, such as images, sounds, or other information, it has been proposed that they may be particularly useful in the treatment of phobic disorders.

Recent studies comparing the 2 technologies appear to support the advantage of AR compared with VR in particular psychiatric conditions [16]. While both VR and AR have shown promising results in treating anxiety disorders, it is thought that AR may have the advantage of providing a less intense form of exposure therapy, since the patient is still in the real world but with the added support of computer-generated elements. Vinci et al [17] discussed AR as an effective novel technology for treating substance dependence and anxiety disorders. In addition, Riva et al [12] and Vinci et al [17] suggested that AR adds to the benefits of normal clinical psychology and discussed the potential effectiveness of AR. However, limited data were available to provide any conclusive results for the efficacy of AR in various types of psychiatric disorders.

The therapeutic alliance is very crucial in the management of psychiatric disorders and is considered a key factor in the success of therapy. A strong therapeutic alliance is characterized by mutual trust, respect, and collaboration between the therapist and client, with the goal of achieving the client's therapeutic goals. Therapeutic alliance can be tested by 3 main domains of a good alliance [18,19]. In addition, Tracey and Kokotovic [18] described the 3 domains: goal, task, and bond. The therapeutic alliance is one of the essentials for a good outcome in psychotherapy; a good quality therapeutic alliance is linked to the higher success of the psychotherapeutic approach [19]. So, to agree on the efficacy of any intervention, a good-quality therapeutic alliance should occur. Based on the current literature, the therapeutic alliance in technology-mediated psychotherapy is still questioned [20].

Objectives

The main objective of this systematic review is to understand how effective augmented reality exposure therapy (ARET) is in treating phobic disorders. In addition, we aimed to investigate the best strategy to be used in exposure therapy with a lower recurrence rate, lower cost, and higher acceptance by the patients. Thereby, therapeutic alliance, cost-effectiveness, and efficacy of AR will be investigated.

Methods

Search Strategy

A comprehensive search included 4 databases. MEDLINE was our primary database to test for the eligibility of the search strategy, and the search strategy was duplicated on Embase, PsycINFO, and Scopus. The keywords included Phobia OR Phobia* OR "Phobic AND disorders" OR Agoraphobia AND "augmented AND reality" OR "augmented AND reality AND exposure AND therapy" OR "ARET" OR "mixed AND reality". Furthermore, gray literature was not included in the search to guarantee better quality papers as the grey literature is not peer-reviewed literature. However, excluding the gray literature can decrease the new studies done in this area and increase the risk of publication bias in this study [21].

Definitions

AR is defined as a technology that superimposes an overlay of simulated objects [22,23]. AR psychotherapy was defined as any psychotherapy that uses AR technology, like ARET. Phobic disorders were defined according to standardized diagnostic criteria, namely Diagnostic and Statistical Manual of Mental Disorders (DSM) or International Classification of Diseases (ICD).

Inclusion and Exclusion Criteria

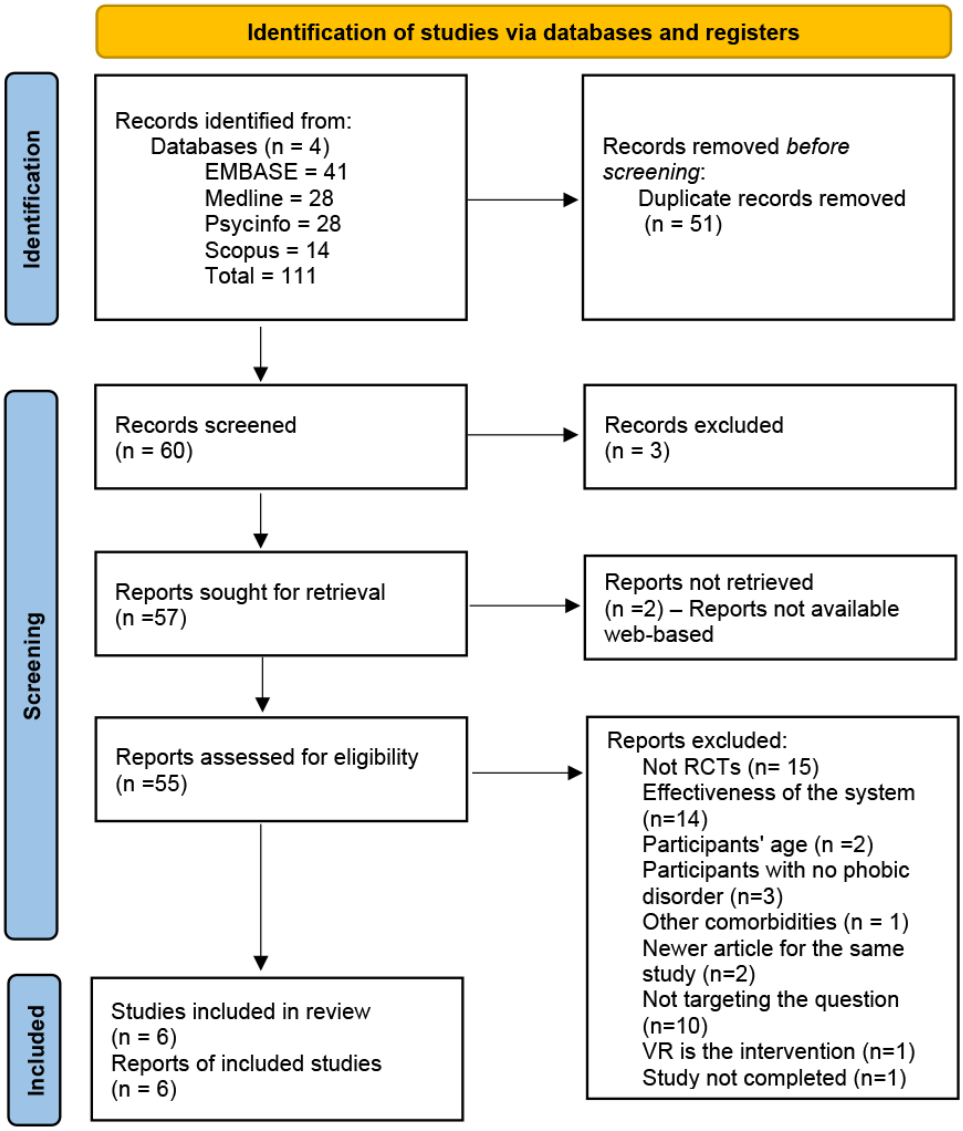
The selection of the studies followed the Population, Intervention, Comparison and Outcomes (PICOS) approach to focus on a specific question [24]. (1) Inclusion criteria: Participants are adolescents and adults between 15 and 75 years of age, those who met the criteria for the diagnosis of specific phobia and are eligible for treatment, and those with no other comorbid disorders. The intervention included any psychotherapy that uses AR technology, like ARET. The comparison groups include groups that use VR technology for exposure therapy, groups treated with the usual real-life exposure therapy (treatment as usual [TAU]), groups treated using CBT or any psychotherapy, or waiting lists. The outcome measures are multiple, including qualitative questionnaires (testing the behaviors of avoidance and the therapeutic alliance) and quantitative measures measuring the physiological effect of fear and anxiety, such as heart rate, electrocardiograms (ECGs), and skin conductance recordings (SCR). Finally,

randomized control trials are the design targeted. (2) Exclusion criteria: uncompleted studies or studies that did not have a comparison group.

Procedure

Initially, the MEDLINE database was searched for using AR technology to treat phobic disorders; a preliminary result of 122 records was retrieved and screened for relevancy. Some synonyms were removed from the search strategy as the results revealed unrelated studies. Later, a second run was done and included relevant studies. The search keywords were duplicated in all the databases. All the records were imported into EndNote 20 (Clarivate Analytics) and were screened for duplication and eligibility for inclusion (as per the criteria above). The abstracts were used for a primary screening; however, if the decision was unclear based on the abstract, a full study was retrieved to determine the eligibility. Later, after the final screening, the included papers' information was extracted into Excel (Microsoft Corp) sheets and Microsoft tables to help in the comparison; the quality assessment tools used helped select the needed data to extract and also, to map out the number of records and reports screened, retrieved, excluded, and included in this study [25]. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram was filled during each step of this study (Figure 1). Furthermore, the review included a qualitative synthesis of the available studies in this research field. The qualitative synthesis included a very detailed assessment of each study's findings. And multiple bias assessments were used to assess for common types of bias, including the Critical Appraisal Skills Programme (CASP) checklist, the Jadad scale, and the bias assessment table by Boland et al [24]. While a quantitative synthesis was supposed to be carried out, it was not done due to the unavailability of the raw data.

Figure 1. PRISMA diagram illustrating the selection procedure. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT: randomized controlled trial; VR: virtual reality.



Results

Overview

A total of 111 studies were found. After the first screening, 60 records remained and were screened, with 3 records being irrelevant, and 2 records being unavailable on the internet; 55 studies were screened for eligibility. A total of 7 studies met the inclusion criteria; however, 1 study met the exclusion criteria and was removed. A total of 208 participants were studied in 6 studies, excluding Toffolo et al [26], with 112 participants, as the study is not yet completed. Furthermore, the control group of the 7 studies was the following: 4 studies (4/6) had IVET as the control group, and 2 studies (2/6) had the waiting list as a control group.

Furthermore, different outcome measures were used to conclude the results, as displayed in Table 1. These include the Behavioral Avoidance Test (BAT), Fear of Spiders Questionnaire (FSQ), Spider Phobia Beliefs Questionnaire (SPBQ), and Working Alliance Inventory-Short (WAI-S). These are the most used validated qualitative questionnaires in studies with spider phobia. BAT evaluates the ability of the participants to change their maladaptive behavior (avoidance). At the same time, FSQ and the SPBQ are ideally used together to cover the gaps in SPBQ [27]. Finally, WAI-S is the short version of the Working Alliance Inventory, which is a validated scale to understand the therapeutic alliance between the therapist and the patients. Moreover, some studies used quantitative outcome measures like SCR; SCR measures the autonomic arousal caused by fear and anxiety. All studies have significant results showing ARET's good efficacy (and continuous efficacy) in treating phobic disorders. In addition, studies concerned about the therapeutic alliance of AR showed a good alliance between therapists and participants with no significant overall differences between ARET's and IVET's therapeutic alliance.

Studies comparing AR to the IVET (4/6) showed nonsignificant results between the 2 modalities in both the efficacy and the therapeutic alliance ($P>.05$), suggesting good effectiveness, feasibility, and a therapeutic alliance of AR. And the studies comparing the novel technology with a waiting list (2/6) showed significant statistical results with a $P<.05$, suggesting a good efficacy of ARET compared to nothing being done. The below Table 1 illustrates summary of the studies: clinical condition, participants, interventions, measures, and outcomes.

Table 1. Summary of the key studies included in the review with details of interventions, measures, and outcomes.

	Clinical condition	Participants	Intervention	Measures	Outcomes
IVET[a] for control					
Botella et al [28], Spain	Specific phobia (cockroaches phobia and arachnophobia)	In vivo exposure (N=31) augmented reality system (N=32)	One intensive session (up to 3 h long)—clinical settings	BAT[b], FSQ[c], SPBQ[d], Fear-MTB[e], and CSS[f]	• Directly posttreatment measures showed IVET superiority in some measures. However, follow-up at 3 and 6 months showed nonsignificant results between IVET and ARET[g]. • ARET is effective in treating small animal phobia. • ARET is accepted and well-tolerated by patients.
Wrzesien et al [29], Spain	Specific phobia (cockroaches phobia and arachnophobia)	N=22 randomly assigned to IVET or ARET	One intensive session (up to 3 h long)—clinical settings	BAT and WAI-S[h]	• Nonsignificant differences between ARET and IVET in the overall therapeutic alliance.
Wrzesien et al [30], Spain	Specific phobia (cockroaches Phobia and arachnophobia)	N=20, randomly assigned to IVET or ARET	One intensive session (up to 3 h long)—clinical settings	TCS[i], WAI-S, and BAT	• Nonsignificant differences between ARET and IVET in the overall therapeutic alliance. • TCS was correlated to WAI-S and BAT scores. • TCS helped in understanding the significant differences of WAI-S subscales with an overall limited effect on the alliance. • TCS can be used in future research as a scale to test for therapeutic alliance in technology mediated therapy.
Wrzesien et al [31], Spain	Specific phobia (cockroaches phobia and arachnophobia)	N=12, randomly assigned to IVET or ARET	One intensive session (up to 3 h long)—clinical settings	Participants' outcome measures: BAT, WAI-S, and scales for anxiety, avoidance, and beliefs. The therapists' measures: capacity of the ARET. Usefulness and the frequency of use. Scale for therapeutic alliance. Sessions recorded.	• ARET and IVET showed similar effectiveness on a clinical basis. • Both groups were able to interact with a live feared object after the session.
Waiting list for control					
Javanbakht et al [32], USA	Specific phobia (arachnophobia)	ARET (n=13) waiting list control (n=12)	The sessions conducted in the clinics aimed to gradually increase the intensity until the patient SUDS[j] score was below 4. This was followed with nonclinical settings for 1 month on patients' own base of exposure using AR[k] app.	FSQ, SPBQ, CEQ[l], BAT, SUDS, biosignal measures: SCR[m]	• ARET group were able to touch a live tarantula or the tank after 1 session, while waiting list group stayed a minimum of 1 meter away from the tank. • ARET efficacy on decreasing the symptoms remained the same or improved at 1 month follow.

	Clinical condition	Participants	Intervention	Measures	Outcomes
Zimmer et al [33], Switzerland	Specific phobia (arachnophobia)	Gamified AR spider exposure app (n=33), control group: waiting list (n=33)	Patients were trained in a clinical setting using 1 session of gradual exposure, then they were sent home with the app access to do 6 sessions of 30 minutes each, within 2 weeks period.	SUDS, FSQ, BAT, SBQ[n], GSE[o], 1 question to assess self-reported fear reduction of spiders.	• The repeated use of nonclinical settings reduced subjective fear in all measures and was effective at a low cost. • Nonclinical settings exposure was short and time-saving. • Exposure to a simulated object was safe in a nonclinical setting.

[a]IVET: in-vivo exposure therapy.

[b]BAT: Behavioral Avoidance Test.

[c]FSQ: Fear of Spiders Questionnaire.

[d]SPBQ: Spider Phobia Beliefs Questionnaire.

[e]Fear-MTB: Main Target Behavior (Fear).

[f]CSS: Clinician Severity Scale.

[g]ARET: augmented reality exposure therapy.

[h]WAI-S: Working Alliance Inventory-Short.

[i]TCS: Therapeutic Collaboration Scale.

[j]SUDS: Subjective Units of Distress Scale.

[k]AR: augmented reality.

[l]CEQ: Credibility/Expectancy Questionnaire.

[m]SCR: skin conductance recordings.

[n]SBQ: Spider phobia beliefs questionnaire.

[o]GSE: General Self-Efficacy Scale.

Bias Assessment of the Included Studies

Botella et al [28] and Wrzesien et al [30] studies fulfilled most of the requirements for high-quality research in the CASP checklist and rated 4 out of 5 on the Jadad scale. However, Wrzesien et al [29] and Wrzesien et al [31] did not report most of the sections in the CASP checklist and rated 1 out of 5 and 2 out of 5 on the Jadad scale, respectively; a lower quality and higher level of bias results are expected. Javanbakht et al [32] and Zimmer et al [33] scored a 5 out of 5 on the Jadad scale and filled out the CASP checklist perfectly; this means both are high-quality studies that can be reliable in their results with a lower chance of bias.

Discussion

AR Efficacy Compared With Other Modalities

Based on the results of this review, despite knowing that the spiders are simulated and unreal, ARET showed a statistically significant outcome in treating small animal phobia. Participants being treated by ARET in all 6 completed studies showed a substantial improvement in their phobic symptoms; each study used different outcome measures, but all the measurements were validated in testing the phobic symptoms. ARET showed no statistical differences (P>.05) compared to IVET's efficacy in 4 individual studies. This suggests a comparable efficacy with the standardized therapy of phobic disorders [8]. The improvement in the ARET compared to the waiting list group in 2 individual studies showed statistically significant results (P<.05). Thus, ARET leads to a positive impact on phobic symptoms.

In addition, this study suggests the need for longer sessions in cases of exposure therapy using extended reality like VRET or ARET [16,34]. However, Javanbakht et al [32] demonstrated the ability of AR to be effective in less than 1-hour sessions. As pointed out in the results section, the mean time needed for an ARET session to give the participants the ability to walk in a room with moving large spiders with Subjective Units of Distress Scale (SUDS) below 4 was 38 (SD 12.13) minutes.

Shiban [9] discussed the added benefits of ARET over traditional exposure therapy and how the therapist can control the scenarios more easily than the IVET. Wrzesien et al [31] and Botella et al [28] supported this theoretical opinion with evidence-based inferences. Part et al [31] found that the ARET group could interact in a greater number of situations with the phobic stimulus. Unlike IVET, the ARET group could observe the phobic stimulus moving on their personal belongings, observe dead objects, put their foot near the object, and find the phobic object under different artifacts. Moreover, Botella et al [28] discussed the different variety of spiders (size, number, types, and color) and the spiders' behaviors that were easier achieved in the ARET group. Correlating the findings to the current knowledge suggests the superiority of the ARET over the IVET as patients get higher and more intense exposure due to the ability of ARET to put the patient in more scenarios in different contexts easily and without real danger. The current review of the literature suggests a better outcome in cases with more intensity and a higher number of scenarios with the feared object [35,36].

Furthermore, exposure to an increased number of scenarios can be a reason for the long-term efficacy of the treatment and less

frequent recurrence [37,38]. However, it is not fully understood as Shiban et al [39] noted that an increased number of contexts would lose the positive effect after 15 days of the intervention completion. Based on our results, only Botella et al [28] had follow-up measures for the participants comparing the ARET with IVET; Botella et al [28] followed up with the patients 3 and 6 months after the completion of the treatment. Botella et al's [28] results revealed a positive effect on the long-term efficacy of ARET, with regression in the impact of IVET.

No randomized control trials comparing ARET to VRET were identified from the search. Thus, the superiority of ARET to VRET cannot be evaluated. However, Tsai et al [16] noted a higher presence in the use of ARET compared to VRET, with statistically significant differences in the physiological effect of the use of ARET to VRET, which suggests a better outcome of ARET.

AR and Therapeutic Alliance

As displayed in the results section, 3 studies used WAI-S as one of their outcome measures. The results showed a promising alliance between therapists and participants. These 3 compared the therapeutic alliance between ARET and IVET; the results displayed no statistically significant differences between the 2 groups in all 3 studies. However, the intervention group's results suggested a good relationship between the therapist and the participants. Compared to the traditional therapeutic approach, AR technology did not negatively or positively impact the therapeutic alliance. These results suggest ARET's good efficacy in treating small animal phobia.

Part et al [31] noted that the visual attention of the therapist using ARET was more focused on the notes and the phobic stimulus than on the patient. However, the overall attention given to the patient was nonsignificant compared to the IVET group. The nonsignificant results can be due to the small number of participants or the lack of a real effect. The current literature in this area suggests the problem in the therapeutic alliance in technology-mediated psychotherapy is due to the therapist's concerns about the alliance [20]. Part et al [31] findings propose the need for more training to use technology-mediated exposure therapy to limit attention distraction.

AR Limitations

While AR has the potential to be a valuable tool in treating phobias, it is important to acknowledge some of its limitations to ensure the safe and effective use of AR technology. These include technical challenges, cost-effectiveness and ethical considerations.

Wrzesien et al [31] discuss some of the limitations of ARET. The interaction with the object with a stick, pen, or paper was achieved in the IVET, not the ARET group. This is one of the ARET limitations where interaction with the visual object is limited. This limitation can be overcome with mixed reality technology instead; mixed reality technology allows interaction with a visual object, but no studies were found regarding mixed reality. Moreover, Wrzesien et al [31] noted some functional issues with the system used. However, this is not a limitation of ARET but the system used in this study.

AR Cost-Effectiveness

Baus and Bouchard [40], Bras et al [41], and Albakri et al [11] discussed the possibility of having a lower cost using AR instead of VR since fewer computer-generated objects are needed to be created; VR will need an entirely simulated environment while for AR only the stimulus object is created. However, this review failed to investigate the cost-effectiveness of AR, as no studies have been done looking into the different costs between the modalities. However, this can help us understand the need for studies looking into the cost of each modality to help us find the most cost-effective modality.

Strengths

This study aimed to provide evidence-based inferences to a focused question looking into the efficacy of novel technology in treating phobic disorders. The use of a focused question helped in narrowing and guiding the search. In addition, it helped develop a comprehensive search strategy to cover the search area without missing related studies. Furthermore, using the PICOS approach helped standardize the selection procedure; standardization gave a low heterogenicity between the included studies.

Furthermore, all included studies are published peer-reviewed studies; using peer-reviewed studies can lower the possibility of having low-quality studies.

Limitations

The number of studies the search reveals is small and can suggest a limitation. However, considering the comprehensive search strategy and the use of 4 massive databases, the small number revealed by the search is not due to a limitation in the search strategy. But instead, the problem is the small number of studies done in this area. The limited number of studies carried out in this area can be considered as the main limitation of this review.

Furthermore, the exclusion of grey literature in our study may be considered a limitation. It is expected that a wide range of unpublished work related to AR exists, increasing the potential of publication bias [21].

Conclusions

Our literature review demonstrated clinically significant efficacy and a promising therapeutic alliance of ARET. However, the currently available data do not provide evidence regarding the cost-effectiveness of ARET. Further research is warranted to ascertain ARET's cost-effectiveness and examine its efficacy in other populations and anxiety conditions.

Acknowledgments

No funding was provided from any pharmaceutical or technology companies to carry out this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[PDF File (Adobe PDF File), 69 KB - mental_v10i1e51318_app1.pdf]

References

1. International Classification of Diseases-11 (ICD-11) for mortality and morbidity statistics. World Health Organization. URL: https://icd.who.int/browse11/l-m/en#/http%3a%2f%2fid.who.int%2ficd%2fentity%2f1336943699 [accessed 2022-07-02]
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders DSM-5™, Fifth Edition. Washington, D.C: American Psychiatric Association Publishing; 2013.
3. McCabe RE, Stetin MB. Specific phobia in adults: epidemiology, clinical manifestations, course, and diagnosis. UpToDate. 2022. URL: https://www.uptodate.com/contents/specific-phobia-in-adults-epidemiology-clinical-manifestations-course-and-diagnosis [accessed 2022-02-08]
4. McCabe RE, Bui E, Stein MB. Approach to treating specific phobia in adults. UpToDate. 2022. URL: https://www.uptodate.com/contents/114843 [accessed 2022-02-08]
5. Baldwin DS, Anderson IM, Nutt DJ, Allgulander C, Bandelow B, den Boer JA, et al. Evidence-based pharmacological treatment of anxiety disorders, post-traumatic stress disorder and obsessive-compulsive disorder: a revision of the 2005 guidelines from the British Association for Psychopharmacology. J Psychopharmacol 2014;28(5):403-439 [FREE Full text] [doi: 10.1177/0269881114525674] [Medline: 24713617]
6. Choy Y, Fyer AJ, Lipsitz JD. Treatment of specific phobia in adults. Clin Psychol Rev 2007;27(3):266-286. [doi: 10.1016/j.cpr.2006.10.002] [Medline: 17112646]
7. Viña C, Herrero M, Rivero F, Álvarez-Pérez Y, Fumero A, Bethencourt JM, et al. Changes in brain activity associated with cognitive-behavioral exposure therapy for specific phobias: searching for underlying mechanisms. Rev Neurol 2020;71(11):391-398 [FREE Full text] [doi: 10.33588/rn.7111.2019487] [Medline: 33205385]
8. Johnstone EC, Owens DC, Lawrie SM, McIntosh AM, Sharpe M. Companion to Psychiatric Studies. Edinburgh: Elsevier Health Sciences; 2010.
9. Shiban Y. Virtual reality exposure therapy for anxiety disorders. Nervenarzt 2018;89(11):1227-1231. [doi: 10.1007/s00115-018-0596-z] [Medline: 30167721]
10. Types of therapy. Anxiety and Depression Association of America. URL: https://adaa.org/find-help/treatment-help/types-of-therapy#:~:text=A%20form%20of%20CBT%2C%20exposure,obsessive%2Dcompulsive%20disorder%20and%20phobias [accessed 2022-04-08]
11. Albakri G, Bouaziz R, Alharthi W, Kammoun S, Al-Sarem M, Saeed F, et al. Phobia exposure therapy using virtual and augmented reality: a systematic review. Appl Sci 2022;12(3):1672 [FREE Full text] [doi: 10.3390/app12031672]
12. Riva G, Baños RM, Botella C, Mantovani F, Gaggioli A. Transforming experience: the potential of augmented reality and virtual reality for enhancing personal and clinical change. Front Psychiatry 2016;7:164 [FREE Full text] [doi: 10.3389/fpsyt.2016.00164] [Medline: 27746747]
13. Giglioli IAC, Pallavicini F, Pedroli E, Serino S, Riva G. Augmented reality: a brand new challenge for the assessment and treatment of psychological disorders. Comput Math Methods Med 2015;2015:862942 [FREE Full text] [doi: 10.1155/2015/862942] [Medline: 26339283]
14. Adnan NH, Ahmad I, Abdullasim N. Systematic review on augmented reality application for autism children. J Adv Res Dyn Control Syst 2018;10(11):26-32 [FREE Full text]
15. Ng YL, Ma F, Ho FK, Ip P, Fu K. Effectiveness of virtual and augmented reality-enhanced exercise on physical activity, psychological outcomes, and physical performance: a systematic review and meta-analysis of randomized controlled trials. Comput Hum Behav 2019;99:278-291. [doi: 10.1016/j.chb.2019.05.026]
16. Tsai CF, Yeh SC, Huang Y, Wu Z, Cui J, Zheng L. The effect of augmented reality and virtual reality on inducing anxiety for exposure therapy: a comparison using heart rate variability. J Healthc Eng 2018;2018:6357351 [FREE Full text] [doi: 10.1155/2018/6357351] [Medline: 30595830]
17. Vinci C, Brandon KO, Kleinjan M, Brandon TH. The clinical potential of augmented reality. Clin Psychol (New York) 2020;27(3):e12357 [FREE Full text] [doi: 10.1111/cpsp.12357] [Medline: 33223628]
18. Tracey TJ, Kokotovic AM. Factor structure of the Working Alliance Inventory. Psychol Assess: J Consult Clin Psychol 1989;1(3):207-210. [doi: 10.1037/1040-3590.1.3.207]
19. Stubbe DE. The therapeutic alliance: the fundamental element of psychotherapy. Focus (Am Psychiatr Publ) 2018;16(4):402-403 [FREE Full text] [doi: 10.1176/appi.focus.20180022] [Medline: 31975934]
20. Lopez A, Schwenk S, Schneck CD, Griffin RJ, Mishkind MC. Technology-based mental health treatment and the impact on the therapeutic alliance. Curr Psychiatry Rep 2019;21(8):76. [doi: 10.1007/s11920-019-1055-7] [Medline: 31286280]

21. Pappas C, Williams I. Grey literature: its emerging importance. J Hosp Librariansh 2011;11(3):228-234. [doi: 10.1080/15323269.2011.587100]
22. Tremosa L. Beyond AR vs. VR: what is the difference between AR vs. MR vs. VR vs. XR? Interaction Design Foundation. URL: https://www.interaction-design.org/literature/article/beyond-ar-vs-vr-what-is-the-difference-between-ar-vs-mr-vs-vr-vs-xr [accessed 2022-07-10]
23. What is augmented reality? The Franklin Institute. 2022. URL: https://www.fi.edu/what-is-augmented-reality [accessed 2022-06-24]
24. Boland A, Cherry MG, Dickson R. Doing a Systematic Review: A Student's Guide, 2nd Edition. Los Angeles: SAGE Publications; 2017.
25. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
26. Toffolo MBJ, Fehribach JR, van Klaveren CPBJ, Cornelisz I, van Straten A, van Gelder JL, et al. Automated app-based augmented reality cognitive behavioral therapy for spider phobia: study protocol for a randomized controlled trial. PLoS One 2022;17(7):e0271175 [FREE Full text] [doi: 10.1371/journal.pone.0271175] [Medline: 35830423]
27. Szymanski J, O'Donohue W. Fear of Spiders Questionnaire. J Behav Ther Exp Psychiatry 1995;26(1):31-34. [doi: 10.1016/0005-7916(94)00072-t] [Medline: 7642758]
28. Botella C, Pérez-Ara M, Bretón-López J, Quero S, García-Palacios A, Baños RM. In vivo versus augmented reality exposure in the treatment of small animal phobia: a randomized controlled trial. PLoS One 2016;11(2):e0148237 [FREE Full text] [doi: 10.1371/journal.pone.0148237] [Medline: 26886423]
29. Wrzesien M, Bretón-López J, Botella C, Burkhardt JM, Alcañiz M, Pérez-Ara M, et al. How technology influences the therapeutic process: evaluation of the patient-therapist relationship in augmented reality exposure therapy and in vivo exposure therapy. Behav Cogn Psychother 2013;41(4):505-509. [doi: 10.1017/S1352465813000088] [Medline: 23410820]
30. Wrzesien M, Burkhardt JM, Botella C, Alcañiz M. Evaluation of the quality of collaboration between the client and the therapist in phobia treatments. Interact Comput 2012;24(6):461-471. [doi: 10.1016/j.intcom.2012.09.001]
31. Wrzesien M, Burkhardt JM, Alcañiz M, Botella C. How technology influences the therapeutic process: a comparative field evaluation of augmented reality and in vivo exposure therapy for phobia of small animals C3—lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). 2011 Presented at: 13th IFIP TC 13 International Conference on Human-Computer Interaction, INTERACT 2011; September 5-9, 2011; Lisbon, Portugal. [doi: 10.1007/978-3-642-23774-4_43]
32. Javanbakht A, Madaboosi S, Grasser LR. Real-life contextualization of exposure therapy using augmented reality: a pilot clinical trial of a novel treatment method. Ann Clin Psychiatry 2021;33(4):220-231. [doi: 10.12788/acp.0042] [Medline: 34672925]
33. Zimmer A, Wang N, Ibach MK, Fehlmann B, Schicktanz NS, Bentz D, et al. Effectiveness of a smartphone-based, augmented reality exposure app to reduce fear of spiders in real-life: a randomized controlled trial. J Anxiety Disord 2021;82:102442 [FREE Full text] [doi: 10.1016/j.janxdis.2021.102442] [Medline: 34246153]
34. Miloff A, Lindner P, Dafgård P, Deak S, Garke M, Hamilton W, et al. Automated virtual reality exposure therapy for spider phobia vs. in-vivo one-session treatment: a randomized non-inferiority trial. Behav Res Ther 2019;118:130-140. [doi: 10.1016/j.brat.2019.04.004] [Medline: 31075675]
35. Norberg MM, Newins AR, Jiang Y, Xu J, Forcadell E, Alberich C, et al. The scarier the better: maximizing exposure therapy outcomes for spider fear. Behav Cogn Psychother 2018;46(6):754-760. [doi: 10.1017/S1352465818000437] [Medline: 30001760]
36. Craske MG, Mohlman J, Yi J, Glover D, Valeri S. Treatment of claustrophobias and snake/spider phobias: fear of arousal and fear of context. Behav Res Ther 1995;33(2):197-203. [doi: 10.1016/0005-7967(94)p4441-v] [Medline: 7887879]
37. Shiban Y, Pauli P, Mühlberger A. Effect of multiple context exposure on renewal in spider phobia. Behav Res Ther 2013;51(2):68-74. [doi: 10.1016/j.brat.2012.10.007] [Medline: 23261707]
38. Dunsmoor JE, Ahs F, Zielinski DJ, LaBar KS. Extinction in multiple virtual reality contexts diminishes fear reinstatement in humans. Neurobiol Learn Mem 2014;113:157-164 [FREE Full text] [doi: 10.1016/j.nlm.2014.02.010] [Medline: 24583374]
39. Shiban Y, Schelhorn I, Pauli P, Mühlberger A. Effect of combined multiple contexts and multiple stimuli exposure in spider phobia: a randomized clinical trial in virtual reality. Behav Res Ther 2015;71:45-53. [doi: 10.1016/j.brat.2015.05.014] [Medline: 26072451]
40. Baus O, Bouchard S. Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: a review. Front Hum Neurosci 2014;8:112 [FREE Full text] [doi: 10.3389/fnhum.2014.00112] [Medline: 24624073]
41. Brás S, Soares SC, Cruz T, Magalhães T, Marques B, Dantas C, et al. The feasibility of an augment reality system to study the psychophysiological correlates of fear-related responses. Brain Behav 2018;8(9):e01084 [FREE Full text] [doi: 10.1002/brb3.1084] [Medline: 30136383]

Abbreviations

AR: augmented reality

ARET: augmented reality exposure therapy
BAT: Behavioral Avoidance Test
CASP: Critical Appraisal Skills Programme
CBT: cognitive behavioral therapy
DSM: Diagnostic and Statistical Manual of Mental Disorders
ECG: electrocardiogram
FSQ: Fear of Spiders Questionnaire
ICD: International Classification of Diseases
IVET: in-vivo exposure therapy
PICOS: Population, Intervention, Comparison and Outcomes
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SCR: skin conductance recordings
SPBQ: Spider Phobia Beliefs Questionnaire
SUDS: Subjective Units of Distress Scale
TAU: treatment as usual
VR: virtual reality
VRET: virtual reality exposure therapy
WAI-S: Working Alliance Inventory-Short

Edited by J Torous; submitted 27.07.23; peer-reviewed by M Hirota, M Eckert; comments to author 29.10.23; revised version received 06.11.23; accepted 07.11.23; published 30.11.23.

Please cite as:
Hasan S, Alhaj H, Hassoulas A
The Efficacy and Therapeutic Alliance of Augmented Reality Exposure Therapy in Treating Adults With Phobic Disorders: Systematic Review
JMIR Ment Health 2023;10:e51318
URL: https://mental.jmir.org/2023/1/e51318
doi:10.2196/51318
PMID:38032710

Review

Self-Guided Mental Health Apps Targeting Racial and Ethnic Minority Groups: Scoping Review

Fiby Saad[1], MSc; Mia Eisenstadt[1,2], PhD; Shaun Liverpool[1,2,3], PhD; Courtney Carlsson[4], MBA; Isabella Vainieri[1,2,5], PhD

[1]Clinical, Educational and Health Psychology, Division of Psychology and Language Sciences, University College London, Faculty of Brain Sciences, London, United Kingdom

[2]Evidence Based Practice Unit, Anna Freud National Centre for Children and Families, London, United Kingdom

[3]Department of Social Work & Wellbeing, Edge Hill University, Faculty of Health, Social Care and Medicine, Ormskirk, United Kingdom

[4]Paradym, London, United Kingdom

[5]Department of Psychology, Royal Holloway, University of London, Egham, Surrey, United Kingdom

Corresponding Author:
Isabella Vainieri, PhD
Clinical, Educational and Health Psychology, Division of Psychology and Language Sciences
University College London
Faculty of Brain Sciences
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 44 2039872684 ext 2684
Email: ucjuiva@ucl.ac.uk

Abstract

Background: The use of mental health apps (MHAs) is increasing rapidly. However, little is known about the use of MHAs by racial and ethnic minority groups.

Objective: In this review, we aimed to examine the acceptability and effectiveness of MHAs among racial and ethnic minority groups, describe the purposes of using MHAs, identify the barriers to MHA use in racial and ethnic minority groups, and identify the gaps in the literature.

Methods: A systematic search was conducted on August 25, 2023, using Web of Science, Embase, PsycINFO, PsycArticles, PsycExtra, and MEDLINE. Articles were quality appraised using the Mixed Methods Appraisal Tool, and data were extracted and summarized to form a narrative synthesis.

Results: A total of 15 studies met the inclusion criteria. Studies were primarily conducted in the United States, and the MHAs designed for racial and ethnic minority groups included ¡Aptívate!, iBobbly, AIMhi- Y, BRAVE, Build Your Own Theme Song, Mindful You, Sanadak, and 12 more MHAs used in 1 study. The MHAs were predominantly informed by cognitive behavioral therapy and focused on reducing depressive symptoms. MHAs were considered acceptable for racial and ethnic minority groups; however, engagement rates dropped over time. Only 2 studies quantitatively reported the effectiveness of MHAs among racial and ethnic minority groups. Barriers to use included the repetitiveness of the MHAs, stigma, lack of personalization, and technical issues.

Conclusions: Considering the growing interest in MHAs, the available evidence for MHAs for racial and ethnic minority groups appears limited. Although the acceptability seems consistent, more research is needed to support the effectiveness of MHAs. Future research should also prioritize studies to explore the specific needs of racial and ethnic minority groups if MHAs are to be successfully adopted.

(JMIR Ment Health 2023;10:e48991) doi:10.2196/48991

KEYWORDS

mental health apps; racial and ethnic minority groups; self-guided; mental health; culturally appropriate technology

Introduction

Background

Mental health apps (MHAs) are frequently used as self-guided tools to help people with various mental health conditions, including anxiety [1] and depression [2]. More than 10,000 MHAs are currently available for smartphone users [3], and this number is increasing daily [4] due to a high interest in MHAs among the public [5], which peaked during the COVID-19 pandemic [6]. For instance, the number of MHA downloads increased by 2 million during the COVID-19 pandemic compared with prepandemic levels [6].

Despite the overall increase in downloads over the years, MHAs appear to appeal to certain populations more than others. For instance, people who have had a previous diagnosis of mental illness [7] or those who are more symptomatic [8] may be more likely to download MHAs. Interest in MHAs is especially high among younger generations, with studies reporting that younger participants (aged 18-22 years) were more interested in MHAs than older participants (≥23 years) [7,9]. This increased interest in MHAs among younger generations could be attributed to the incorporation of smartphone technology in their daily lives [5], as well as the increase in mental health conditions among young people [10]. Another reason is self-monitoring and tracking of progress over time, as it can influence an individual's motivation to continue psychological treatment and enhance feelings of control, which is especially important in young people [11].

The COVID-19 pandemic has significantly impacted the mental health of young people. For example, the Opinions and Lifestyle Survey conducted by the Office for National Statistics revealed that the prevalence of anxiety and depression increased by almost 11% between June 2019 and March 2020 in people aged 16 to 39 years compared with prepandemic levels. However, studies have shown that 50% to 80% of young adults who struggle with mental health issues do not seek treatment [12,13] Some experts argue that stigma around mental illness is a key barrier when accessing face-to-face (FTF) therapy, leaving some young people to express a preference for MHAs [14,15]. Overall, younger age and high self-stigma are associated with a low mental health help–seeking attitude and a negative attitude toward FTF therapy [16].

Despite their popularity, MHAs present some challenges for app users. First, there seems to be a high turnover rate of MHAs. Larsen et al [17] found that apps targeted for depression were unavailable to access approximately every 3 days, leading to difficulties for users to commit to one app and see any long-term benefits. Another major issue with MHAs is the level of user engagement because people rarely use MHAs as a long-term solution [18,19]. For instance, studies have shown that the median duration of app use was only 3 hours over an 8-week treatment period [19] and the median retention rate was 5.5 days (across 8 studies) [20].

However, the most important issue with MHAs is the lack of evidence of their effectiveness. A recent review showed that only 2 out of the 73 apps targeting common mental health symptoms provided direct evidence to support the use and effectiveness of their app [17]; this highlights that app developers might use scientific jargons to lure users into using the app despite no evidence supporting their claims. Even apps that are approved by public authorities report little evidence of their effectiveness. Another review found that only 15% of the MHAs in the UK National Health Service library provided evidence of effectiveness [21], highlighting the need for regulations to ensure that MHAs meet specific standards of care [22].

Despite these challenges, there are a range of benefits that have contributed to the rapid growth and popularity of MHAs. First, MHAs can be accessed anywhere and at any given time. By contrast, traditional therapy occurs at set hours or in specific settings. Furthermore, services may have increased waiting times [23], which raises major risks for individuals, such as self-harm or suicide [24]. Second, unlike FTF therapy, MHAs can be used by any number of people. Third, unlike publicly funded therapy, in which an individual requires a diagnosis or a basis for referral, MHAs generally have no requirements or criteria for use. Overall, MHAs can be used outside clinical settings or as adjunct support to help people manage everyday stress [25].

The ability to access mental health aid outside clinical settings can be especially helpful for people from racial and ethnic minority backgrounds. For this review, *racial and ethnic minority group* refers to any racial and ethnic group with national or cultural traditions different from those of the main majority. Evidence shows that people from racial and ethnic minority backgrounds experience higher levels of stressors than the majority population; these stressors such as low socioeconomic status, discrimination, and racism can negatively affect mental health outcomes [26-28]. People from racial and ethnic minority backgrounds also experience increased barriers when engaging with mental health services [29-31] and are less likely to self-report and receive treatment [31]. This is possibly due to personal and environmental barriers such as the inability to recognize and accept mental health problems, embarrassment, confidentiality concerns, preference for self-reliance, social stigma against mental health, and financial factors [32-36]. Other factors are related to health care providers, such as language barriers, cultural naivety, insensitivity, and discrimination toward the needs of racial and ethnic minority service users [36]. Overall, individuals from racial and ethnic minority backgrounds are exposed to increased risk factors for poor mental health and experience inequalities in accessing mental health care.

MHAs can offer opportunities to access mental health support and overcome some of the abovementioned barriers encountered by racial and ethnic minority populations. For instance, MHAs provide a sense of safety to some users, increasing their ability to disclose and share their feelings [37], as they enable access to services from their homes, and more importantly, they avoid the stigma associated with disclosing a mental health problem [38,39]. This is particularly important for racial and ethnic minority populations, as evidence suggests that mental health stigma is higher in people from racial and ethnic minority backgrounds than in the majority population [33]. Furthermore, the consequences of mental health stigma are higher among

racial and ethnic minority populations, as they often experience other social adversities that negatively affect mental health, leading to untreated mental health problems as well as poorer mental health outcomes [33]. Kern et al [7] conducted a survey of college students in the United States to explore their openness, use, and attitudes toward MHAs. Out of 565 respondents, 179 were of racial and ethnic minority background, and they found that participants from this background preferred downloading an MHA instead of going to therapy. Similarly, Lungu and Sun [40] found that Asian American youth endorsed seeking help on the web rather than going to professionals in an FTF setting. Although this is promising, interest does not always correlate with actual use [41]. Furthermore, a recent systematic review of MHAs found that there was an absence of diverse samples, with many studies using majority White populations, whereas the effectiveness, acceptability, and use of MHAs in racial and ethnic minority groups remain poorly understood [42].

Objectives

We conducted a scoping review of the literature to (1) describe the purposes of using MHAs in racial and ethnic minority groups, (2) examine the acceptability of MHAs among those groups, (3) examine the effectiveness of MHAs with these groups, (4) identify the barriers to MHA use within these groups, and (5) identify the gaps in the literature. We will only focus on self-guided MHAs that users can use without additional help (eg, video chat and text messaging), as they offer a more sheltered environment for the user, further removing the issue of stigma [43]. Due to the recent interest in MHAs among young people and the need for a comprehensive overview of the literature focusing on racial and ethnic minority groups, this study covered a wide age range of 14 to 36 years. This age range also captures 3 main age groups that have been found to have high smartphone use:14 to 18 [15,44,45], 18 to 21 [7,43], and 25 to 36 years [46].

Methods

This scoping review was conducted in accordance with the Joanna Briggs Institute methodology for scoping reviews [47] and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [48]. Refer to Multimedia Appendix 1 [49] for the PRISMA-ScR checklist.

Search Strategy

A systematic search was conducted in the following databases: Embase; PsycINFO; PsycArticles; PsycExtra; MEDLINE ALL, via OVID; and Web of Science. See Multimedia Appendix 2 for a complete list of search terms. The search algorithm was defined including concepts related to *mobile phone apps*, *mental health*, and *racial and ethnic minority groups*. The search was conducted on August 25, 2023, with no limit placed on the publication year.

Eligibility Criteria

Studies were included if they fulfilled all the following criteria: (1) most participants were from a racial and ethnic minority background (ie, more than 50%); (2) the study explored "self-guided" MHAs, meaning that the participants used the apps alone without outside help; (3) participants' age range was between 14 and 36 years; (4) the study focused on mental health issues; and (5) the study was written in English. Studies were excluded if they were solely used for adherence to medication or other lifestyle changes such as diet or exercise.

Selection Process

The CADIMA software package (Julius Kühn-Institut) was used to facilitate the review processes, including screening and data extraction [50]. The titles and abstracts were independently screened by 2 reviewers (FS and IV), and those that met our inclusion criteria were used for full-text screening. All the full texts were screened in parallel by the same 2 reviewers. Any inconsistencies between the reviewers were discussed before reaching an agreement.

Data Extraction and Quality Assessment

The extracted data included (1) study design (eg, qualitative, quantitative, or mixed methods); (2) participants' demographic details (eg, age, ethnicity, and occupation); (3) geographic location; (4) the intervention used, including theoretical basis, purpose, and duration of use; (5) data regarding the acceptability of MHAs; (6) data related to the effectiveness of the intervention; and (7) any barriers to MHA use. Acceptability was defined as "a multi-faceted construct that reflects the extent to which people delivering or receiving a health care intervention consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention" [51].

Critical appraisal was conducted following the Mixed Methods Appraisal Tool checklist [52]. The Mixed Methods Appraisal Tool consists of 2 general screening questions and 5 questions for each type of study design. Each question was answered by responding "yes," "no," or "can't tell" and scored 1 for "yes" and 0 for "no," resulting in the maximum score of 7 for each study. Quality assessment was conducted independently by 2 reviewers (FS and IV). Any discrepancies were discussed, and if necessary, a third team member was consulted to reach a final decision.

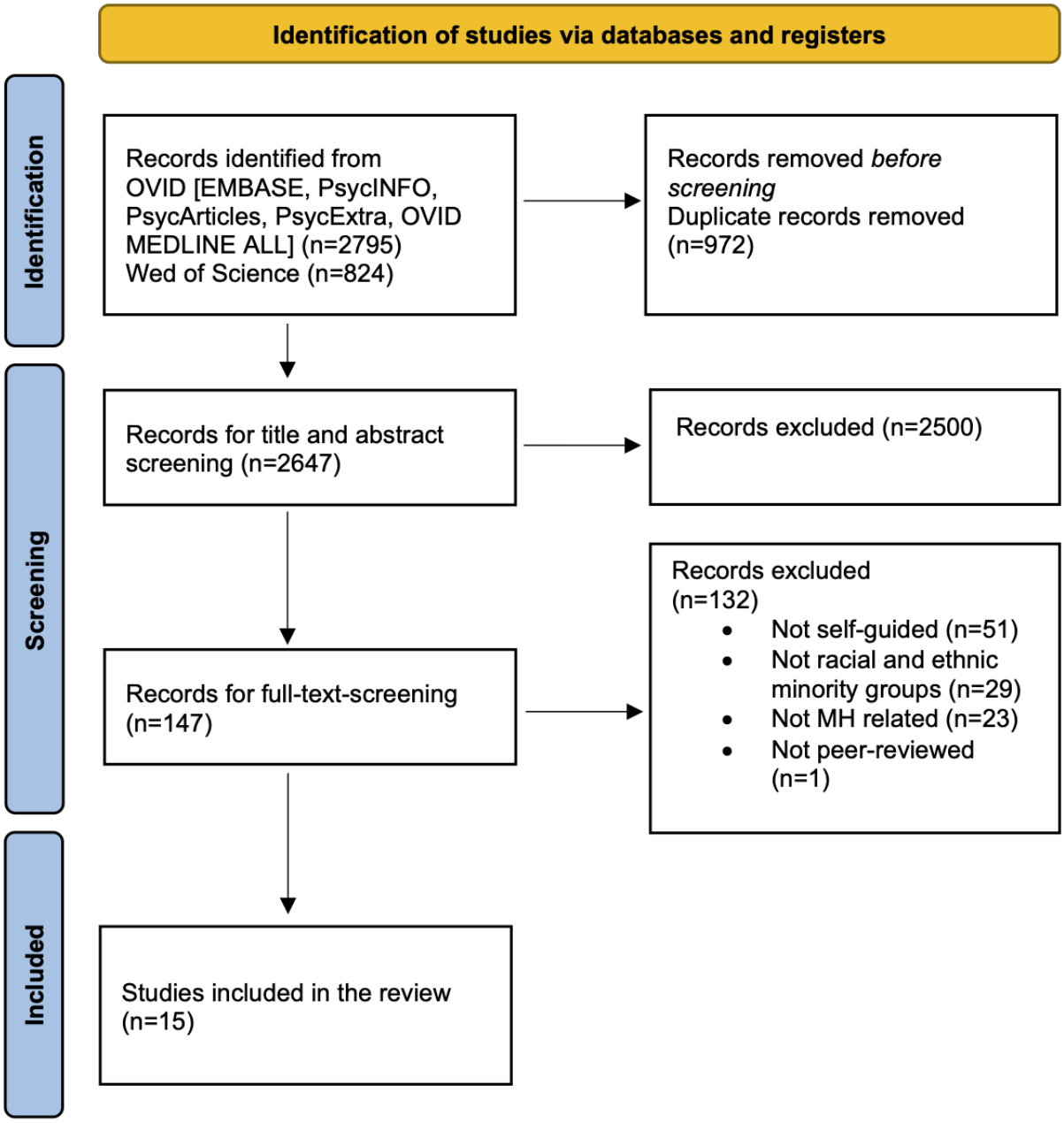
Data Analysis and Data Synthesis

First, the study and its population characteristics were charted to provide an overall description of the body of evidence. Second, a narrative synthesis, supported by thematic and content analysis as outlined by Popay et al [53], was conducted to provide an overall narrative to address the aims of the review.

Results

A total of 15 studies were eligible for inclusion in this scoping review. A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram [49] is shown in Figure 1 to illustrate the flow of information and the identified records at each phase of the scoping review.

Figure 1. Flowchart highlighting the key stages of the screening process. MH: mental health.



Characteristics of Included Studies

Out of the 15 publications that met the inclusion criteria for this review [40,54-67], 10 (67%) studies were conducted in the United States [40,54,56,58-61,63,66,67], with publication dates ranging from 2016 to 2023. Overall, 13 (87%) of the 15 studies focused on specific MHAs [54-61,63-67], including 2 MHA prototypes [58,67]. Of the app-specific studies, 4 had a mixed methods design [55,58,60,63], 7 were quantitative [54,56,57,59,61,65,67] and 2 were qualitative studies [64,66]. The last 2 studies did not focus on a specific MHA; instead, they assessed preference for web-based help versus FTF mental health help using surveys, both were quantitative [40,62]. Table 1 reports the characteristics of the reviewed studies, including study name, study design, sample size, mean age of participants, the racial and ethnic minority group, MHA name (if applicable), targeted mental health condition, duration of use, outcome measure, study location, and critical appraisal score.

Table 1. Summary of included studies.

Study (year)	MHA[a] name	Targeted mental health condition	Duration of use	Study design	Outcome measure	Ethnic minority group	Mean age (years)	Sample size	Geographic location	Critical appraisal score
Dahne et al [54], (2019)	• ¡Aptívate!	• Depression	8 wk	Quantitative	• Spanish language Beck Depression Inventory—II • Semistructured interview	Hispanic	• 36.05	• 42	United States	6
Tighe et al [55], (2020)	• iBobbly	• Depression	6 wk	Mixed methods	• Semistructured interview and RCT[b] data	Aboriginal and Torres Strait Islander	• 24.15	• Interviews (n=18) • RCT (n=61)	Canada	6
Pratap et al [56], (2018)	• iPST[c]	• Depression	12 wk	Quantitative	• PHQ-9[d] • Sheehan Disability Scale	Hispanic or Latino	• 34.90	• Hispanic (n=106) • Non-Hispanic (n=239)	United States	6
Lungu and Sun [40], (2016)	—[e]	—	—	Quantitative	• 18-item web-based survey • MHI-21[f]	Asian American	• 18.7	• 572	United States	5
Tighe et al [57], (2017)	• iBobbly	• Suicidal ideation • depression • psychological distress • impulsivity	6 wk	Quantitative	• DSI-SS[g] • PHQ-9 • K10[h] • BIS-11[i]	Indigenous Australians	• 26.25	• 61	Australia	6
McCall et al [58], (2021)	—	• Anxiety and depression	—	Mixed methods	• QUIS[j] • Tobii (eye-tracking software)	African American	• 29	• 15	United States	6
Rushing et al [59], (2021)	• BRAVE	• Promote help-seeking behavior, mental health, and cultural resilience	8 wk	Quantitative	• None used	American Indian and Alaska Native	• 15-24	• 1030	United States	7
Stephens et al [60], (2020)	• BRAVE	• Promote help-seeking behavior, mental health, and cultural resilience	8 wk	Mixed methods	• None used	American Indian and Alaska Native	• 15-24	• 1030	United States	

Study (year)	MHA[a] name	Targeted mental health condition	Duration of use	Study design	Outcome measure	Ethnic minority group	Mean age (years)	Sample size	Geographic location	Critical appraisal score
Wrobel et al [61], (2022)	• BRAVE	• Promote help-seeking behavior, mental health, and cultural resilience	8 wk	Quantitative	• Mobile Commons tracks message engagement • Questionnaires and surveys	American Indian and Alaska Native	• 15-24	• 1030	United States	7
Maloney et al [62], (2020)	—	—	—	Quantitative	• Surveys	Jamaican	• 10-19	• 107	United Kingdom	6
Neal-Barnett et al [63], (2019)	• BYOTS[k]	• Anxiety and negative thinking	1 wk	Mixed methods	• Focus groups	Black and biracial	• 12-15	• 72	United States	7
Povey et al [64], (2016)	• iBobbly	• Depression	1 wk	Qualitative	• Focus groups	Aboriginal or Torres Strait Islander	—	—	Australia	5
Röhr et al [65], (2021)	• Sanadak	• PTSD[l]	4 wk	Quantitative	• PDS-5[m] • PHQ-9 • GAD-7[n] • PHQ-15 • GSE[o] • SSMIS-SF[p] • SSMIS-AW[q] • SSMIS-AG[r] • SSMIS-AP[s] • SSMIS-HS[t] • RS-13[u] • LSNS-6[v] • ESSI[w] • EQ-5D-5L • EQ-VAS[x] • PGI[y]	Syrian	• Control: 33.67 • Intervention: 32.98	• 133	Germany	6
Agapie et al [66], (2022)	• Mindshift • Sanvello • Woebot • Headspace • Insight Timer • Shine • Smiling Mind • Covid Coach • Daylio • Moodflow • Talk Life		1 wk for each category	Qualitative	• Follow-up survey and Mobile App Rating Scale	Hispanic and Black	• 18	• 5	United States	5

Study (year)	MHA[a] name	Targeted mental health condition	Duration of use	Study design	Outcome measure	Ethnic minority group	Mean age (years)	Sample size	Geographic location	Critical appraisal score
		• CBT[z] (n=4) • Mindfulness (n=4) • Miscellaneous apps (n=4), which did not pertain to 1 category, including a coping app, journaling app, mood-tracking app, and peer support app								
Watson-Singleton [67], (2023)	• Mindful You	• Mindfulness	2 wk	Quantitative	• FFMQ[aa] • MSES[ab] • ATMS[ac] • MBUS[ad] • Mindfulness Knowledge Scale • PSS[ae] • Difficulties in Emotion Regulation Scale	African American	• 31.1	• 39	United States	4

[a]MHA: mental health app.
[b]RCT: randomized controlled trial.
[c]iPST: internet-based problem-solving therapy.
[d]PHQ: Patient Health Questionnaire.
[e]Not available.
[f]MHI-21: Mental Health Inventory.
[g]DSI-SS: Depressive Symptom Inventory–Suicidality Subscale.
[h]K10: The Kessler Psychological Distress Scale.
[i]BIS-11: Barratt Impulsivity Scale.
[j]QUIS: Questionnaire for User Interface Satisfaction.
[k]BYOTS: Build Your Own Theme Song.
[l]PTSD: posttraumatic stress disorder.
[m]PDS: Posttraumatic Diagnostic Scale for DSM-5.
[n]GAD-7: Generalized Anxiety Disorder, 7 items.
[o]GSE: General Self-efficacy.
[p]SSMIS-SF: Self-Stigma of Mental Illness Scale–Short Form.
[q]SSMIS-AW: Self-Stigma of Mental Illness Scale– Stereotype Awareness.
[r]SSMIS-AG: Self-Stigma of Mental Illness Scale–Stereotype Agreement.
[s]SSMIS-AP: Self-Stigma of Mental Illness Scale–Stereotype Application.
[t]SSMIS-HS: Self-Stigma of Mental Illness Scale–Harm to Self-esteem.

[u]RS-13: Resilience Scale.
[v]LSNS-6: Lubben Social Network Scale (social isolation).
[w]ESSI: ENRICHD Social Support Inventory.
[x]VAS: Visual Analog Scale.
[y]PGI: Posttraumatic Growth Inventory.
[z]CBT: cognitive behavioral therapy.
[aa]FFMQ: Five Facet Mindfulness Questionnaire.
[ab]MSES: Mindfulness Self-Efficacy Scale.
[ac]ATMS: Attitudes Toward Mindfulness Scale.
[ad]MBUS: Mindfulness Behavior Usage Scale.
[ae]PSS: Perceived Stress Scale.

Purposes of Using MHAs in Racial and Ethnic Minority Groups

The most common purpose for using MHAs was depression (¡Aptívate!, iPST [internet-based problem-solving therapy], iBobbly, and a prototype app by McCall et al [58]) [54,56-58]. The iBobbly MHA also addressed impulsivity; however, this was not the main purpose of use.

Two studies focused on overall psychological well-being [59,66]. The BRAVE app was used to promote overall mental well-being by including help-seeking behaviors, general mental health, and cultural resilience as outcome measures. Agapie et al [66] included a mix of MHAs, with the aim of measuring their effect on psychological well-being using qualitative methods.

The other apps in this review had various purposes. The Build Your Own Theme Song (BYOTS) app was aimed at reducing anxiety and negative thoughts. The Sanadak app [65] aimed to reduce posttraumatic stress disorder (PTSD) symptoms, whereas Mindful You [67] aimed to reduce stress.

Intervention Characteristics

A total of 20 apps were investigated in this review; 8 MHAs (including 2 prototypes) were included in 14 of the 15 studies [54,56-59,63-67], whereas 1 study [66] included 12 self-help MHAs that were qualitatively investigated.

Of the 20 apps, 8 (40%) were based on cognitive behavioral therapy (CBT) [65,66] or variations of the CBT approach, such as problem-solving therapy [56], acceptance-based therapy [57], and behavioral activation therapy [54]. Five (25%) apps were based on mindfulness [66,67], and the 7 (35%) apps could be described as miscellaneous: the prototype by McCall et al [58], which included the elements of CBT and psychotherapy [58]; the BYOTS app, which is based on musical cognitive restructuring [63]; and the BRAVE app, which is based on offering information and role model videos aimed at providing coping skills [59]. The remaining 4 miscellaneous apps were described as “wellness hacks” by Agapie et al [66]: Covid Coach, Daylio, Moodflow, and Talk Life. The full list of MHAs categorized by therapeutic approach is shown in Textbox 1.

Textbox 1. Therapeutic approach of the 20 mental health apps included in this review.

Cognitive behavioral therapy–based apps

- Sanadak [65]
- Internet-based problem-solving therapy [56]
- ¡Aptívate! [54]
- iBobbly [57]
- Mindshift [66]
- Sanvello [66]
- Woebot [66]
- Wysa [66]

Mindfulness-based apps

- Headspace [66]
- Insight Timer [66]
- Shine [66]
- Smiling Mind [66]
- Mindful You [67]

Miscellaneous apps

- Prototype by McCall et al [58]
- Build Your Own Theme Song [63]
- BRAVE [59]
- Covid Coach [66]
- Daylio [66]
- Moodflow [66]
- Talk Life [66]

Consideration of Racial and Ethnic Minority Groups While Developing MHAs

Of all the MHAs mentioned, 7 apps targeted racial and ethnic minority groups specifically: ¡Aptívate!, iBobbly, BRAVE, BYOTS, the prototype app by McCall et al [58], Sanadak, and Mindful You. The inclusion of racial and ethnic minority groups was ensured by codeveloping the app with the target population, by using workshops [56], interviews [55,59], usability trials [58], or working with culturally informed organizations [63]. ¡Aptívate! [54] was developed in Spanish language to be acceptable to the Hispanic population. Both Sanadak and Mindful You were developed with the specific needs of racial and ethnic minority groups in mind and tailored to the type of material used in the apps.

Examining the Acceptability of MHAs Among Racial and Ethnic Minority Groups

To measure the acceptability of MHAs among racial and ethnic minority groups, 8 studies referred to app use and interactive data [54,56,57,59-61,65,67] in 6 MHAs. Studies reported good adherence to the specified MHAs throughout the set duration period in Hispanic, Aboriginal, Torres Strait Islander, and American Indian and Alaska Native individuals. Adherence data ranged from 81.8% to 91.2% of participants using and interacting with the app.

For the ¡Aptívate! app, participants were asked to use the app within the 8 weeks provided. The retention rate was 100% in the first week but decreased to 50% by the eighth week. This study suggests that the 50% drop in retention can be explained by the local versus remote recruitment of Latina participants. Those who attended baseline visits in person were more likely to use the app more frequently than those who did remotely. Dahne et al [54] also reported that 50% of Hispanic participants who continued using ¡Aptívate! 2 months after enrollment showed a high level of acceptability. Pratap et al [56] also recruited Hispanic and Latina participants and conducted a randomized controlled trial for 3 months to evaluate the iPST app. Engagement and retention rates were assessed based on the number of completed surveys. The study reported 34.4% dropouts in the Hispanic and Latino population. Of those who dropped out, more than half reported making ≤US $20,000 annually. Of those who used the app, Hispanic and Latina participants showed a 50% decrease in engagement from week 1 to week 4. It is important to note that this is based on the completion of the assessment and, therefore, is not an accurate representation of app use.

The BRAVE app was also used in an 8-week trial [60], with an overall retention rate of 87%. Among the participants in the BRAVE arm, 41 American Indian and Alaska Native participants opted out during the intervention and 25 opted out at crossover. This suggests a dropout rate of only 13% [59].

The iBobbly app was used by Aboriginal and Torres Strait Islander participants in a 6-week trial. The app had the lowest dropout rate (3%) of all other MHAs in this review. The study argues that this was due to technical issues and speculated that some participants might have felt self-conscious about sharing their use data [57].

Sanadak was designed for Syrian participants, and they were asked to use the app regularly for 4 weeks. The retention rate was 87.2%, with a dropout rate of 12.8%, where most participants refused to continue. Upon further analysis, Röhr et al [65] claimed that there was no significant difference between participants who completed the study and those who did not. Finally, Watson-Singleton et al [67] explored the Mindful You app. African American participants were asked to use the app for 2 weeks. The study reported a dropout rate of 45%, which the study found difficult to explain because the app was designed specifically for African Americans. Participants who continued using the app felt positive about Mindful You, giving the app 4.38 stars out of 5.

Overall, dropout rates were significantly higher among Hispanic and Latino participants than among non-Hispanic participants, with the latter staying on average 18.5 days longer.

Qualitative studies measured the acceptability of MHAs using interviews [55], workshops and focus groups [63], surveys [40,62], and questionnaires [58].

Participants were interviewed about the iBobbly app in terms of acceptability, cultural appropriateness, and whether the app provided help with their feelings and created distractions. The Aboriginal and Torres Strait Islander participants reported that iBobbly was acceptable, especially in terms of accessibility. Moreover, the participants felt a sense of privacy that was valued more than talking with a therapist or a family member. Participants also spoke of the "shame" attached to young Aboriginal people when asking for help, and so the iBobbly app was seen as culturally appropriate. Povey et al [64] also explored the iBobbly app and compared it with a therapist-led app. Aboriginal and Torres Strait Islander participants showed enthusiasm when they helped design the AIMhi- Y app [43]. Barnett et al [63] conducted focus groups regarding the BYOTS app designed for Black and biracial girls, and they were prompted to use the app 3 times a day for 1 week. This study did not mention dropout rates. However, using focus groups, the study showed that Black and biracial girls found the BYOTS app acceptable and useful in daily situations.

Agapie et al [66] explored a variety of apps with Hispanic and Latina participants. They were asked to use 1 app from each category weekly for a 5-week period and then complete a focus group at the end of each week. Every week participants would use a different app and then the last week they chose their favorite. There were no official use data; participants were asked to report how often they used the app, and they were most likely to report "a few days a week." It seems that mindfulness apps were more acceptable, with 60% of participants reporting continued use. During the focus groups, participants noted that the apps were easy to use and accessible. However, the participants generally preferred to use the apps with more free content. Some participants reported concern about whether the app was validated by professionals and expressed the need to feel safe.

Both Lungu and Sun [40] and Maloney et al [62] used a questionnaire to assess the acceptability of MHAs in general. Of the 75.3% Asian American young adults who endorsed seeking mental health help on the web, only 22% were interested in MHAs [40]. Asian American participants were more likely to be in the "No therapy" and "Online only" groups compared with White participants. Similarly, 56% of the Jamaican participants were interested in using MHAs. However, shame, stigma, and embarrassment were reported to be the major barriers to seeking help. However, using a questionnaire, McCall et al [58] found that African American women reported that the prototype app was easy to use and provided culturally helpful information for anxiety and depression.

Overall, the apps were acceptable both quantitatively through use data and qualitatively, as participants described their engagement with and enthusiasm for the apps. However, dropout rates among some racial and ethnic minority groups remain high, and there is some discrepancy in the measurement of acceptability.

Examining the Effectiveness of MHAs With Racial and Ethnic Minority Groups

Outcome Measures and Study Design

Nine of the 15 studies included in this review were quantitative [40,54,56,57,59,61,62,65,67], 4 used mixed methods [55,58,60,63], and 2 were qualitative [64,66].

In terms of quantitative studies, 3 studies [54,56,57] assessed depression levels in Hispanic and Indigenous Australian individuals. Two studies assessed mental health resilience in American Indian and Alaska Native populations [59,61]. One study assessed the levels of PTSD in Syrian refugees [65], and the other focused on mindfulness in Black African Americans [67]. The last 2 measured the receptiveness of web-based mental health support and MHAs with Jamaican [62] and Asian American [40] participants.

Four studies adopted mixed methods designs [55,58,60,63] and assessed Aboriginal and Torres Strait Islander, Black and biracial, and Indigenous Australian population. Three of these studies explored named MHAs: BRAVE [60], iBobbly [55], and BYOTS [63]. The remaining was an unnamed app, and the study explored its usability [58]. All studies, except for the one by Stephens et al [60], used surveys [55,58,62,63] and focus groups [63] or interviews [55,56] or cognitive walkthrough and think-aloud methods [58]. Stephens et al [60] did not use any measures, as they reported lessons learned from recruiting and engaging participants from the previous BRAVE study [59]. Outcome measures that were used by Tighe et al [55] and Povey et al [64] were appropriately translated to and validated in other

languages to suit the ethnicity of the sample. Surveys, workshops, and interviews were developed and approved by mental health professionals of the target ethnicity.

Finally, the 2 qualitative studies focused on measuring the acceptability of the respective MHAs and discussing barriers to continued use [64,66]. All the outcome measures are presented in Table 1.

Effectiveness

Studies that explored a specific app assessed its effectiveness by using either weekly assessments [54,56,65] or pre- and postintervention changes in outcome measure scores [57,59,61,63,67]. Outcome measures were divided into clinical outcomes (eg, depression, anxiety, and suicidality) and other behavioral outcomes (eg, distress, resilience, and self-efficacy).

Quantitative studies on clinical outcomes that measured effectiveness using weekly assessments had inconclusive results. The ¡Aptívate! app [54] reported significantly lower depressive symptoms in Hispanic adults than in the no-treatment group; however, depressive symptoms did not differ on average across time between the 2 groups. Pratap et al [56] found improvement in depression scores among Hispanic and non-Hispanic participants, regardless of the treatment arm and ethnicity. However, they noted no evidence of any clinically meaningful changes between the iPST and the control group. The authors noted that only participants who reported severe depressive symptoms showed the greatest decline; however, this only lasted until week 4 of the study. Tighe et al [57] reported a decline in depressive symptoms among Aboriginal and Torres Strait Islander participants, but no significant reduction was observed in the primary outcome of suicidality in Indigenous Australian participants. In addition, no significant relationship between use time and any of the outcome measures was observed. The Sanadak app [65] also showed no significant differences in PTSD symptoms between the intervention and control groups after 4 weeks and 4 months of follow-up.

Other behavioral outcomes were also explored by quantitative studies. For instance, Pratap et al [56] explored functional impairment in addition to depressive symptoms and found no difference in disability outcomes across treatment arms and no difference between Hispanic and non-Hispanic participants. Tighe et al [57] explored psychological distress and impulsivity as secondary outcomes. The iBobbly app was associated with a significant decrease in Kessler Psychological Distress Scale scores after 6 weeks; however, there was no significant change in impulsivity [57]. Rushing et al [59] included the following secondary outcomes: self-efficacy, self-esteem, resilience, coping strategies, substance use, and cultural identity. They found that American Indian and Alaska Native participants who reported better health on average at baseline were more likely to report stronger cultural identity, cultural resilience, and positive coping strategies. No significant differences emerged in any of the primary outcomes of the BRAVE app (help seeking, self-efficacy related to mental health, and negative coping) [59,61]. A surprising finding for the BRAVE app was that higher scores on help-seeking attitude at baseline were associated with a decrease in the number of clicks or engagement with the app. However, it is important to note that Wrobel et al [61] reported that the engagement data were highly skewed, with some participants clicking an average of 3.4 times, but some users clicked 49 times. Finally, Röhr et al [65] included the secondary outcomes: self-efficacy, self-stigma, and resilience. They found that after using the Sanadak app, Syrian refugees showed no differences in any of the secondary outcomes, except for self-stigma. Syrian refugees reported lower levels of self-stigma following the use of the Sanadak app.

Quantitative evidence from mixed methods studies also showed inconclusive results on both clinical and behavioral outcomes. Neal-Barnett et al [63] concluded that Black and biracial girls who used the BYOTS app reported significantly lower negative and anxious thoughts on day 7 than on day 1. Although this study showed a positive result, the app was used for only 1 week, so there is still uncertainty regarding whether these improvements would last. Watson-Singleton et al [67] reported that Black African American participants who used Mindful You showed a significant decrease in stress levels after 2 weeks. They also showed increased capacity for emotional regulation and a significant increase in self-efficacy and mindfulness behaviors. However, there were no significant differences in the endorsements of mindfulness attributes, attitudes, or knowledge.

Finally, of the 2 qualitative studies, the one by Agapie et al [66] used a focus group to ask about the perceived effectiveness of the different apps that the participants used. Hispanic and Latina participants reported that all the apps used had small positive impacts on their mental health. Miscellaneous apps were ranked as the most effective in improving mental health well-being, followed by CBT apps and mindfulness apps. Povey et al [64] focused only on acceptability, whereas McCall et al [58] explored usability rather than the effectiveness of the app.

Barriers to MHA Use Within Racial and Ethnic Minority Groups

Several barriers to MHAs were reported by the studies, ranging from cost to cultural appropriateness. Four studies did not explicitly report any barriers; however, they did highlight that not all clients may respond to self-guided treatment [54,58,63,65].

One of the most common barriers to using MHAs was the lack of personal touch. Aboriginal and Torres Strait Islanders, who used the iBobbly app, reported the need for more cultural content that related to their community [64]. This barrier was also true for American Indian and Alaska Native participants who used the BRAVE app [59]. Rushing et al [59] reported that due to the lack of representation in the media, participants reacted positively to both study arms, as they both contained cultural content. Participants who used the iBobbly app reported that such apps were not given enough community awareness and were therefore less likely to be used [64]. Rushing et al [59] also found that those with higher help-seeking tendencies were less likely to use the BRAVE app, which they hypothesized was because they were more likely to have support from people around them. The need for a personal touch was common even across the multiple apps explored by Agapie et al [66], with Hispanic and Latina participants reporting that the content was not specific enough for them.

The second most common barrier was stigma. Islander participants who used the iBobbly app reported that others may not engage with the app due to stigma surrounding mental health [55]. Pratap et al [56] also noted similar concerns among Hispanic participants when using the iPST app. Jamaican participants have gone as far as to describe using MHAs as embarrassing, relating to the stigma attached to receiving mental health support [62].

The third barrier was the cost. This was not so common but was mentioned by both Islander and Hispanic participants [64,66]. Agapie et al [66] found that Hispanic participants were more likely to use apps that were richer in free content compared with those that required a subscription. Other barriers included the repetitiveness of the MHA that was described by Rushing et al [59] as "message fatigue," as lack of engagement was evident after the 10th text sent by the BRAVE app. Furthermore, literacy and language barriers were brought up by Islander participants who argued that some people in their community may not be comfortable using English [64]. Finally, technical issues were also identified as barriers to using MHAs. For instance, Tighe et al [55] failed to gather use data for 21 out of the 61 participants due to internet connectivity issues, a technical problem with their device, or an uncharged battery. Stephens et al [60] also noted that some participants lost access to their mobile phones and were thus unable to interact with the content of the BRAVE app.

Discussion

Principal Findings

This scoping review aimed to (1) describe the purposes of using MHAs in racial and ethnic minority groups, (2) examine the acceptability of MHAs among those groups, (3) examine the effectiveness of MHAs with the groups, (4) identify the barriers to MHA use within the groups, and (5) identify the gaps in the literature. Overall, our research pooled findings from 15 publications and highlighted important findings regarding the evidence related to MHA use among the racial and ethnic minority groups. Overall, MHAs were used for different purposes such as improving depression, decreasing psychological distress, increasing cultural resilience, and promoting help-seeking behavior. Fundamentally, most MHAs targeting racial and ethnic minority groups are underpinned by CBT and focus on depressive symptoms. In terms of acceptability, MHAs appear to be of interest among racial and ethnic minority groups; however, there is limited and mixed evidence of their effectiveness. Barriers to use include intervention-specific characteristics (eg, repetitiveness of the tasks), user-specific characteristics (eg, stigma), and technology-specific characteristics (eg, internet connectivity). Finally, several gaps in the literature, namely, the participant pool, MHAs design, study design, and study location, were identified. Taken together, these findings need to be considered to deepen our knowledge of MHA use and experiences among racial and ethnic minority groups.

Regarding evidence based on the intended purpose of using MHAs in racial and ethnic minority groups, most of the apps included in our review focused on depression and psychological distress. Although depression is one of the most common mental health disorders with a high prevalence among young people [68], the fact that it is one of the main purposes of the use of MHAs in this population is relevant. For instance, people from racial and ethnic minority backgrounds experience increased challenges compared with the majority population including social inequities, discrimination, and disparities in living conditions and work environments that may increase the risk of developing depression and psychological distress [26-28]. Islander participants who used the iBobbly app noted the need of more cultural content that was specific for them to increase engagement with the MHA [64]. However, Watson-Singleton et al [67] reported a 45% dropout rate from Mindful You despite having created a culturally specific app for Black African American population. Thus, there seems to be uncertainty about what specific changes would keep racial and ethnic minority groups engaged in MHAs. Another common challenge experienced by racial and ethnic minority groups is the stigma against mental health, which can form a barrier to accessing mental health support [33-35]. Stigma was a common barrier to using MHAs across Islander, Hispanic, and Jamaican participants [56,62,64]. Only one app in this review addressed this issue and focused on improving help-seeking behaviors [59]. However, the app reported no significant improvement in help-seeking behavior. In contrast, Röhr et al [65] found that a secondary outcome of the Sanadak app was reduced self-stigma in Syrian refugees. Therefore, future studies should aim to improve help-seeking behaviors and reduce mental health stigma in people from racial and ethnic minority backgrounds.

This review noted two critical observations regarding acceptability: (1) the measurements used and (2) attitudes of racial and ethnic minority groups toward MHAs. We followed the definition by Sekhon et al [52] for measuring acceptability: the willingness to participate and the adherence to the MHA. In our review, 8 studies measured acceptability using use or interactive data [54,56,57,59-61,64,67], and 7 used qualitative methods such as interviews, workshops, and surveys [40,55,58,62-64,66]. How acceptability is measured in these studies is essential, as it can affect how an MHA is perceived. For instance, in our review, the iBobbly app was investigated using both use data [55] and qualitative methods [57]. The findings showed that iBobbly was not acceptable in terms of use data; however, qualitative evidence showed that Aboriginal and Torres Strait Islanders reported that the iBobbly app was acceptable and culturally appropriate, and it reduced stigma surrounding mental health issues. Our review further highlights the heterogeneity in the definition and measurement of acceptability, making it difficult to draw conclusions.

Second, there seems to be ambivalence around racial and ethnic minority groups in terms of the acceptability of MHAs. Hispanic and Latina participants showed a high willingness to use MHAs [54,56] but showed a lack of engagement and high dropout rates [54,56]. In contrast, Agapie et al [66] found that 60% of Hispanic participants used the mindfulness apps even after the trial. Similarly, this review shows how many among Black and African American participants find MHAs useful and acceptable [62,63,67]. However, in the study by Watson-Singleton et al [67], almost half of the Black American participants dropped

out for no given reason. In the study by Maloney et al [62] Jamaican participants explained stigma and embarrassment as major barriers to use. Similarly, the BRAVE app was found to be helpful for American Indian and Alaska Native participants; however, upon closer examination of the interactive data, Wrobel et al [61] found that engagement was lower than expected. The study that used the iBobbly app showed that Aboriginal and Torres Strait Islander participants were highly willing to use MHAs [55]; however, there was still a lack of interest in MHAs among them [55]. The other 2 racial and ethnic minority groups included in this study were Syrian refugees and Asian American individuals, who both showed high interest in MHAs and high dropout rates [66] or would rather use Facebook [40].

Overall, despite the high willingness of racial and ethnic minority groups to use MHAs, evidence reports an overall mixed view of engagement. More research adopting appropriate and standardized methods for measuring acceptability should be considered in the future.

Of note, 2 user-specific factors are related to the acceptability of MHAs. First is the level of psychological distress among racial and ethnic minority groups. In our review, we observed that Indigenous Australians with higher levels of distress were more likely to use MHAs and adhere to them [57]. However, in qualitative studies, Indigenous Australians and Aboriginal and Torres Strait Islanders reported that in extreme distress, they might not benefit from MHAs and FTF therapy would be more appropriate [55,64].

The second factor is help-seeking behavior as shown in the BRAVE study. Stephens et al [60] and Wrobel et al [61] found that surprisingly, those who scored high on help-seeking behavior showed less engagement with the BRAVE app. It was suggested that these participants might already have had their own ways to deal with distress and, therefore, were less likely to use other methods such as the BRAVE app. This corroborates the findings of Lungu and Sun [40], who suggest that some people from ethnic minority groups prefer to seek other forms of support (eg, Facebook). Facebook is not a MHA; however, some participants were more comfortable to reveal information on Facebook than attend FTF therapy. It would be helpful to understand the ways in which participants adapted to seek help, as it seems to influence engagement.

Regarding the effectiveness of the apps, the review found only 2 quantitatively effective apps: BYOTS and Mindful You [63,67], which corroborates the findings from previous reviews that reported limited or mixed evidence of the effectiveness of MHAs [17,21,69]. Previous reviews revealed that most MHAs claim effectiveness; however, there is no scientific evidence supporting their claims. This highlights the dire need for regulations on MHAs that are available on app stores. The BYOTS and Mindful You apps were also trialed for only 1 and 2 weeks, respectively; therefore, we cannot confidently assume that they will be effective for longer periods [63]. However, the fact that these MHAs were effective in reducing negative and anxious thoughts in Black and African Americans is in line with previous literature that showed that when given access to treatment, Black Americans benefit and engage more from therapy than White Americans [70]. Qualitative data found that Aboriginal youth in the study by Tighe et al [55] reported enjoying the iBobbly app even if it did not improve their clinical symptoms. Similarly, the participants who used the BRAVE app showed a significant positive improvement, but it was not different from those who received science, technology, engineering, and mathematics (STEM) messages [59]. Moreover, Sanadak app, which was used by Röhr et al [65], did not significantly decrease PTSD symptoms; however, participants' self-stigma toward mental health was notably reduced. Overall, little is known about how users interact with MHAs in clinically meaningful ways.

In terms of app-specific factors that affect effectiveness, evidence shows that users prefer using mobile apps in short bursts of time [19] highlighting that long-term use might result in repetition for app users. A possible solution might be to have users regularly engage with the app to improve its benefits [22]. Stephens et al [60] suggested creating a "pause" in the BRAVE messages so that users continue to be engaged and avoid "text fatigue." Future studies should investigate the features that can encourage engagement among MHAs users. This review highlights the importance of co-design approaches and cultural adaptations. Ramos et al [71] noted that culturally inspired MHAs may be more appealing to racial and ethnic minority groups and can lead to increased intervention uptake. All the MHAs included in this study, except ¡Aptívate! [54] and iPST [56], were designed with the guidance of racial and ethnic minority groups. The inclusion of these groups in the process of creating the app prevents stereotyping and ensures the most culturally relevant factors to the user [71]. ¡Aptívate! [54] and iPST [56] only included accessible language as a culturally adaptive factor in apps. This is in line with the review by Ramos et al [71], who found that almost 58% of the MHAs included only 1 criterion, suggesting that the inclusion of culturally relevant criteria is far from the norm. Our review also showed that a common barrier was that there was not enough cultural content, even for apps specifically designed for racial and ethnic minority groups [64]. Future studies should consider the impact of cultural factors on the effectiveness of MHAs. Furthermore, future studies would benefit from exploring these factors from a qualitative perspective for more insights, as this review shows only 1 effective MHA despite including many culturally adaptive factors.

Observations and Gaps in the Field With Suggestions for Future Research

Four important observations emerged around the potential gaps in the literature: participant pool, MHA design, study design, and location of the study.

Participant Pool

The participants recruited in the studies that we reviewed were primarily Hispanic and Latina or Black and biracial. Therefore, there is a need to recruit participants in MHA research from a wider racial and ethnic minority background. Another important observation in the review is that only 2 studies [54,64] included a greater number of unemployed than employed participants. The remaining studies included either employed participants or those who attended colleges or schools. Future research should

also include low-income racial and ethnic minority populations to help find ways to effectively incorporate MHA technology as an accessible mental health support tool.

MHA Design

Regarding the MHA design, the apps in this review were primarily CBT based [54,64] or inspired by it using acceptance-based therapy [55,57], problem-solving therapy [56], cognitive musical restructuring [63], or a mixture of psychotherapy and CBT [58,59,66]. Therefore, more research is needed to explore different theoretical underpinnings to identify what works for whom, in what context, and among different cultures.

Study Design

Most of the included studies were quantitative, which arguably did not explain why the apps were ineffective. Strategies such as interviews or workshops might help better explore the barriers experienced by participants and help tailor targeted interventions. For instance, studies with mixed methods design offered valuable insights into the strengths and barriers of MHAs [55,58,60,63]. Although qualitative research takes time, future MHA research should consider qualitative research as the beneficial next step to progress in the field of MHAs for racial and ethnic minority populations.

Study Location

A total of 10 studies were conducted in the United States [40,54,56,58-61,63,66,67], 2 in Australia [57,64], 1 in Canada [55], 1 in the United Kingdom [62],1 in Germany [65]. Overall, more studies are needed globally to achieve generalizability of the findings and improve our understanding of MHA use among people of racial and ethnic minorities. More research is needed to explore whether MHA might be incorporated into existing services as a source of additional support to help overcome some of the existing barriers to service receipt among racial and ethnic minority groups.

Limitations

This review benefited from independent screening by 2 researchers, and this minimized selection bias. Similarly, 2 reviewers were involved in quality appraisal, thereby reducing any bias in the assessments. However, this study has some limitations. Although our search terms were guided by previous systematic reviews including racial and ethnic minority groups, this is not an extensive list of all terminology related to racial and ethnic minority groups; therefore, the review was limited to the search terms used. Moreover, as the researchers involved could only read English, several studies that may have been relevant to this review were excluded. However, despite not imposing limitations on the country of origin and an extensive list of racial and ethnic minority group–related search terms, we were only able to include 15 studies; this demonstrates a dearth of evidence of MHAs among racial and ethnic minority groups, which highlights the need for further investigation.

Conclusions

In this review, we aimed to explore the use of MHAs among racial and ethnic minority groups. This review synthesized data from 15 publications and reviewed 7 interventions. Although acceptability seems fairly consistent, more research is needed to support MHA effectiveness and overcome existing barriers. Overall, the literature on MHAs among racial and ethnic minority groups is still scarce, and there is still much left to understand. Future app developers should consider including racial and ethnic minority groups' input in the development of MHAs as well as widening the scope of MHAs to focus on a range of disorders and use different theoretical approaches.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist for scoping reviews.
[PDF File (Adobe PDF File), 516 KB - mental_v10i1e48991_app1.pdf]

Multimedia Appendix 2
Scoping review search terms.
[DOCX File , 28 KB - mental_v10i1e48991_app2.docx]

References

1. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: There is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
2. Gould CE, Kok BC, Ma VK, Zapata AM, Owen JE, Kuhn E. Veterans Affairs and the Department of Defense mental health apps: a systematic literature review. Psychol Serv 2019 May;16(2):196-207. [doi: 10.1037/ser0000289] [Medline: 30431306]
3. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 May 01;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
4. Torous J, Wisniewski H, Liu G, Keshavan M. Mental health mobile phone app usage, concerns, and benefits among psychiatric outpatients: comparative survey study. JMIR Ment Health 2018 Nov 16;5(4):e11715 [FREE Full text] [doi: 10.2196/11715] [Medline: 30446484]

5. Bauer M, Glenn T, Geddes J, Gitlin M, Grof P, Kessing LV, et al. Smartphones in mental health: a critical review of background issues, current status and future concerns. Int J Bipolar Disord 2020 Jan 10;8(1):2 [FREE Full text] [doi: 10.1186/s40345-019-0164-x] [Medline: 31919635]
6. Perez S. Meditation and mindfulness apps continue their surge amid pandemic. TechCrunch. 2020 May 28. URL: https://techcrunch.com/2020/05/28/meditation-and-mindfulness-apps-continue-their-surge-amid-pandemic/ [accessed 2023-11-28]
7. Kern A, Hong V, Song J, Lipson SK, Eisenberg D. Mental health apps in a college setting: openness, usage, and attitudes. Mhealth 2018 Jun 30;4:20 [FREE Full text] [doi: 10.21037/mhealth.2018.06.01] [Medline: 30050916]
8. Lipschitz JM, Connolly SL, Miller CJ, Hogan TP, Simon SR, Burdick KE. Patient interest in mental health mobile app interventions: demographic and symptom-level differences. J Affect Disord 2020 Feb 15;263:216-220. [doi: 10.1016/j.jad.2019.11.083] [Medline: 31818779]
9. Temkin AB, Schild J, Falk A, Bennett SM. Mobile apps for youth anxiety disorders: a review of the evidence and forecast of future innovations. Prof Psychol Res Pract 2020 Aug;51(4):400-413. [doi: 10.1037/pro0000342]
10. Oswalt SB, Lederer AM, Chestnut-Steich K, Day C, Halbritter A, Ortiz D. Trends in college students' mental health diagnoses and utilization of services, 2009-2015. J Am Coll Health 2020 Jan;68(1):41-51. [doi: 10.1080/07448481.2018.1515748] [Medline: 30355071]
11. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):89-103. [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
12. Eisenberg D, Nicklett EJ, Roeder K, Kirz NE. Eating disorder symptoms among college students: prevalence, persistence, correlates, and treatment-seeking. J Am Coll Health 2011 Nov;59(8):700-707 [FREE Full text] [doi: 10.1080/07448481.2010.546461] [Medline: 21950250]
13. Gruttadaro D, Crudo D. College students speak: a survey report on mental health. National Alliance on Mental Illness. 2012. URL: https://www.nami.org/Support-Education/Publications-Reports/Survey-Reports/College-Students-Speak_A-Survey-Report-on-Mental-H [accessed 2023-11-28]
14. Kosyluk KA, Conner KO, Al-Khouja M, Bink A, Buchholz B, Ellefson S, et al. Factors predicting help seeking for mental illness among college students. J Ment Health 2021 Jun;30(3):300-307 [FREE Full text] [doi: 10.1080/09638237.2020.1739245] [Medline: 32191145]
15. Ross SG, Bruggeman B, Maldonado M, Deiling M. Examining personal, perceived, treatment, and self-stigma in college students: the role of parent beliefs and mental health literacy. J College Stud Psychother 2019 Mar 01;34(3):183-197. [doi: 10.1080/87568225.2019.1580657]
16. Ibrahim N, Amit N, Shahar S, Wee LH, Ismail R, Khairuddin R, et al. Do depression literacy, mental illness beliefs and stigma influence mental health help-seeking attitude? A cross-sectional study of secondary school and university students from B40 households in Malaysia. BMC Public Health 2019 Jun 13;19(Suppl 4):544 [FREE Full text] [doi: 10.1186/s12889-019-6862-6] [Medline: 31196033]
17. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019 Mar 22;2(1):18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
18. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
19. Zhang R, Nicholas J, Knapp AA, Graham AK, Gray E, Kwasny MJ, et al. Clinically meaningful use of mental health apps and its effects on depression: mixed methods study. J Med Internet Res 2019 Dec 20;21(12):e15644 [FREE Full text] [doi: 10.2196/15644] [Medline: 31859682]
20. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3(1):21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
21. Leigh S, Flatt S. App-based psychological interventions: friend or foe? Evid Based Ment Health 2015 Nov 12;18(4):97-99. [doi: 10.1136/eb-2015-102203] [Medline: 26459466]
22. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
23. Beck A, Burdett M, Lewis H. The association between waiting for psychological therapy and therapy outcomes as measured by the CORE-OM. Br J Clin Psychol 2015 Jun 26;54(2):233-248. [doi: 10.1111/bjc.12072] [Medline: 25425180]
24. Bruen AJ, Wall A, Haines-Delmont A, Perkins E. Exploring suicidal ideation using an innovative mobile app-strength within me: the usability and acceptability of setting up a trial involving mobile technology and mental health service users. JMIR Ment Health 2020 Sep 28;7(9):e18407 [FREE Full text] [doi: 10.2196/18407] [Medline: 32985995]
25. Ly KH, Asplund K, Andersson G. Stress management for middle managers via an acceptance and commitment-based smartphone application: a randomized controlled trial. Internet Interv 2014 Jul;1(3):95-101 [FREE Full text] [doi: 10.1016/j.invent.2014.06.003]

26. Paradies Y, Ben J, Denson N, Elias A, Priest N, Pieterse A, et al. Racism as a determinant of health: a systematic review and meta-analysis. PLoS One 2015 Sep 23;10(9):e0138511 [FREE Full text] [doi: 10.1371/journal.pone.0138511] [Medline: 26398658]
27. Williams DR. Stress and the mental health of populations of color: advancing our understanding of race-related stressors. J Health Soc Behav 2018 Dec 28;59(4):466-485 [FREE Full text] [doi: 10.1177/0022146518814251] [Medline: 30484715]
28. Mossakowski KN. Dissecting the influence of race, ethnicity, and socioeconomic status on mental health in young adulthood. Res Aging 2008 Sep 10;30(6):649-671. [doi: 10.1177/0164027508322693]
29. Tribe R. Mental health of refugees and asylum-seekers. Adv Psychiatr Treat 2018 Jan 02;8(4):240-247 [FREE Full text] [doi: 10.1192/apt.8.4.240]
30. Fazel M, Wheeler J, Danesh J. Prevalence of serious mental disorder in 7000 refugees resettled in western countries: a systematic review. Lancet 2005;365(9467):1309-1314. [doi: 10.1016/S0140-6736(05)61027-6] [Medline: 15823380]
31. Harwood H, Rhead R, Chui Z, Bakolis I, Connor L, Gazard B, et al. Variations by ethnicity in referral and treatment pathways for IAPT service users in South London. Psychol Med 2023 Feb;53(3):1084-1095 [FREE Full text] [doi: 10.1017/S0033291721002518] [Medline: 34334151]
32. Alvidrez J, Snowden LR, Kaiser DM. The experience of stigma among Black mental health consumers. J Health Care Poor Underserved 2008 Aug;19(3):874-893. [doi: 10.1353/hpu.0.0058] [Medline: 18677076]
33. Eylem O, de Wit L, van Straten A, Steubl L, Melissourgaki Z, Danışman GT, et al. Stigma for common mental disorders in racial minorities and majorities a systematic review and meta-analysis. BMC Public Health 2020 Jun 08;20(1):879 [FREE Full text] [doi: 10.1186/s12889-020-08964-3] [Medline: 32513215]
34. Brown A, Rice SM, Rickwood DJ, Parker AG. Systematic review of barriers and facilitators to accessing and engaging with mental health care among at-risk young people. Asia Pac Psychiatry 2016 Mar;8(1):3-22. [doi: 10.1111/appy.12199] [Medline: 26238088]
35. Cook BL, Trinh NH, Li Z, Hou SS, Progovac AM. Trends in racial-ethnic disparities in access to mental health care, 2004-2012. Psychiatr Serv 2017 Jan 01;68(1):9-16 [FREE Full text] [doi: 10.1176/appi.ps.201500453] [Medline: 27476805]
36. Memon A, Taylor K, Mohebati LM, Sundin J, Cooper M, Scanlon T, et al. Perceived barriers to accessing mental health services among Black and minority ethnic (BME) communities: a qualitative study in Southeast England. BMJ Open 2016 Nov 16;6(11):e012337 [FREE Full text] [doi: 10.1136/bmjopen-2016-012337] [Medline: 27852712]
37. Ardi Z, Sukmawati IS, Ifdil I, Afdal A, Rangka IB, Suranata K. Exploring the acceptability of internet-based mental health mobile app services using network psychometrics analysis. J Phys Conf Ser 2018 Dec 07;1114:012106 [FREE Full text] [doi: 10.1088/1742-6596/1114/1/012106]
38. Fraser HS, Blaya J. Implementing medical information systems in developing countries, what works and what doesn't. AMIA Annu Symp Proc 2010 Nov 13;2010:232-236 [FREE Full text] [Medline: 21346975]
39. Pagalday-Olivares P, Sjöqvist BA, Adjordor-van de Beek J, Abudey S, Silberberg AR, Buendia R. Exploring the feasibility of eHealth solutions to decrease delays in maternal healthcare in remote communities of Ghana. BMC Med Inform Decis Mak 2017 Dec 02;17(1):156 [FREE Full text] [doi: 10.1186/s12911-017-0552-z] [Medline: 29197391]
40. Lungu A, Sun M. Time for a change: college students' preference for technology-mediated versus face-to-face help for emotional distress. Telemed J E Health 2016 Dec;22(12):991-1000. [doi: 10.1089/tmj.2015.0214] [Medline: 27223266]
41. Torous J, Vaidyam A. Multiple uses of app instead of using multiple apps – a case for rethinking the digital health technology toolbox. Epidemiol Psychiatr Sci 2020 Jan 31;29:E100 [FREE Full text] [doi: 10.1017/s2045796020000013]
42. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 08;8(11):e31170 [FREE Full text] [doi: 10.2196/31170] [Medline: 34747713]
43. Levin ME, Krafft J, Levin C. Does self-help increase rates of help seeking for student mental health problems by minimizing stigma as a barrier? J Am Coll Health 2018;66(4):302-309. [doi: 10.1080/07448481.2018.1440580] [Medline: 29447600]
44. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]
45. Mascheroni G, Ólafsson K. Net children go mobile: risks and opportunities. Second edition. Educatt. 2014. URL: https://netchildrengomobile.eu/ncgm/wp-content/uploads/2013/07/DEF_NCGM_SecondEdition_Report.pdf [accessed 2023-11-28]
46. Rashid A, Zeb MA, Rashid A, Anwar S, Joaquim F, Halim Z. Conceptualization of smartphone usage and feature preferences among various demographics. Cluster Comput 2020 Feb 05;23(3):1855-1873. [doi: 10.1007/s10586-020-03061-x]
47. Peters M, Mcinerney P, Soares CB, Parker D, Godfrey CM, Khalil H. Methodology for JBI scoping reviews. In: Joanna Briggs Institute Reviewers' Manual: 2015 Edition. Adelaide, Australia: The Joanna Briggs Institute; 2015.
48. Page MJ, Moher D. Evaluations of the uptake and impact of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and extensions: a scoping review. Syst Rev 2017 Dec 19;6(1):263 [FREE Full text] [doi: 10.1186/s13643-017-0663-8] [Medline: 29258593]
49. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]

50. Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environ Evid 2018 Feb 1;7(1):8 [FREE Full text] [doi: 10.1186/s13750-018-0115-5]
51. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
52. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo MP, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
53. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme. Lancaster University. 2006 Apr. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2023-11-28]
54. Dahne J, Collado A, Lejuez CW, Risco CM, Diaz VA, Coles L, et al. Pilot randomized controlled trial of a Spanish-language behavioral activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among United States Latinx adults with limited English proficiency. J Affect Disord 2019 May 01;250:210-217 [FREE Full text] [doi: 10.1016/j.jad.2019.03.009] [Medline: 30870770]
55. Tighe J, Shand F, McKay K, Mcalister TJ, Mackinnon A, Christensen H. Usage and acceptability of the iBobbly app: pilot trial for suicide prevention in aboriginal and Torres Strait islander youth. JMIR Ment Health 2020 Dec 01;7(12):e14296 [FREE Full text] [doi: 10.2196/14296] [Medline: 33258782]
56. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
57. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
58. McCall T, Ali MO, Yu F, Fontelo P, Khairat S. Development of a mobile app to support self-management of anxiety and depression in African American women: usability study. JMIR Form Res 2021 Aug 17;5(8):e24393 [FREE Full text] [doi: 10.2196/24393] [Medline: 34133313]
59. Craig Rushing S, Kelley A, Bull S, Stephens D, Wrobel J, Silvasstar J, et al. Efficacy of an mHealth intervention (BRAVE) to promote mental wellness for American Indian and Alaska native teenagers and young adults: randomized controlled trial. JMIR Ment Health 2021 Sep 15;8(9):e26158. [doi: 10.2196/26158] [Medline: 34524092]
60. Stephens D, Peterson R, Singer M, Johnson J, Rushing SC, Kelley A. Recruiting and engaging American Indian and Alaska native teens and young adults in a SMS help-seeking intervention: lessons learned from the brave study. Int J Environ Res Public Health 2020 Dec 16;17(24):9437 [FREE Full text] [doi: 10.3390/ijerph17249437] [Medline: 33339260]
61. Wrobel J, Silvasstar J, Peterson R, Sumbundu K, Kelley A, Stephens D, et al. Text messaging intervention for mental wellness in American Indian and Alaska native teens and young adults (BRAVE study): analysis of user engagement patterns. JMIR Form Res 2022 Feb 25;6(2):e32138 [FREE Full text] [doi: 10.2196/32138] [Medline: 35212633]
62. Maloney CA, Abel WD, McLeod HJ. Jamaican adolescents' receptiveness to digital mental health services: a cross-sectional survey from rural and urban communities. Internet Interv 2020 Sep;21:100325 [FREE Full text] [doi: 10.1016/j.invent.2020.100325] [Medline: 32455121]
63. Neal-Barnett A, Stadulis R, Ellzey D, Jean E, Rowell T, Somerville K, et al. Evaluation of the effectiveness of a musical cognitive restructuring app for Black inner-city girls: survey, usage, and focus group evaluation. JMIR Mhealth Uhealth 2019 Jun 27;7(6):e11310 [FREE Full text] [doi: 10.2196/11310] [Medline: 31188130]
64. Povey J, Mills PP, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of mental health apps for Aboriginal and Torres Strait Islander Australians: a qualitative study. J Med Internet Res 2016 Mar 11;18(3):e65 [FREE Full text] [doi: 10.2196/jmir.5314] [Medline: 26969043]
65. Röhr S, Jung FU, Pabst A, Grochtdreis T, Dams J, Nagl M, et al. A self-help app for Syrian refugees with posttraumatic stress (Sanadak): randomized controlled trial. JMIR Mhealth Uhealth 2021 Jan 13;9(1):e24807 [FREE Full text] [doi: 10.2196/24807] [Medline: 33439140]
66. Agapie E, Chang K, Patrachari S, Neary M, Schueller SM. Understanding mental health apps for youth: focus group study with Latinx youth. JMIR Form Res 2022 Oct 18;6(10):e40726 [FREE Full text] [doi: 10.2196/40726] [Medline: 36256835]
67. Watson-Singleton NN, Pennefather J, Trusty T. Can a culturally-responsive mobile health (mHealth) application reduce African Americans' stress?: a pilot feasibility study. Curr Psychol 2021 Mar 02;42(2):1434-1443. [doi: 10.1007/s12144-021-01534-9]
68. Thapar A, Eyre O, Patel V, Brent D. Depression in young people. Lancet 2022 Aug 20;400(10352):617-631. [doi: 10.1016/S0140-6736(22)01012-1] [Medline: 35940184]
69. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]

70. Baum A, Revenson TA, Singer J. Race and health: racial disparities in hypertension and links between racism and health. In: Handbook of Health Psychology Second Edition. New York, NY: Psychology Press; 2011.
71. Ramos G, Ponting C, Labao JP, Sobowale K. Considerations of diversity, equity, and inclusion in mental health apps: a scoping review of evaluation frameworks. Behav Res Ther 2021 Dec;147:103990 [FREE Full text] [doi: 10.1016/j.brat.2021.103990] [Medline: 34715396]

Abbreviations

BYOTS: Build Your Own Theme Song
CBT: cognitive behavioral therapy
FTF: face-to-face
iPST: internet-based problem-solving therapy
MHA: mental health app
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PTSD: posttraumatic stress disorder
STEM: science, technology, engineering, and mathematics

Edited by J Torous; submitted 15.05.23; peer-reviewed by E Brondolo, W Bramer; comments to author 24.08.23; revised version received 14.10.23; accepted 21.10.23; published 06.12.23.

Please cite as:
Saad F, Eisenstadt M, Liverpool S, Carlsson C, Vainieri I
Self-Guided Mental Health Apps Targeting Racial and Ethnic Minority Groups: Scoping Review
JMIR Ment Health 2023;10:e48991
URL: https://mental.jmir.org/2023/1/e48991
doi:10.2196/48991
PMID:38055315

Review

Digital Cognitive Assessment Tests for Older Adults: Systematic Literature Review

Claudio Cubillos[1*], BA, MSc, PhD; Antonio Rienzo[2*], BA, MSc

[1]Escuela de Ingeniería Informática, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

[2]Escuela de Ingeniería Biomédica, Universidad de Valparaiso, Valparaíso, Chile

[*]all authors contributed equally

Corresponding Author:
Antonio Rienzo, BA, MSc
Escuela de Ingeniería Biomédica
Universidad de Valparaiso
General Cruz 222
Valparaíso, 2362905
Chile
Phone: 56 998291207
Email: antonio.rienzo@uv.cl

Abstract

Background: The global health pandemic has affected the increasing older adult population, especially those with mental illnesses. It is necessary to prevent cases of cognitive impairment in adults early on, and this requires the support of information and communication technologies for evaluating and training cognitive functions. This can be achieved through computer applications designed for cognitive assessment.

Objective: In this review, we aimed to assess the state of the art of the current platforms and digital test applications for cognitive evaluation, with a focus on older adults.

Methods: A systematic literature search was conducted on 3 databases (Web of Science, PubMed, and Scopus) to retrieve recent articles on the applications of digital tests for cognitive assessment and analyze them based on the methodology used. Four research questions were considered. Through the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology, following the application of inclusion and exclusion criteria, a total of 20 articles were finally reviewed.

Results: Some gaps and trends were identified regarding the types of digital applications and technologies used, the evaluated effects on cognitive domains, and the psychometric parameters and personal characteristics considered for validation.

Conclusions: Computerized tests (similar to paper-and-pencil tests) and test batteries (on computers, tablets, or web platforms) were the predominant types of assessments. Initial studies with simulators, virtual environments, and daily-life activity games were also conducted. Diverse validation methods and psychometric properties were observed; however, there was a lack of evaluations that involved specific populations with diverse education levels, cultures, and degrees of technology acceptance. In addition, these evaluations should consider emotional and usability aspects.

(JMIR Ment Health 2023;10:e47487) doi:10.2196/47487

KEYWORDS

cognitive digital test; systematic review; cognitive screening; digital interventions; older adults

Introduction

Background

According to the United Nations, a country is considered old when ≥7% of its population is aged >60 years [1,2]. However, some countries exceed this percentage [3-5]. The increase in life expectancy and the growing population of older adults represent some of the most significant demographic changes that society is experiencing today. In all countries, individuals aged ≥60 years are beginning to constitute a large segment of the population. It is estimated that the proportion of this population group will increase 2-fold, rising from 11% to 22% by the year 2050 [6]. Although there are significant variations between countries and continents, the segment of the population aged 60 years is the fastest growing [7]. In addition, 10.4% of adults aged >60 years have cognitive impairments associated

with aging. In this group, 20.9% of people aged >80 years have this condition [8]. This affects the social and economic aspects of a country [9,10].

However, it is not the process of aging itself that is causing alarm among current governments and societies, but rather the burden of dementia that is associated with this aging population. The probability of developing dementia increases with age, doubling approximately every 5 years. In general, it is estimated that there are approximately 40 million people with dementia worldwide, with approximately 9 million people in Europe alone. Furthermore, this global figure is estimated to double every 2 decades, reaching 131.5 million by 2050 [11,12].

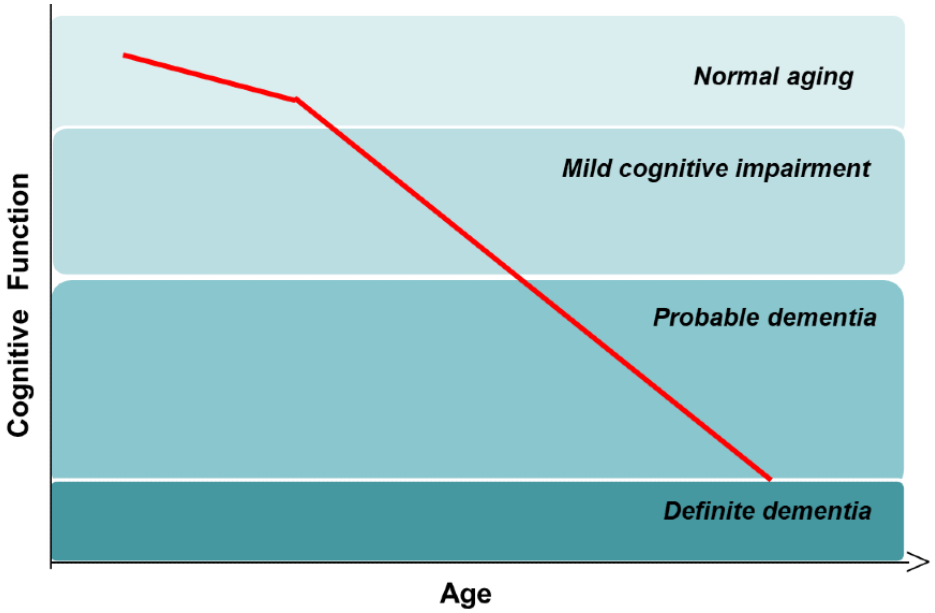
Due to the aforementioned reasons, there has been an increasing interest in cognitive training and other interventions that can mitigate or reverse these degenerative changes in older adults. There are several research papers and recent literature on the outcomes of cognitive interventions for brain training in older adults [12-15], but their scientific positions vary, and all authors emphasize the need for more empirical evidence. Furthermore, cognitive decline and changes in cognitive status [16] can easily go unnoticed in clinical settings. Cognitive assessments are often time-consuming and require a trained health care specialist, such as a neurologist or gerontologist [17,18], to provide detailed information on the patient's health [19]. However, conducting these evaluations poses significant access barriers, as many older adults are unable to attend promptly due to physical or cognitive limitations, fear of going out (due to the pandemic), long waiting times, or long travel distances. These barriers constitute important obstacles in determining the initial stages of cognitive decline.

Thus, this study aimed to provide an updated literature review of the main specialized digital cognitive tests for older adults. This review classified the tests based on their types and characteristics. In addition, it included comparative tables that highlight the technological aspects, cognitive domains evaluated, tasks, activities, and psychometric parameters used in each test.

Cognitive Problems in Older Adults

As a normal component of aging, many people experience a decline in their cognitive functions. When the decline becomes more significant, pathological processes may occur. Different levels of cognitive impairment were observed. As cognitive degeneration progresses, cognitive and functional decline reach a threshold, and the person is clinically diagnosed with probable dementia [20]. At present, it is unlikely that neuronal damage in the brain can be reversed, and most of the recent treatments available only provide symptom relief rather than a cure for the disease. However, the disease progression can be effectively controlled if dementia is detected at an early stage. Therefore, the most effective strategy is to detect dementia in its early stages and initiate an intervention. Theoretical changes in cognitive function in a person as a function of age (toward possible dementia) are shown in Figure 1 [20].

Figure 1. Changes in cognitive functions in a person.



In recent years, the concept of mild cognitive impairment (MCI) has been applied to describe the transitional stage between normal aging and the early stages of dementia [21]. It refers to a "mild" condition in which there is objective memory impairment without functional impairment [21,22]. In general, individuals with MCI have a high probability of gradually progressing to dementia. This means that they are at a higher risk of developing dementia than healthy individuals. Approximately 12% to 15% of individuals with MCI develop clinical dementia with functional disability within 1 year; therefore, the early detection of dementia depends on an accurate diagnosis [21,22].

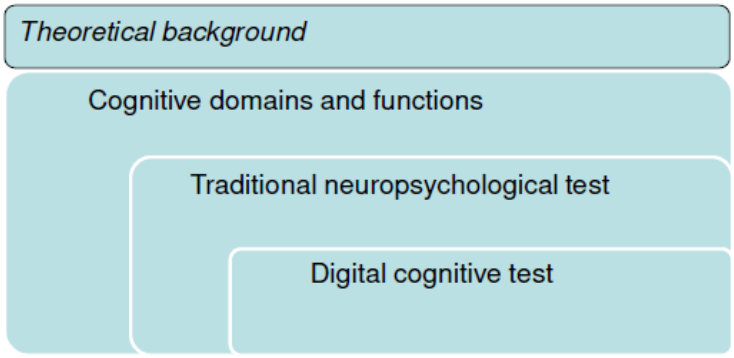
According to specialists [1], there are various types of cognitive impairment. Currently, amnesic and nonamnesic MCI are distinguished. This distinction is based on the presence or absence of deterioration in the mnemic function. In addition, it is possible to differentiate MCI according to the number of affected cognitive domains. Some individuals have unidomain MCI, whereas others have multidomain MCI, which involves

impairment in >1 cognitive domain. Although memory impairment is the most representative symptom of MCI, several cognitive domains other than memory are compromised in most individuals with MCI [20,23,24].

Theoretical Background

A summary is presented on the theoretical background of cognition and cognitive functions, and then it delves into the traditional evaluation tests as well as the digital cognitive tests that are available. The scheme of the work is shown in Figure 2.

Figure 2. Structure of the theoretical background.



Cognitive Domains and Functions

For specialists, cognition is defined as "the set of psychological skills that account for all mental life," and it is composed of "cognitive domains" (sets of cognitive processes or functions). Cognitive functions are mental processes that allow us to perform any task. They enable an individual to actively participate in the processes of receiving, selecting, transforming, processing, storing, and retrieving information, thus enabling them to function effectively in their surroundings. Cognitive skills are continuously used to learn and remember information, integrate personal history and identity, manage information related to the individual's location and destination, maintain and distribute attention, recognize different sounds, process different stimuli, perform calculations, and mentally represent an object [25,26].

According to the literature, the most important cognitive functions include attention [25,27,28], orientation [27,29], memory [25,27], perception (or gnosis) [25,26,29], executive functions [26,29], praxis [25,26,29], language [25,29], social cognition [26,29], and visuospatial skills [26,29]. Cognitive disorders affect the cognitive functions of individuals who experience them.

A special area of interest (for older adults) turns out to be memory, which deserves further explanation. It is the most frequently mentioned cognitive function, the lack of which is attributed to making most daily errors. Apparently, everything is attributed to the "lack" of memory, which is difficult to define. However, in general terms, memory refers to the ability to acquire, store, and retrieve various types of information [25,30]. At each stage of the memory process, the entire brain is involved [27], encompassing several phases: registration, encoding, storage, recall, and recognition of information. There are different forms and types of memory [25,27], including sensory or iconic memory; short-term memory (STM); working memory (also called operational memory); and long-term memory, which can be divided into 2 groups [29-31]: explicit or declarative memory and implicit or nondeclarative memory.

Traditional Neuropsychological Tests

Neuropsychological evaluation is used to determine an individual's cognitive status. It is conducted when there are symptoms of cognitive syndromes such as aphasia or dementia. The evaluation consists of performing cognitive tests to establish the presence of these syndromes. For example, if it is necessary to determine a patient's language proficiency, a cognitive test assessing language skills should be conducted. Similarly, if it is necessary to determine a patient's memory status, a cognitive test evaluating their memory should be performed. Therefore, neuropsychological evaluation can determine the presence of cognitive or behavioral syndromes and suggest the etiology of the pathological condition as well as the possible evolution. This knowledge is crucial for determining appropriate rehabilitation measures for patients [32].

Several cognitive assessment techniques have been developed to assess adults in clinical settings. They can be distinguished between tests for cognitive screening, which diagnose possible levels of cognitive impairment or degrees of dementia, and specialized tests that evaluate one or more specific cognitive domains [16,33]. A previous literature review established that the most common screening types are the Montreal Cognitive Assessment (MoCA) [34,35], the Mini-Mental State Examination (MMSE) [30,36,37], the Addenbrooke's Cognitive Examination-Revised [31], the Mini-Cog [35] and the Abbreviated Mental Test (AMT) [37,38]. Among the tests used to evaluate cognitive abilities in specific domains and in patients with special characteristics are the following: the Stroop test, which is an attentional test that detects neurological and brain problems and assesses the ability to classify information from the environment and react selectively to it [39]; the Corsi Cubes test, which evaluates visuospatial STM and allows the study of the effect of emotions on STM [40]; the Trail Making Test (TMT), a neuropsychological test that measures visual attention and task switching (consisting of 2 parts) [41]; and the Rey-Osterrieth Complex Figure test, which provides information about a person's neuropsychological functioning [42] in terms

of attention, concentration, coordination, and visuospatial abilities, among others.

Most cognitive tests are evaluated by specialists by using psychometric parameters. Psychometry is a branch of experimental psychology that is responsible for measuring and quantifying a person's psychological processes and cognitive abilities [43]. The most commonly used measures to assess the quality of psychometric instruments, as well as the measurement parameters that are normally used in cognitive tests, include reliability [44], validity [45], sensitivity [44,46], specificity [44,46], and receiver operating characteristic (ROC) curves [47]. These measures are briefly explained in Multimedia Appendix 1.

Digital Cognitive Tests

New computing technologies and platforms, including tablets and smartphones, offer many opportunities to create interactive tasks and experiences that can be used to infer the cognitive status. Cognitive assessment software packages are available [48,49], and they offer computerized versions of traditional tests that can be self-administered. In addition, various websites perform tests and training for certain cognitive abilities. Examples include Cognifit [50], NeuronUP [51], and Stimulus [52].

An analysis of previous literature reviews yielded 4 papers by the following authors: Zygouris and Tsolaki [53] in 2015, Aslam et al [54] in 2017, Marques-Costa et al [55] in 2018, and Tsoy et al [56] in 2021. These reviews included 11 to 17 papers that were published in 2012, 2015, 2017, and 2019, respectively. The most recent study by Tsoy et al [56] focused on only 3 cognitive domains: attention, memory, and language. From the above, it can be deduced that there is a need for an up-to-date state-of-the-art technology. Previous reviews have mainly focused on computer tests and digital test batteries, neglecting the inclusion of emerging technologies such as virtual reality (VR), video games, gamification, and artificial intelligence (AI). Tsoy et al [56] used this as an exclusion criterion for his systematic review.

Previous reviews provide different levels of detail regarding the characteristics of different digital tests. These include the hardware used [53,54,56], input mode or data capture [53,54], test time [53-56], and administration modality [53,54]. However, none of them offer details on the instructions and how they are delivered to the patient, the environment or place where the test is performed, or the sequence of tasks that the participants must perform. This lack of information makes it difficult to compare the results within the same cognitive domain.

One aspect analyzed by all these reviews was the quality of the cognitive tests and the psychometric properties applied. They agree on highlighting important drawbacks in terms of the replicability of the studies [54], existence of well-structured psychometric data [55], and evaluation of various psychometric properties [53]. However, a limitation of these reviews is that they do not include the sample size in the tests, details of how many participants had MCI or dementia, or psychometric parameters used for each test in their comparative tables. All these aspects are addressed in this review.

The main specialized digital cognitive tests for older adults are presented, classifying and comparing the domains and cognitive tasks evaluated. This complements the missing characteristics in the analyses of the previous reviews.

Methods

Objectives and Research Questions

To investigate the current state of applications or digital tests for cognitive evaluation in older adults, 4 research questions were formulated:

- RQ1: What are the different technological alternatives that are currently used in digital devices to assess the cognitive abilities of older adults?
- RQ2: What are the different types and characteristics of computerized (digital) cognitive tests?
- RQ3: What are the main characteristics of the subtests and tasks used in different digitized cognitive tests?
- RQ4: What are the main effects, personal traits, and psychometric parameters considered for validating digitized cognitive tests for older adults?

Eligibility Criteria

A literature review was conducted on digital systems for detecting cognitive problems. The aim was to gather updated information on technological solutions that could help overcome the possible barriers and difficulties of traditional psychometric tests. To do this, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology was used. Through a standardized review process, information is delivered in a flowchart that considers 4 stages: identification, screening, eligibility, and inclusion.

To search primary studies (articles), we searched 3 databases based on the search strategies: Web of Science, PubMed, and Scopus. These databases were chosen because they have peer review processes in which experts approve the publications. We combined the keywords with logical operators to obtain the following search expression:

("cognitive assessment" OR neuropsych*) AND (computer* OR web OR "digital test" OR evaluation) AND ("older adult" OR adult) AND ("cognitive impairment") AND valid*.

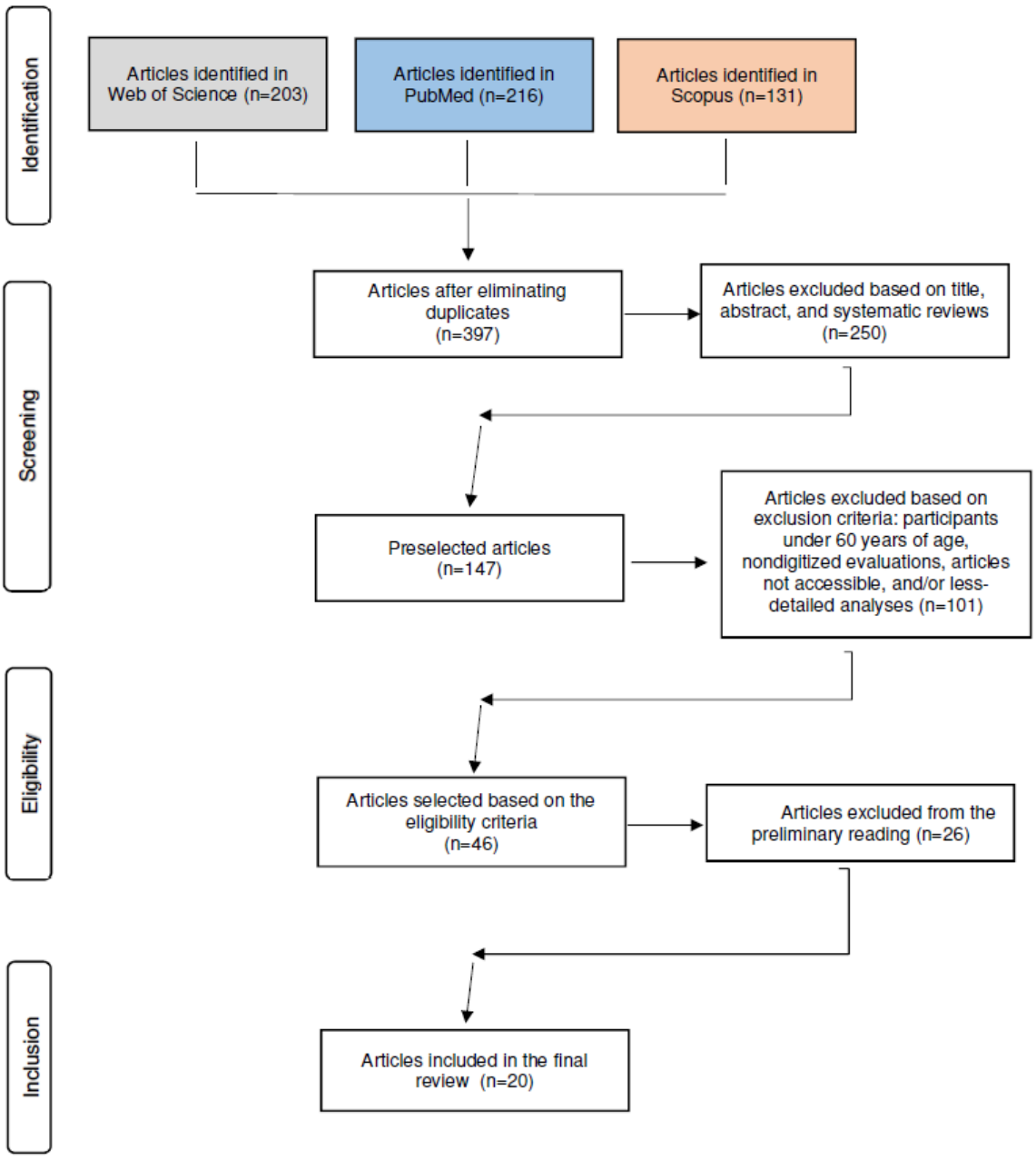
Due to the high number of publications (although a large percentage of articles appear on several sites), the following criteria were considered for the first selection and review of the most relevant articles:

- Inclusion criteria were articles and research papers that were published 2015 onward and were written in both English and Spanish; those articles in which titles and summaries (abstracts) included terms that addressed any of the research questions. The articles should be published by publishers with a website and should fall under the categories of scientific articles, conferences (proceedings), or book chapters.
- Exclusion criteria were articles in which titles and abstracts were not related to the objective of the study or the research questions, those that were repeated in another language, and those that were not related to older adults.

Data Collection

On the basis of the proposed methodology, we searched for articles from the 3 databases. The process is shown in the flowchart in Figure 3. The explanation for each phase of the PRISMA methodology is as follows:

Figure 3. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) scheme of the systematic review (own elaboration).



Identification

A systematic search of the literature was performed using 3 detailed databases, resulting in the retrieval of 203 articles from Web of Science, 216 articles from PubMed, and 151 articles from Scopus.

- Screening: duplicates were eliminated, leaving 397 articles. The searches were then filtered by title, abstract, and systematic review, resulting in a total of 147 reports.
- Eligibility: after conducting the first superficial reading of the articles (n=101), those that did not meet the inclusion criteria were excluded, resulting in 46 studies.
- Inclusion: these 46 documents were reviewed again to identify points that did not align with the study objectives. Finally, 20 articles were selected. All articles corresponded to scientific articles; there were no conference proceedings, book chapters, or university theses.

In a previous search, various validated methodologies were found concerning the so-called “gold standard,” or reference test. These tools allow for different methodologies to be used in conducting structured tests to quantify the affected cognitive domain. Table 1 was created using 20 articles discovered. Each article included the first author’s name, year of publication, country, instrument name, and the technological alternative that met certain quality criteria.

Table 1. Articles analyzed.

Study, year	Country	Article title	Types of technology
Köstering et al [57], 2016	Germany	Analyses of rule breaks and errors during planning in computerized tower tasks: insights from neurological patients	Computerized cognitive tests
Scharre et al [58], 2017	United States	Digitally translated Self-Administered Gerocognitive Examination (eSAGE): Relationship with its validated paper version, neuropsychological evaluations, and clinical assessments	Digital version
Wong et al [59], 2017	China	Computerized Cognitive Screen: a self-administered computerized test for screening for cognitive impairment in community social centers	Computerized cognitive battery
Valladares-Rodriguez et al [60], 2017	Spain	Design process and preliminary psychometric study of a video game to detect cognitive impairment in senior adults	Computerized cognitive game
Rapp et al [61], 2018	United States	Computer simulations for assessing cognitively intensive instrumental activities of daily living in older adults	Simulator of daily activity
Possin et al [62], 2018	United States	The Brain Health Assessment for detecting and diagnosing neurocognitive disorders	Cognitive web platform
Groppell et al [63], 2019	United States	A rapid, Mobile Neurocognitive Screening Test to aid in identifying cognitive impairment and dementia (BrainCheck): cohort study	Cognitive web platform
Khaligh-Razavi et al [64], 2019	United States	Integrated Cognitive Assessment: speed and accuracy of visual processing as a reliable proxy to cognitive performance	Computerized cognitive tests
Eraslan Boz et al [65], 2019	Turkey	A new tool to assess amnestic mild cognitive impairment in Turkish older adults: Virtual Supermarket	Virtual reality environment
Takahashi et al [66], 2019	Japan	Development and validity of the Computer-Based Cognitive Assessment Tool for intervention in community-dwelling older individuals	Computerized cognitive battery
Ichii et al [67], 2019	Japan	CogEvo, a cognitive function balancer, is a sensitive and easy psychiatric test battery for age-related cognitive decline	Computerized cognitive battery
Schulz-Heik et al [68], 2020	United States	Evaluation of adding the CANTAB[a] computerized neuropsychological assessment battery to a traditional battery in a tertiary care center for veterans	Computerized cognitive battery
Cahn-Hidalgo et al [69], 2020	United States	Validity, reliability, and psychometric properties of a computerized, cognitive assessment test (Cognivue)	Computerized cognitive battery
Tsoy et al [70], 2020	United States	BHA-CS[b]: A novel cognitive composite for Alzheimer's disease and related disorders	Cognitive web platform
Chin et al [71], 2020	Korea	A validation study of the Inbrain CST[c]: a tablet computer-based Cognitive Screening Test for Elderly People with cognitive impairment	Computerized cognitive battery
Lunardini et al [72], 2020	Italy	Supervised digital neuropsychological tests for cognitive decline in older adults: usability and clinical validity study	Computerized cognitive tests
Noguchi-Shinohara et al [73], 2020	Japan	A new computerized assessment battery for cognition (C-ABC) to detect mild cognitive impairment and dementia around 5 min	Cognitive web platform
Chan et al [74], 2020	China	Electronic cognitive screen technology for screening older adults with dementia and mild cognitive impairment in a community setting: development and validation study	Cognitive web platform
Rodriguez-Salgado et al [75], 2021	Cuba	A brief digital cognitive assessment for detection of cognitive impairment in Cuban older adults	Cognitive web platform
Bottiroli et al [76], 2021	Italy	The smart aging platform for assessing early phases of cognitive impairment in patients with neurodegenerative diseases	Computerized cognitive game

[a]CANTAB: Cambridge Neuropsychological Test Automated Battery.

[b]BHA-CS: Brain Health Assessment-Cognitive Score.

[c]CST: Cognitive Screening Test.

Results

Overview

Some preliminary statistics could be obtained regarding the publication dates of the articles, such as the countries where the studies were conducted (Table 1). Subsequently, the technological alternatives were classified and explained. For example, differentiating between digital tests, that is, if they are cognitive batteries and computerized cognitive tests, or computerized platforms (web), or simulators of daily activity, using a VR environment, or computerized cognitive games. Possible factors that influenced each psychometric instrument were also characterized. For example, the technical aspects of

the tests such as the administration time, data capture mode, modality (technology used), operation (instructions) of the software, and how the tests are administered.

In addition, comparative tables were created to provide a synthesized description of the digitized cognitive tests, differentiating some characteristics such as the cognitive domains covered, the number of tests or stages available, the activities involved in each test, and the evaluation scoring system. Finally, the articles were analyzed to validate their psychometric properties in comparison with traditional tests.

Statistical data of the articles studied are shown in Tables 2 (based on the year of publication) and 3 (based on the country where the study was conducted).

Table 2. Papers by the year of publication.

Year of publication	Papers (n=20), n (%)
2016	1 (5)
2017	3 (15)
2018	2 (10)
2019	5 (25)
2020	7 (35)
2021	2 (10)

Table 3. Papers by the country of study.

Countries for study	Papers (n=20), n (%)
United States	8 (40)
Europe	5 (25)
Japan	3 (15)
Chinese	2 (10)
Korea	1 (5)
Cuba	1 (5)

RQ1: What Are the Different Technological Alternatives That Are Used in Digital Devices to Assess the Cognitive Abilities of Older Adults?

Overview

To answer the first research question (RQ1), we conducted a preliminary analysis of the selected articles, identified differences in categories of digital tests, and briefly described the identified classes. Table 4 shows the number of articles on each application (according to the type of technology).

Table 4. Papers by the type of digital test (technology).

Types of digital test (technology)	Papers (n=20), n (%)
Test battery	6 (30)
Web platform	6 (30)
Computerized test	3 (15)
Games	2 (10)
Digital version	1 (5)
Simulator IADL[a]	1 (5)
Virtual reality	1 (5)

[a]IADL: Independent Activities of Daily Living.

Cognitive Batteries and Computerized Cognitive Tests

These are a set of tests and tasks that allow the evaluation of multiple cognitive domains, such as language, executive function, attention, and memory [57,59]. The tests are based on and validated using psychometric methods. They present a greater advantage than traditional tests (pencil and paper), as they lead to a detailed cognitive profile, reducing possible errors caused by administration bias [59]. In addition, they provide information about the testing process, such as reaction times

and the sequence of answers (good and not good). In addition, automatic scoring helps professionals improve clinical diagnoses. The small difference between test batteries [59,66-69,71] and computerized tests [57,64,72] is that the former has several tests and tasks designed to assess >4 different cognitive domains, whereas the latter includes fewer tests and focuses on a single cognitive domain. In general, computerized cognitive batteries, which are used on PCs, notebooks, or tablets, serve as supportive tools in clinical and community settings.

Simulators of Daily Activity

At the onset of MCI, the ability to perform daily activities remains unaffected. However, as the deterioration progresses, the performance of these activities decreases. Activities such as shopping, taking medications, and using telephone lines become difficult. To assess the condition, it is necessary to invest in devices and time and to consider the burden it places on the patients. As an alternative, the use of simulators for daily activities offers enormous possibilities. For example, in Simulation-Based Assessment of Cognition [61], the patient can interact with software to complete tasks such as withdrawing money at a virtual automated teller machine or making a call to a virtual pharmacy to indicate the needed medicine. These activities are associated with executive function. Therefore, it is an instrument with many advantages that require further development to include the full range of activities commonly performed by older adults, vary the performance depending on the particular neuropathology, and adapt to different cultural and socioeconomic settings.

Computerized Platforms (Web)

These are cognitive batteries, tasks, or adaptations of traditional tests that allow comprehensive evaluations of cognitive impairment using mobile devices or devices connected to the internet. The cognitive evaluation begins after patients enter their sociodemographic data. Among these options, the platforms can be configured with an algorithm that enables the questions to be displayed and prompts the participant to select the correct answer. Meanwhile, the information is stored on a centralized computer server, which allows access to cognitive detection to be faster, more efficient, and automatic; and it is possible to perform the test in the comfort of one's home (or any appropriate place) [62,63,70,73-75].

Digital Version

These are digital cognitive assessments that are equivalent to paper versions and allow for greater flexibility. In addition, the tool (a single test, not a battery) can increase the screening of individuals who are being evaluated through self-administration using technological devices. The technological solution is built based on the questions from the original version but with the added advantage of automatically measuring the time it takes for participants to answer the questions. In addition, it can determine the frequency with which participants return to previous pages based on the subject. Finally, the evaluation is accessible through web, and the results obtained are delivered in a digital format [58].

VR Environment

Currently, there is an increase in the use of VR technology for evaluating cognitive dysfunction. An individual can enhance their interaction in a simulated environment by following the instructions of a traditional cognitive test. For example, in a Virtual Supermarket (VSM) [65], before the exercise, age, gender, occupation, years of education, and any possible memory complaints are registered. VSM generates a randomized list of products for a daily shopping activity. The individual is expected to locate the items on the list, place them in the shopping cart, take them to the register and pay the correct amount. Furthermore, the participant must navigate the VSM by touching green footprints on the screen while pushing the shopping cart. It is an exercise designed to examine multiple cognitive domains, such as visual and verbal memory, executive functions, attention, and spatial navigation. People with cognitive impairment require more time and make a greater number of errors than healthy individuals [65]. For example, patients with cognitive impairment will not be able to remember a list of instructions. The authors mentioned possible limitations if all the adults had similar experiences and functioning in daily life, especially in tasks related to purchases. They also note that the payment in euro currency could have added additional complexity and cognitive load.

The use of VR technologies allows for reduced costs and decreased administration time due to automatic scoring. In addition, participants may be able to self-test in the comfort of their homes without the supervision by a specialist, who will be consulted only if the test detects signs of possible deterioration. However, further studies that include participants with different degrees of familiarity with new technologies, especially tablets, are needed.

Computerized Cognitive Games

These applications can be categorized into 2D games and 3D scenario generation. Games (or game batteries) are clinically useful resources that allow for the detection of deficiencies in multiple domains. For example, Episode Gamification [60] is a game that involves taking a virtual walk through a medium-sized city where everyday objects are displayed. The challenge is to remember the maximum number of items to be displayed while avoiding any interfering objects. Another example is Smart Ageing [76], a 3D game that features a loft with a kitchen, bedroom, and living room area. Participants use a touch screen monitor to navigate and interact with the environment, performing five tasks related to daily life: (1) find a list of objects in the kitchen after exploring it; (2) water the flowers while listening to the radio, pressing the space bar each time the word "sun" is heard; (3) make a phone call using the phone book and the phone that are placed on the nightstand, remembering to turn on the television after dialing the number; (4) identify the 12 objects presented in task 1 from a 2D screen with 24 images of objects; and (5) find each of the objects searched for in task 1 while being in the kitchen. Therefore, cognitive computer games have emerged as a novel approach for assessing the cognitive state of people, allowing them to simulate recurring tasks and sensory stimuli while collecting information on the patient's reaction time in certain tests [60,76].

One problem with several traditional cognitive tests is that they exhibit a "learning bias" [64], meaning that an individual's cognitive performance improves with repeated exposure to the test, solely because of learning the task, without any actual change in their cognitive ability. Consequently, this bias reduces the reliability of a test when it is used repeatedly (for example, when monitoring performance over time). Computerized cognitive tests can overcome this difficulty by randomly tailoring different task contents to participating adults. In addition, to adopt a psychometric evaluation, it is necessary to normalize the target population according to their specific context and ensure semantic agreement in the tasks, considering both language and culture. For example, in this review, we found only 2 platforms that could be adapted to Latin American population [70,75].

Digital cognitive tests have the potential to be objective, standardized, and most importantly, repeatable. Computerized testing applications provide ideal formats for generating alternative tests, thereby improving test-retest reliability during repeated administration in long-term monitoring. As screening and monitoring tools for serious diseases, computerized cognitive tests are being developed, with emphasis on ensuring their comprehensiveness, validity, and reliability.

Regarding new technologies, although the selected articles were published between 2016 and 2021, only a few used games, virtual or augmented reality, or simulators of daily life for cognitive purposes. Only 4 applications moved in this direction [60,61,65,76], although we expect more to come in the near future. This is especially relevant for cases in which one wants to measure slight cognitive differences over time (whether improvements or deterioration) instead of simply discriminating between a healthy adult, someone with MCI, and a certain degree of cognitive impairment. Our review also found a study that used machine learning techniques [60]. The review by Marques-Costa [55] also observed the need to include item-response theory techniques associated with automatic assessment. The item-response theory can help adapt the difficulty level of cognitive tests to older adults' personal characteristics and context. This makes us expect a greater inclusion of data analysis and AI techniques in future research, especially those aimed at improving the accuracy of diagnosis and instrument reliability.

With regard to VR technology use, we highlight its absence in cognitive tests. From previous reviews [53-56], the only case mentioned was the Computer Assessment of Mild Cognitive Impairment application, which included a VR driving task [53]. In our systematic review, we found only one study based on supermarkets [65]. It can be argued that this approach allows designers to contrast their batteries with classic paper-and-pencil tests and compare their concurrent validity. At the same time, the higher cost and potential risk associated with developing a VR- or gaming-based test from scratch might also influence the decision to adopt a more "conservative" approach and rely on proven testing instruments.

Finally, an important limitation of most digitized cognitive applications is that they require reliable internet connection. It is necessary for these applications to be able to work offline, saving progress and not depending only on synchronous communication. Future work should explore solutions that can function without a stable internet connection to enhance the accessibility of such tools and encourage their use in rural areas. In addition, in the face of any network contingency, applications that can work offline [72] would perform all its functionalities on a local computer in the face of any network contingency.

RQ2: What Are the Different Types and Characteristics of Computerized (Digital) Cognitive Tests?

Overview

On the basis of the collected information, a characterization of the different digital cognitive tests was conducted to answer the second research question (RQ2). Next, the technical aspects of the tests, such as the administration time, data capture mode, modality, mode of operation (or software instructions), administration method, and location, are detailed. Some criteria or parameters allowed for comparing the different tests included in the selected articles, based on which Table 5 was prepared.

Table 5. Main characteristics of the cognitive digital tests.

Cognitive digital test	Time (min)	Input mode	Modality	Software instructions	Exam administration	Location
TOL[a] [57]	8	Touch Screen, peripheral, or PC mouse	Computer with touch monitor	Visual instructions (text and time limit)	Administered by examiner psychologist	Face to face; laboratory room
eSAGE[b] [58]	17	Touch screen	Tablet or web-based	Visual instructions	Self-administration	Face to face; community-clinical settings
CoCoSc[c] [59]	15	Touch screen or headphones	Computer with touch monitor	Visually or verbally (audio)	Self-administered or browser	Face to face; housing and community centers
Episodix Gamification (CVLT[d]) [60]	30-40	Touch screen or PC peripherals (joystick, mouse, and Kinect)	Android or computer (Windows, Linux, or iOS)	Instructions are provided in audio and text format.	Personal clinical support assistance	Face to face; community and university center
SIMBAC[e] (IADL[f]) [61]	10	Touch screen or PC peripheral	Tablet or computer	Modules with visual (text) and verbal (voice recorder) instructions	Self-managed or trained technician	Face to face; medical care center
UCSF[g] Brain Health [62]	10	Touch screen	Software platform or iPad 9.7 inch	Instructions in the software; examiner evaluating	Automated scoring; cannot be self-administered	Face to face; diagnosed in university centers
BrainCheack Inc [63]	21 (mean)	Mobile touch screen	iPad, iPhone, or desktop browser	Instruction by examiner	Research staff	Face to face; community center
CGN_ICA[h] [64]	5	Touch screen	iPad, Raspberry, or web	__[i]	Self-administered or examiner	Face to face or distance; clinic or home
VSM[j] [65]	25	Touch screen and computer peripherals	Tablet (10-inch) or PC	Instructions by examiner	Self-administered or personal assistance in repeating instructions	Face to face; institutes and medical centers
CompBased-CAT[k] [66]	10-15	Touch screen, PC peripherals, or headphones	Tablet (Asus) or computer (Windows 10)	Visual on-screen instructions and voice with external noise-cancelling hearing aid	Self-administered or minimal assistance in instructing	Face to face; institutes and geriatric hospitals
CogEvo [67]	10	Touch screen	Computer OS[l]	Audiovisual with home icon	Administered by examiner	Face to face
CANTA[m] [68]	45-60	Touch screen	Touchscreen computer (Windows)	Verbal instructions from the instructor	Administrated with trained supervisor	Face to face; adult centers
Cognivue [69]	10	PC peripherals	Computer OS	Automated instructions and test subbattery	Self-administered or assisted by nonclinical support staff	Face to face or clinical establishment. No specific place
BHA-CS[n] [70]	10	Touch screen	Software and tablet, TabCAT Pad de 9,7	—	Managed by examiner	Face to face; adult and Alzheimer centers
Inbrain CST[o] [71]	30	Touch screen	Tablet or OS Microsoft Windows 10	Verbal (written)	Minimum attendance	Face to face or distance; private room in clinic
Trail Making Test y Bells Test [72]	5	Touch screen	Tab A6 con S pen or Webserver nube	Verbal assistant, virtual supervision of the test	Unsupervised environment; virtual only	Distance; Geriatric Foundation or Home
C-ABC[p] [73]	5	Touch screen	Computer (OS) with touch screen (80×60)	On-screen text and verbal description with headphones on PC	Assistance by a psychologist if needed	Face to face; Memory Clinic
EC-Screen[q] [74]	5	Touch screen	Web or tablet	Reading questions answer	Autoadministrated or assisted	Face to face or distance; geriatric community settings
BHA [75]	10-5	Touch screen	Tablet or web	Digital survey	Neurologist examiner or neuro psychologist	Face to face or institute community centers

Cognitive digital test	Time (min)	Input mode	Modality	Software instructions	Exam administration	Location
SG[r] (IADL) [76]	10-30	Touch screen	Touch screen computer	Visual instructions (and examiner)	Administered in the presence of a neuropsychologist	Face to face or neuropsychology unit, communities

[a]TOL: Tower of London.
[b]eSAGE: Self-Administered Gerocognitive Examination.
[c]CoCos: Computerized Cognitive Screen.
[d]CVLT: California Verbal Learning Test.
[e]SIMBAC: Simulation-Based Assessment of Cognition.
[f]IADL: Independent Activities of Daily Living.
[g]UCSF: University of California, San Francisco.
[h]CGN-ICA: Cognitivity Neurosciences-Integrated Cognitive Assessment.
[i]Not available.
[j]VSM: Virtual Supermaket.
[k]CompBased-CAT: Computer-Based Cognitive Assessment Tool.
[l]OS: Multiple Operating System
[m]CANTAB: Cambridge Neuropsychological Test Automated Battery.
[n]BHA-CS: Brain Health Assessment-Cognitive Score.
[o]CST: Cognitive Screening Test.
[p]C-ABC: computerized assessment battery for cognition.
[q]EC-Screen: Electronic Cognitive Screen.
[r]SG: serious game.

Time

The administration time of the tools found in the literature varied from 5 to 45 minutes. Compared with the traditional MMSE assessment that relies on professional training, digital cognitive tests take approximately 10 minutes, excluding the time needed to score the participant. In contrast, if the examination is performed by a technician or someone without specialized training, it will take much longer [66]. Computerized batteries allow for more accurate measurements; for example, the computerized assessment battery for cognition [73] can be administered in a short amount of time (approximately 5 min). It is a sensitive battery that detects not only cognitive impairment but also dementia. Finally, positive results should be considered supportive methods and not definitive diagnoses. Therefore, patients should be referred for a more comprehensive evaluation by health professionals [59].

Input Mode (or Capture)

The data-acquisition mode was identified using a touch interface and computer peripherals. In a study of digital games, it was found that older adults preferred a touch interface to computer peripherals such as a keyboard and mouse. In addition, in the digital version of the Self-Administered Gerocognitive Examination (eSAGE) [58], the participants did not use a stylus but instead used their fingers to draw or write the requested answers. Finally, the use of hearing aids allows for the cancellation of external noise and helps patients avoid distractions [66].

Modality (of the Device)

Psychometric instruments were implemented on desktop computers, laptops, tablets, iPads or iPhones. The use of larger screens helps individuals with visual impairments access and complete the tests. These digital methods allow us to capture the response time with better precision. For example, the Electronic Cognitive Screen [74] integrates the clock task, which reflects the speed of processing, and executive function. This test, when displayed on a tablet, allows for the detection of a person's lesser fine motor control; it is easier than the paper version for older adults.

Software Instructions

The mode of operation (or software instructions) can be entered by examiners or more easily integrated into the software. For the "Simulation-Based Assessment of Cognition" simulation software, the instructions consist of voice recordings and text files [58]. Each module incorporates specific instructions. The platforms read the questions and then prompt the participants to select the correct answer. The "TMT and Bells" tests [72] incorporate AI to detect the participants' voices and dictate the task guidelines. In the virtual game "VSM," the guidelines are shown visually and auditorily on the screen for each activity in the game's virtual environment [65]. The administration of the digital version is minimally assisted because each item comes with simple written instructions. Finally, the role of the examiner must be clear because providing clues during the test can introduce bias into the evaluation.

Test Administration

The administration of a psychometric instrument can influence the results. The advantage of administering digital cognitive tests in a laboratory is the potential to reduce the frequency of errors, software failures, and interruptions, but the psychometric instrument can be conditioned to exclude patients [57]. The Computerized Cognitive Screen [59] allows for self-administration with automated scoring while still requiring

minimal assistance for older people. The automated calculation of scores in the "Cognivue" battery [69] is more efficient and consistent than traditional tests. In Episodix [60], standardized administration makes data collection and response time capture more efficient. In the digital version [58], web-based administration is useful for people living in rural regions with limited resources and a lack of access to health care providers. The test supervisor only observed and served as a guide or security during the evaluation. The greatest advantage found in VR environments is allowing older adults to check their cognitive functioning at home and only visit a specialist if necessary [65]. The Integrated Cognitive Assessment visual categorization test [64] was found to be self-administered due to its simple design with no language and culture barriers. The platform [66] has an automatic scoring algorithm that helps alleviate the burden on the professional staff. Thanks to automatic scoring, not requiring individuals with specialized knowledge makes it easier to eliminate human error and reduce the duration of examinations.

Location (or Place)

Participants were recruited from clinics, memory study centers, hospitals, and neurological institutions. Diagnostic tests in community centers are beneficial because they allow digital cognitive tests to be performed on a larger scale and are easily accessible, as they require minimal professional demand. In addition, for some older adults, interacting with a computer can be more pleasant than visiting a health care facility. It can also be administered by individuals with a lower level of education, although guidance from professionals is necessary for those in community centers and primary health care clinics. Home-administered cognitive tests can provide a relaxing, nonintrusive, and familiar environment for patients [64,72]. The flexibility of the applications in their electronic versions also facilitates their administration in the comfort of one's home using a mobile device or tablet. In the case of the TMT-Bell digital test [72], the authors emphasized the need to strengthen the collaboration between technicians and clinicians. They also suggested the development of novel indicators that could further enhance the utility of digitized tests. In addition, digitized tests were performed in a clinical setting under the supervision of a trained professional. Therefore, they proposed conducting more usability and validity studies on their entire platform in a domestic environment. Despite the digitization of cognitive tests, only 4 applications have declared that they can be used remotely (web-based) without the presence of a supervisor [64,71,72,74].

From the reviewed articles, we can conclude that there are clear advantages to conducting (digital) tests using technological devices. These tests allow psychometric evaluations to be executed in the comfort of one's home or community centers [62,63,65,70,73-75], with minimal intervention from a specialist professional. In addition, most psychometric tests were implemented on equipment with touchscreen devices, such as notebooks, tablets, and smartphones. However, there was one exception: the Cognivue application [69] was carried out with a desktop computer. In cases where an older person is not familiar with technology, assistance can be provided by a family member or the clinical supervisor.

In addition, computerized cognitive tests provide new opportunities to remotely monitor cognitive changes and detect early dementia. Without the need to visit a trained practitioner or counselor, computerized cognitive tests can be distributed over the internet and self-administered [58,59,66,71,72,74] or eventually supported by a supervisor [61,64,65,69]. Some tests need to be administered by neurologists, psychologists, or specialized clinical personnel [57,60,62,75,76].

Technology benefits psychometric tests in numerous ways. For example, it significantly reduces the time that elapses between the application of the evaluation instrument and the reporting of results, sometimes providing immediate feedback. In addition, technology helps to eliminate qualification errors that are common among human beings. It is also possible to implement technological security measures, such as using passwords, and finally, it enables the adaptation of the content of the test according to the characteristics of the person who will answer the test [32].

In addition, in the field of psychology, there is a growing need for cultural adaptation tests to prevent or minimize errors that may introduce bias, such as construct, methodological, and item-related errors, resulting from improper use of the tests. Therefore, it is necessary to thoroughly consider language and cultural differences. This will ensure that valid and reliable measurements can be obtained when the tests are applied to different cultural groups [77]. For example, in one study [75], the Brain Health Assessment (BHA) was translated and adapted into Cuban Spanish by a multidisciplinary team of four language experts: 1 geriatrician, 1 neurologist, 1 psychiatrist, and 3 neuropsychologists. The adaptation was performed in accordance with the current guidelines for cross-cultural test development, with the Guidelines for Translating and Adapting Tests [78,79], and the results were validated with the MoCA. In particular, verbal stimuli on the favorites task were adapted to have similar frequency and complexity as the English version and to represent culturally appropriate concepts in the Cuban context. Furthermore, the video game Episodix [60], which has a multilingual design and is supported in English, Spanish, and Galician, was adapted and administered in Spain; the Spanish version of the California Verbal Learning Test (CVLT) is known as test de Aprendizaje Verbal España-Complutense in Spanish [80], and their results were validated with the MMSE.

RQ3: What Are the Main Characteristics of the Subtests and Tasks Used in Different Digitized Cognitive Tests?

A test could be intended to assess a specific cognitive function or several functions, and a particular cognitive function could be assessed using one or more tasks. Table 6 shows the number of articles in terms of the evaluation of cognitive tasks in each application or test. Of the 20 articles, digital applications that included up to 5 different tests were found [61,67,68,76]. Six tests were found in the studies by Wong et al [59] and Takahashi et al [66], whereas 7 tests were found in the studies by Scharre et al [58], Groppell et al [63], and Chin et al [71]. One article included 8 different tests [73], and another article included up to 10 tests [69]. When examining articles that included a task or activity to assess different cognitive functions (Table 7), it

is evident that executive functions and memory are the most commonly addressed, whereas praxis, abstractions, and calculations are less frequently discussed. Regarding memory and its types (Table 8), we found that memory in general was the most extensively covered, assessed by 6 articles, whereas remote and prospective memory were the least covered, assessed by 1 article each.

To answer the third research question (RQ3), we examined the operational performance of all the digitized cognitive tests. Table 9 lists the applications discussed in the 20 articles, with a more detailed overview of the work carried out. This table includes the digital applications and their respective cognitive domains that they evaluate and train. It also provides a summary of how each application works and the sequence of tasks required of the participant.

Table 6. Papers by the number of task of the digital test.

Number of task of the digital test	Papers (n=20), n (%)
1	4 (20)
2	0 (0)
3	2 (10)
4	3 (15)
5	4 (20)
6	2 (10)
7	3 (15)
8	1 (5)
9	0 (0)
10	1 (5)

Table 7. Papers by assessed cognitive functions.

Functions	Papers (n=20), n (%)	References
Executive functions	16 (80)	[57-59,61-63,65-68,70-72,74-76]
Memory	15 (75)	[58-60,62-68,70,71,73-75]
Visuo-spatial	10 (50)	[58,62,65,67,69-71,74-76]
Attention	8 (40)	[59,65-68,71-73]
Language	6 (30)	[62,63,70,71,75,76]
Orientation	6 (30)	[58,59,67,69,71,73]
Processing speed	6 (30)	[62,64,68-70,72]
Learning	2 (10)	[59,64]
Praxis or gnosis	1 (5)	[63]
Abstractions	1 (5)	[58]
Calculations	1 (5)	[58]

Table 8. Papers by types of evaluated memory.

Memory types	Papers (n=20), n (%)	References
Memory (general)	6 (30)	[59,62,63,67,71,74]
Work memory	3 (15)	[59,68,75]
Episodic memory	3 (15)	[60,70,75]
Visual memory	2 (10)	[64,65]
Intermediate memory	2 (10)	[66,73]
Verbal memory	2 (10)	[58,65]
Remote memory	1 (5)	[66]
Prospective memory	1 (5)	[59]

Table 9. Type, domains, and description of tasks of digital applications.

Name of the evaluation, test, or battery; cognitive domains; type	Cognitive assessment tests, n	Brief description by activity
Computerized TOL[a] [57]; planning as an executive function; specific	1	• Tower task (rule breaks and errors during planning): the TOL consists of a series of instructions using colored balls (red, yellow, and blue) that are suspended within 3 bars. A reference configuration is required to be completed with a minimum of movement. Only one ball can be moved at a time; in each bar, only the most prominent ball can be moved, and the balls are only deposited on the bars. The examiner explains the instructions and before starting the test, presents a set of experimental problems to familiarize the participants with the task; in addition, there is a time limit of 1 min per test, with 24 tests. The software records the planning error when the participant tries to select a blocked ball that is under another ball, places a ball on a bar that has the number of full balls, or selects a position other than the final goal.
Digital Gerocognitive Examination (eSAGE[b]) [58]; orientation, executive-level, memory language, abstraction, calculations, and visuospatial skills; screening	7	• Date: the participant is instructed to write the date of the day. • Picture naming and verbal fluency: words are presented backward, and these describe the names of the images that appear on the screen (eg, piano and volcano). • Delayed recall and written instruction: the instructions for the tests must be memorized, and at the end of the last page, "I finished" must be written. • Modified Trail B and problem-solving task: they are then asked to draw a line from one circle to another, starting at 1 and alternating numbers and letters in order before ending in the letter (1, A, 2, B, 3, C, etc). • Determining similarities: later, they are asked to answer questions by comparing objects. • Word problem calculation: a problem must be solved with letters. • Copy 3D construction and clock drawing: a 3D square should be copied, and a clock face is drawn with the digits from 1 to 12. Then, the hands for 5 min past 11 o'clock (11:05) are asked to be drawn, and finally, the individual should place an L on the hour hand and an S on the min hand.
CoCoSc[c] [59]; learning and memory, executive functions, orientation, attention and working memory and PM[d]; screening	6	• Time-based PM and even-based PM: participants are instructed to perform an objective action; then a digital clock is displayed on the screen to control the time. The clock disappears if there is success, or after 30 s if no action has been executed. In the second PM test, the same procedure is followed but with different objectives. • Conflict inhibition (ringing doorbell): patients are asked to touch the screen twice in response to the sound of a doorbell and only once to the double sound of the doorbell (6 practice tests are administered before the evaluated activity). • Word List Learning: a list of 6 words of 2 syllables is presented at a speed of 1 s, and then the person remembers the word and mentions it aloud, thus facilitating the encoding of the memory. • Orientation to year, month, and day of the week and orientation to place: chronological orientation is measured by sets of questions and selection answers ("Yes," "No," "I don't know"), then answers to questions from the region, district, residential area, and location of the test are chosen. ("I don't know" and "Don't" are incorrect answers.) • Attention and working memory (test analogous to the Wechsler memory scale): the participant must reproduce on the screen of the device the sequential order (forward and backward) of the location where rabbits appear. • Delayed word list memory: there are 12 words of 2 syllables, which are classified into objective and distracting words; the participant indicates whether the word is objective or distracting.
Episodix, CVLT[e] [60]; episodic memory, regarding the verbal; specific	1	• CVLT: it is a game based on learning and a memory word list from the CVLT test. It covers many more elements of episodic memory than just those involve remembering words verbally. • The CVLT is a word set task that is displayed as a shopping list for a day of the week, and the participant performs the test and must remember all the items on the list. A second list of interferences is then presented for the next shopping day. After waiting for a while (the recovery phase) and producing a third list that includes 2 d of the week with new items, the participant must recognize the items in the list. The game is presented in a virtual village, where the participant moves to find objects that are displayed visually and audibly. There are 3 collections of objects; as in the original test, they are denoted as lists A, B and C, where list A allows the main learning, list B is interference, and list C is the recognition list. In addition, the recovery phase is time based, similar to the CVLT.

Name of the evaluation, test, or battery; cognitive domains; type	Cognitive assessment tests, n	Brief description by activity
SIMBAC[f] [61]; IADL[g] executive functions; specific	5	• Facial recognition and memory of names: an image of a human face is observed for 5 s, followed by a series of facial images of different sexes and ages. The image is paired with a unique name. Then several images are displayed, and the individual must select the photo that is paired with the image seen and the name. Six trials (3 males and 3 females) are presented; the difficulty grows with 2 to 4 name and face trials. • Filling a pillbox: a 7-d pillbox with 14 compartments is shown; in 3 pillbox containers, the instructions "take a pill overnight" are explained. They must correlate the shape and color of the pills within the pillbox, selecting the pill with the corresponding medication compartment. • Using an ATM[h]: the instructions indicate that money is withdrawn from a virtual ATM. An ATM card must be inserted, the PIN and the amount to be withdrawn must be entered, the money is withdrawn, and the card is recovered. • Automated prescription renewal by phone: participants are requested to renew the prescription of a drug in a pharmacy over phone using the drug label and the call system. The participant taps the phone and listens to a voice recording with instructions.
UCSF[i] BHA[j] [62]; memory, executive function, processing speed, visual-spatial, and language; screening	4	• Favorites (memory domain): participants are asked to indicate their favorite people, foods, and animals, and then they are shown 4 different faces 2 times. Each face is paired with a favorite food or animal. The pair is shown to them for 5 s and will reappear in random order. After each learning test, faces reappear, and participants are asked to remember the food and animal associated with each face. • Match (executive and speed domain): participants were shown a fixed caption of numbers 1 through 7, with corresponding simple abstract images appearing just below each number. The participant was instructed that whenever a number appeared in the middle of the screen, he should tap the corresponding image at the bottom of the screen as quickly as possible. After each answer, a new number appeared. Accurate responses were totaled in 2 min. • Line orientation (visuospatial domain): participants were shown 3 lines on a black background, a 2×50-mm white line shifted vertically above two 3×25-mm orange lines. One orange line was parallel to the white line, and another line was at a different angle. The white line was randomly placed at any angle between −60° and +60° from the vertical. Participants were asked to hit the orange line, which was parallel to the white line. The difficulty was manipulated from test to test by varying the angle difference. • Animal fluency (language; widely used test): participants name different animals as fast as they can for 1 min.

BrainCheck memory assessment Inc [63]; memory, language, praxis, gnosis, and executive functions; screening	7	• They are standardized tests of psychometric manuals.
		  • Immediate recall: A stimulus is presented to the participant, and he is asked to remember it immediately after having presented it.
		  • Delayed recall: the participant will be given some material to remember, either a list of words or a paragraph. Moreover, at a later time, individuals will be asked to remember those items.
		  • TMT-A[k]: participants are asked to draw a line connecting 25 numbers in sequential order.
		  • TMT-B[l]: this requires participants to connect, in sequential and alternating order, 13 numbers and 12 letters (1, A, 2, B... N and 13). The indicator is the error when connecting the sequence and the time to perform the action.
		  • Stroop task: the names of the colors were displayed incongruously; for example, the color red was written in blue font. Participants had to say the color of the font and not read the word, among several options. The total number of correct answers for each of the 20 items in the test was recorded.
		  • Digital symbol substitution task: consists of pairs of digits and symbols (eg, 1/-, 2/⊥ ... 7/Λ, 8/X, 9/=) followed by a list of digits. Below each digit, the participant must write the corresponding symbol within the allowed time.
		  • Matrix problems task: boxes are shown with certain schemes or figures, in which the individual has to add a missing part, and the individual must choose that part from several options.

CGN_ICA[m] [64]; it correlates with processing speed, learning and visual memory; screening	1	• ICA test: it is a visual characterization task that presents on-screen 100 black-and-white images of animals and places. Each image is displayed for 100 ms, followed by an interval of 20 ms, then a mask with white noise for 250 ms, and to finish, the participant must respond to the image he or she observed. The speed and accuracy of the test are quantified. Four experiments were performed, and only 2 were considered for our study.

Name of the evaluation, test, or battery; cognitive domains; type	Cognitive assessment tests, n	Brief description by activity
VSM[n] [65]; visual memory and verbal, executive functions, attention, and spatial navigation; screening	1	• VSM is based on a daily buying activity. A shopping list appears in the upper-right corner of the screen during exercise. The person is expected to locate the items on this list, place them in the shopping cart, take them to the checkout counter, and pay the correct amount for purchases. The participant is asked to move the shopping cart and navigate inside the VSM by tapping green footprints on the screen. It is an exercise designed to examine multiple cognitive domains (as noted).
CompBased-CAT[o] [66]; attention and concentration, immediate memory, remote memory, executive function, and selective attention and perception of space; screening	6	• Digit span forward and digital span backward: In task 1, participants were asked to immediately remember a set of numbers that were presented in random order on the PC screen. In task 2, participants remembered the same set of numbers in reverse order. The number of digits gradually increased (3, 5, 7, 8). The maximum number of digits a participant could remember in the correct order (forward or backward) was recorded. • Memory of item names and tasks and memory recall of item names: In task 3, participants were asked to memorize 10 target words that were presented on the PC screen. Subsequently, 20 words were shown, including 10 target words and 10 distracting words, and participants were asked to select the 10 target words. This was repeated at 2 attempts, and the number of correct answers was recorded on the second attempt. In addition, participants were instructed to remember the 10 target words after all other tests, and the total number of remembered target words was recorded. • Stroop task: the names of the colors were displayed incongruously; for example, the color red was written in blue font on the PC screen. Participants had to say the color of the font and not read the word among 4 options. The total number of correct answers for each of the 20 items of the task was recorded. • Figure recognition task: in task 6, the blocks were esthetically stacked and displayed on the PC screen. Participants were asked to indicate the number of blocks. We recorded the total number of correct answers for each of the 8 items in the task.
Cognitive Function Balancer (CogEvo) [67]; orientation, attention, memory, executive function, and spatial cognition; specific	5	• Orientation: it is the task of selecting the day, week, and time of the exam and randomly displaying options for questions of the day, week, and time. • Follow the order: the purpose of the task is to select numbers or letters, then touch the screen again, but this time alternating digits with characters according to the logical order. For example, 1, 2, 3; ABC and alternates 1, A, 2, B, 3, C. In total, each question consists of 6 digits, 12 characters and 8 combinations. • Flashing light: a random pattern of lights (red, blue, green, and yellow) must be memorized to follow a certain sequence. The test difficulty level increases, up to 16 combinations of lights that depend on the correct answers. The score is calculated based on the response rate and accuracy. • Route 99: squares are drawn from the starting point to the goal, followed by digits shown at random from 1 to 10. It is forbidden to pass obliquely or travel through the same area. It can be 16, 36, or 64 squares (N×N). • Same shape: the task is to choose a central figure of 6, which is located around. A total of 4 questions are selected, and for each question, a circle must be marked in the figure.
CANTAB[p] [68]; attention and working memory, psychomotor or processing speed, and executive functions; specific	5	• The rapid visual processing subtest: a sustained attention test involving a series of numbers that progress steadily over 6 min, during which the participant must press a button each time a specified 3-digit sequence occurs. • The spatial working memory task: tests patients' ability to retain spatial information and manipulate items remembered in working memory and to do so strategically. It requires the participant to "open" a series of colored boxes to find a tab and then perform a new search to find the next tab. • The reaction time index: measures the amount of time that elapses between the appearance of a stimulus on the screen and the release of a button by the participant (reaction time) and from releasing the button to touching the screen (movement time). • The delayed matching task: it involves a visual pattern that is displayed and then removed, and after an interval of 4-12 s, 4 patterns are displayed, from which the participant must choose the pattern that matches the initial pattern. • The paired frames learning task: it is a series of boxes that appear on the screen, each of which "opens" in turn, and some of which contain a pattern. Subsequently, the participant is shown the patterns and must select the boxes that contain them.
Cognitive Assessment Test (Cognivue) [69]; scores correlate with domains of verbal processing, speed, visual acuity, visuospatial function, orientation, and sequencing; screening	3	

Name of the evaluation, test, or battery; cognitive domains; type	Cognitive assessment tests, n	Brief description by activity

		• Adaptive motor control test: evaluates visuomotor responsiveness using speed and accuracy measurements (adaptive motor control test and visual salience test).
		• Perception processing: measures the perceptual processing of the participant in different ways despite adding increasing patterns of disorder.
		  • Letter discrimination: the participant must discriminate real English letters from others that are similar in a variety of ways to those letters.
		  • Word discrimination: the participant should discriminate real 3-letter words from 3-letter nonwords.
		  • Shape discrimination: discriminates a circle filled with a common shape from the rest of the screen filled with other common shapes.
		  • Motion discrimination: discriminates a circle filled with one direction of point movement from the rest of the full screen with another direction of point movement.
		• Memory processing: evaluates memory using specialized sets of visual stimuli.
		  • Letter memorization: measures the participant's ability to remember which letter was presented as a preindication and then select that letter from a sample of alternative elements, despite the addition of increasing amounts of clutter, and the participant must select the correct letter of the English alphabet.
		  • Word memory: select the correct 3-letter word.
		  • Shape memory: select the correct shape.
		  • Motion memory: select the correct direction of movement.

BHA-CS[q] [70]; episodic memory, executive functions, processing speed, language generation and visuospatial; screening	4	• Favorites (memory domain): measures immediate verbal and visual associative memory and delayed visual memory. Participants are asked to remember people and their favorite foods and animals. In each of the 2 learning tests, 4 different faces were shown twice, once with a favorite food and once with a favorite animal. Accuracy is assessed through 2 immediate recovery attempts and a 10-min delayed recovery test. • Match (executive and speed domain): a series of digits is displayed in the center of the screen, and participants are asked to tap the corresponding image at the bottom of the screen as quickly as possible. Accurate responses are totaled in 2 min. • Line orientation (visuospatial domain): several lines are displayed on the screen, and participants are asked to tap the orange line, which is parallel to the white line. The "angle difference" between the mismatched orange line and the white line is scaled based on the accuracy of the answer, and the scores estimate the angle difference when the probability of a correct answer is between 71% and 75%. • Animal fluency (language): measure categorical verbal fluency. Participants are asked to name as many animals as they can in 1 min.
Inbrain CST[r] [71]; attention, languages, memory (and temporal orientation), executive function, and visuospatial; screening	7	• VST[s]: Nine squares are located on the screen. The squares briefly change color in sequence from 2 to 8, and participants must touch the figure in the same order (task forward) or reverse order (task backward). One point per correct sequence is added, and the score range varies from 1 to 14. The task stops when the participant fails twice. • DNT[t] (confrontation names): it is a task of 15 drawn elements that are relatively difficult to pronounce due to their low frequency of use. • Semantic test: the participant lists as many fruits as possible, and Phonemic test: the participant must list as many words as possible, in both cases, for 1 min. • Block design: a pattern is displayed on the screen, using 6 types of 2-color squares or a combination of them. The participant must reproduce the same pattern by dragging one of the 6 squares. There are 10 patterns, and the difficulty depends on the number of squares and time. • Spatial orientation: the participant must indicate the year, month, date, and day of the week of the current period. WPAT[u]: Nine words are displayed in a 3×3 grid in a given sequence. The examinee must memorize the words and their location in the grid, and then the participant is asked to remember the words immediately and after 10 min and to recognize both the word and its location. • K-TMT- A: the participant is asked to connect numbers (1-15) in ascending order as quickly as possible, using an "S pen" on the tablet. The time taken to complete the test is measured. • K-TMT-B: the participant is asked to connect numbers and the day of the week, alternately and in order, as quickly as possible.
	3	

Name of the evaluation, test, or battery; cognitive domains; type	Cognitive assessment tests, n	Brief description by activity
Trail Making Test y (Bells Test) [72]; psychomotor speed and mental flexibility and executive functions and attention; screening		• TMT-A: participants are asked to draw a line connecting 25 numbers in sequential order. • TMT-B: requires participants to connect, in sequential and alternating order, 13 numbers and 12 letters (1, A, 2, B... N, and 13). The indicator is the error when connecting the sequence and the time to perform the action. • Bells: this test allows the assessment of the attention span through a visual search task; the participant is asked to search and mark 35 objects (black ink bells). The indicator is the number of correct targets.
New C-ABC[v] [73]; sensor-motor ability, attention, orientation, immediate memory, and an arithmetic problem; screening	8	• Touching a moving target: the circle lens was presented in different places on the screen, one at a time, and the participant was asked to touch the circle lens as quickly as possible. • The digits order: 9 digits (1-9) were presented in random positions on the screen, and the participant was asked to tap the digits in sequential order as quickly as possible. • Time orientation: the participant was asked to choose today's date (day, month, year, Japanese era name, and day of the week) from a list of candidates on the screen. • The letters-recognition memory: 4 Japanese letters of hiragana ("ri," "na," "ku," and "me") with meaningless relationships were presented on the screen for 5 s. The participant was then asked to select the 4 recognized Japanese letters from the syllabary. • The numbers-recognition memory: 3 numbers without serial numbers were presented one by one on the screen. After 5 s, the participant was asked to select the 3 numbers in the correct order on the license plate. • The figures-recognition memory: 4 figures with different conditions in color and shape are presented on screen. After 5 s, the participant was asked to select the 4 recognized figures from a set of 12 candidates. • The arithmetic problem: a shopping story was presented on the screen; the participant was asked about the total number of products purchased at 2 stores, and then the participant was asked to select the correct number from a set of options. • Detecting a digit test: a table of random sequences of digits is presented on the screen. It is asked to detect and touch the digit (3) of the table, and there are 4 elements of digit 3.
EC-Screen[w] [74]; executive functions, visuospatial skills, mental flexibility, and memory function; screening	3	• Clock test: requires the participant to place a specific time by moving the hands of a digital clock. • Proof of history: it is a fact-conversion test based on the history of a known landmark in the country. The platform reads a short story, and the participant must remember the facts and identify the facts based on questions. • Deferred recognition test: the participant must learn and remember 5 words from 2 syllables that are read by the software.

BHA [75]; episodic memory, attention and working memory, executive functions, visuospatial skills, and language; screening	4	• The same tests on BHA battery were already mentioned [70]:
		  • Favorites (memory domain): measures immediate verbal and visual associative memory and delayed visual memory. Participants are asked to remember people (faces) and their favorite foods and animals. Accuracy is assessed through 2 immediate recovery attempts and a 10-min delayed recovery test.
		  • Match (executive and speed domain): a series of digits is displayed in the center of the screen, and participants are asked to tap the corresponding image at the bottom of the screen as quickly as possible. Accurate responses are totaled in 2 min.
		  • Line orientation (visuospatial domain): several lines are displayed on the screen, and participants are asked to tap the orange line, which is parallel to the white line. The scores estimate the "angle difference" between the mismatched orange line and the white line when the probability of a correct answer is between 71% and 75%.
		  • Animal fluency (language): measure categorical verbal fluency. Participants are asked to name as many animals as they can in 1 min.

Smart Aging Platform (serious game; IADL) [76]; executive, verbal, spatial function; screening	5	

Name of the evaluation, test, or battery; cognitive domains; type	Cognitive assessment tests, n	Brief description by activity
		• It is a game where the participant experiences a virtual environment resembling a loft; the objective of the activity is to perform tasks, and the system records the data of the positions, times, actions, etc. To quantify cognitive functions, the system provides scores and calculates indexes (accuracy, time, and distance). Time is timed from start to finish. The distance is the number of meters traveled. Usually, people who take longer to perform tasks have a diagnosis of mild cognitive impairment. • Object search: the participant is asked to search for a list of objects. • Water flowers while listening to the radio: participants are asked to turn on the radio and press the space bar each time they hear the word "sun" while watering the flowers. • Make a phone call: the participant is asked to make a phone call using the phone book and then turn on the television. • Choose the right object: a screen displays 24 images of objects. The participant must identify the 12 objects presented in task 1. • Find the objects: the participant is placed in front of the kitchen and asked to look for each of the objects.

[a]TOL: Tower of London.

[b]eSAGE: Self-Administered Gerocognitive Examination.

[c]CoCosc: Computerized Cognitive Screen.

[d]PM: prospective memory.

[e]CVLT: California Verbal Learning Test.

[f]SIMBAC: Simulation-Based Assessment of Cognition.

[g]IADL: Independent Activities of Daily Living.

[h]ATM: automated teller machine.

[i]UCSF: University of California, San Francisco.

[j]BHA: Brain Health Assessment.

[k]TMT-A: Trail Making Test-A.

[l]TMT-B: Trail Making Test-B.

[m]CGN_ICA: Cognitivity Neurosciences-Integrated Cognitive Assessment.

[n]VSM: Virtual Supermarket.

[o]CompBased-CAT: Computer-Based Cognitive Assessment Tool.

[p]CANTAB: Cambridge Neuropsychological Test Automated Battery.

[q]BHA-CS: Brain Health Assessment-Cognitive Score.

[r]CST: Cognitive Screening Test.

[s]VST: visual span test.

[t]DNT: difficult naming test.

[u]WPAT: word place association test.

[v]C-ABC: computerized assessment battery for cognition.

[w]EC-Screen: Electronic Cognitive Screen.

The different applications also differ depending on whether the tests are of the screening type or are evaluation tests of specific cognitive domains. The number of tests varies depending on the domains required for clinical evaluation. In general, at least 2 cognitive domains, such as language, praxis, gnosis, or executive functions, are necessary to identify dementia. In addition, a test can involve subdomains. For example, the Episodic Assessment is a modified version of the CVLT and incorporates gamification techniques to assess episodic memory. The advantage of the game is that it allows for the coverage of more elements of executive function rather than those involving simply remembering words verbally [60,63,64].

RQ4: What Are the Main Effects, Personal Traits, and Psychometric Parameters Considered for the Validation of Digitized Cognitive Tests for Older Adults?

Main Effects

All the articles focused on the capacity and accuracy of the test to measure the cognitive level of the patient. However, it has been overlooked in most cases that unlike paper-and-pencil tests, computerized cognitive tests are software programs. Therefore, usability (ease of use) and human-computer interaction dimensions are crucial to their overall effectiveness. Only 2 articles from the review used usability questionnaires [60,72]. Thus, the inclusion of usability tests in cognitive evaluation software is expected in future work, either through usability tests with users or through the evaluation of experts through appropriate usability heuristics.

In addition, digital tests could be affected by the level of digital literacy of older adults or by their aversion to new technologies. People have different levels of skills and use of information and communication technologies, which makes it necessary to assess the extent to which they accept technology. Depending on the user, the degree of information and communication technology acceptance can positively or negatively affect the use of a particular piece of software, such as a shopping website, a video game, or, in our specific scenario, a digital cognitive test. Although this factor may be more relevant when dealing with older adults, it is worth noting that only one of the analyzed studies [72] addressed this aspect. It would be interesting to include the Model of Acceptance and Adoption of Technology in Older Adults, known as Senior Technology Acceptance Model [81,82], in future studies. This model measures various aspects such as intention to use, perceived usefulness, ease of learning, and actual use. By evaluating how diverse levels of technology acceptance affect digital cognitive test performance, we can gain valuable insights.

Other related effects that can be measured in older adults include psycho-emotional aspects such as motivation, effort, tension, and anxiety. On the one hand, there is the overall emotional level of the person, and on the other hand, there is the emotional level that is generated when using a cognitive digital test. One factor to consider is an individual's level of motivation on a given day. Another factor is the degree of motivation specifically for taking a digital cognitive test, as well as the level of effort exerted when performing each task within the test. The study by Valladares-Rodriguez et al [60] included a motivation test and that by Eraslan Boz et al [65] applied a test for anxiety and depression, but only descriptively. They did not evaluate the effects of these emotional factors on cognitive test performance nor did they compare the results between conditions with different levels of anxiety, motivation, or depression. Therefore, we consider it important to measure and evaluate the possible moderating role of diverse emotional factors in future studies, particularly when examining games, simulators, and other immersive software for cognitive evaluation.

Another aspect that has received little attention is the accessibility of software, including sensory (vision and hearing) and physical accessibility. It is necessary to consider potential users who may have visual impairment, illiteracy, or movement disorders that can make it difficult to administer the digital test through a tablet or mobile device [63].

Personal Traits

Regarding the main personal traits considered in cognitive tests, the first criterion is to group the participants based on their cognitive ability (eg, normal, MCI, and dementia). Previous surveys mentioned the need to cover the different personal traits of older adults. The most frequently mentioned traits were age, educational level (years), and gender. However, less attention has been given to language, ethnicity, and cultural aspects. For example, the review by Aslam et al [54] includes a comprehensive analysis of the population, considering cognitive level and nationality as well as demographic factors such as age, gender, and educational level. The review by Zygouris and Tsolaki [53] highlights the difficulty that some tests are not adapted for languages other than English and that the results of certain tests are influenced by an individual's educational level. Tsoy et al [56] includes statistics about the age and years of education of the selected papers but does not consider sex or cultural aspects, although it does highlight that the different tests are not available in several languages. The survey by Marques-Costa et al [55] included data about age, gender, and educational level, but this information was presented in narrative form rather than a comparative table. This suggests that further research is needed to examine specific patient populations of different age groups and educational levels. In addition, the review by Chan et al [83] does not include demographic or educational data.

When analyzing the reviewed papers in our study, we categorized the 20 papers into 5 groups. The first group [60,62,63] did not include any demographic variables. The second group [57,66,72,76] includes demographic and comparative tables of results by cognitive condition but lacks further analysis or discussion of the participants' demographic characteristics. Other studies (third group) conducted an initial analysis of the personal traits of older adults without making statistical adjustments or comparing these characteristics with associated groups. For example, the study by Khaligh-Razavi et al [64] included a demographic table and raised the issue of test dependency on the level of education. In addition, Cahn-Hidalgo et al [69] mentioned that there were biases related to education, language, gender, and culture, but they claimed that their proposed cognitive test was free from these biases. However, no further empirical evidence was provided to support these test features.

A fourth group of studies statistically adjusted the results of the cognitive tests based on demographic characteristics. For example, the studies by Scharre et al [58], Rapp et al [61], and Eraslan Boz et al [65] included a demographic table and considered adjustments based on age and education. However, the study by Rapp et al [61] did not include the years of study in the table. In the study by Schulz et al [68], the results were adjusted for age using Heaton-Revised Norms, while the study by Chan et al [74] adjusted for participants with fewer years of education and provided an "illiterate" version of the test, allowing for more time during administration. In addition, Rodriguez-Salgado et al [75], on the basis of sample controls, they adjusted for age, education, and sex using a regression-based approach, arguing that assessments are complicated by linguistic, ethnic, cultural, and socioeconomic diversity.

Finally, the fifth group performed statistical comparisons by considering different demographic characteristics (older adults' traits). For example, Wong et al [59] presented a comparative table showing the results of cognitive subtests based on the level of education. Ichii et al [67] compared test results based on age and cognitive groups. Eraslan Boz et al [65] evaluated the effects of age, gender, and education. Tsoy et al [70] considered adjustments for age, education, sex, and language, in addition to performing a linear regression considering sex and age. In the study Chin et al [71], ANOVA and Pearson χ^2 test were used to analyze the age, education, and gender of the 3 cognitive

groups. They mentioned that the results were significantly affected by age, gender, and years of education (especially in normal participants). Although Noguchi-Shinohara et al [73] applied the Kruskal-Wallis test and χ^2 test to compare age groups among cognitive conditions, they did not observe any significant differences between gender, age, and education. To summarize, diverse personal traits have been recognized in the existing literature as relevant factors that affect cognitive test performance. Some of the reviewed studies assessed the effects of age, gender, and educational level, whereas none of them considered statistical comparisons of cultural factors. Therefore, further studies are needed to delve deeper into the effects of diverse personal traits on cognitive test performance.

Psychometric Parameters

The main psychometric parameters considered were the different validation processes and measurement parameters presented in this study. A comparative matrix was created for the 20 selected articles by considering the available data and the mentioned psychometric parameters. The data included the type of population under observation, sensitivity, specificity, criteria for discriminating and comparing types of participants, and possible correlations for assessing validity.

Concerning the psychometric validation process, Table S1 Multimedia Appendix 1 [42,44-47,49,57-76] displays the key parameters, including sensitivity, specificity, area under the curve, and correlation validity, with some traditional tests (on paper) such as the MoCA, MMSE, and CVLT. This table was the most difficult to prepare because some tests had nonhomogeneous validation parameters.

In addition, other comparison variables are as follows (columns of the table in Appendix 1):

1. N (normal) is the number of people diagnosed and categorized as a control group based on clinical criteria (or people who are normally healthy or without cognitive impairment).
2. MCI indicates the number of people diagnosed with MCI based on clinical criteria (such as traditional psychometric tests or interviews).
3. DM (dementia) denotes the number of people evaluated and diagnosed with a certain degree of dementia based on clinical criteria (such as traditional psychometric tests or interviews).
4. OT (other pathologies) is the number of people diagnosed with other pathologies, such as Parkinson disease, Alzheimer disease, or stroke.
5. CA-DISCR (capacity to discriminate) indicates the objective of the digital test, which is to assess the instruments' capacity to discriminate and differentiate people to validate the tests, for example, normal versus dementia or normal versus MCI.
6. SENSI: sensitivity (the proportion of people with a disease to be diagnosed who will have a positive result) has already been explained.
7. SPEC: specificity (the proportion of people without a disease to be diagnosed who will have a negative result) has already been explained.
8. ROC area is the statistical method (based on sensitivity and specificity) to evaluate the ability of a test to discriminate between individuals with and without a disease.
9. COM-VAL (comparison and validation): the traditional test (or tests), such as MoCA, MMSE, or others, used in the validation process of the digital test are identified. The instrument proposed in the article is validated by comparing it with traditional tests through statistics of the correlation coefficient used (such as Pearson, linear regression, or others). A correlation is then established (for the validity of the construct).
10. VALID-CO (validity and correlation): the correlation coefficients indicate the relationship of the test with the various (traditional) measurement instruments and their criteria, as well as the relationship that exists between the test and the construct. Construct validity refers to the adequacy of inferences made from observations or measurements (often the test results). Specifically, it determines whether a test measures the intended construct. The validity of a test indicates the extent to which it accurately measures the theoretical construct it is designed to measure and whether it can be used effectively for its intended purpose. A test is considered valid if it accurately measures what it claims to measure.

Although numerous modalities or methodologies are available to evaluate a patient, some of these applications lack sufficient studies on validity, reliability, and precision. In addition, the population sample for these modalities may be very small or may simply lack studies to support their implementation. Three basic, but complex, statistical tables were created, and they aimed to compare the validity parameters and the correlation of the digital applications with their traditional equivalent tests. Heterogeneity among the methods was significant. Most of the studies performed previous tests to categorize the participants into groups (using tools such as the MMSE, MoCA, or other specialized instruments) and subsequently validated the results through a correlation analysis. However, some applications are evaluated according to specific tests (tasks) rather than groups of patients with different characteristics. Some applications evaluated their area under the curve on age ranges [73], whereas others assessed it based on groups categorized by pathologies (such as healthy, MCI, dementia, or other combinations) [58,61,62,67,71,73-75].

In summary, the most usual values for validity cases were as follows:

- Sensitivity ranged from 0.63 to 0.95. With the exception of 2 very extreme cases, the lowest error rate was observed with the Tower of London application [57] when comparing errors made among patients with MCI (0.379), whereas the highest error rate was observed with the University of California, San Francisco application [62] when comparing healthy patients to those with dementia (1.00). The applications with the highest rates of sensitivity were the computerized platforms BHA [70,75], as well as the digital version eSAGE [58].
- The specificity ranged from 0.54, as reported by the CogEvo computerized battery [67], to 1.00, as observed in the eSAGE application [58], when comparing healthy

individuals to those with dementia. The highest average among the different comparisons corresponded to the TMT-A and TMT-B tests [72].
- The computerized cognitive platforms that presented a higher ROC curve were the University of California, San Francisco [62] and BHA [75] applications.

For the correlations, the most frequent cases of validation were observed with the traditional tests, specifically 10 with the MMSE and 8 with the MoCA. Some researchers validated their results using specialized tests such as the CVLT [62], Addenbrooke's Cognitive Examination-Revised [64], Repeatable Battery for the Assessment of Neuropsychological Status [68], and other specific measures [69,75]. The cases with the highest correlations corresponded to the computerized assessment battery for cognition [73], eSAGE [58], BHA [75], Inbrain [71] and Cognivue [69] applications, with average values of 0.75, 0.77, 0.77, 0.85, and 0.86, respectively. It is noteworthy that in 4 articles, evaluations of the sensitivity and specificity of the results were not included [64,68,69,76].

Discussion

Principal Findings

In this work, the systematic review and subsequent analysis of 20 recent articles were performed with technological alternatives to determine "the state of the art" of cognitive assessment tests in digital format. Of the 20 articles selected from 2015 onward, 14 (70%) were from 2019 to 2021, which demonstrates the growing interest in developing research on digital cognitive evaluation and training, particularly for older adults.

The *Results* section includes a discussion of each of the 4 research questions. An analysis of previous literature reviews yielded 4 papers published from 2015 to 2021. Our work provides a more up-to-date state-of-the-art approach, focusing on a wider range of cognitive domains. In addition to considering computer tests and digital test batteries, we included some initial studies on emerging technologies, such as VR, video games, gamification, and AI.

Computerized tests (and tasks), similar to paper-and-pencil tests, were found to generally assess the same domains but with different modalities. The most commonly used type is test batteries, which are administered on computers, tablets, or web platforms. There were also a few cases involving daily-life activity simulators, digital games, VR, and machine learning (AI). The use of these new technologies is expected to increase in the near future. Digital tests present improvements in completion place and time, do not require a structured system, and are portable, allowing cognitive evaluations to be carried out in the comfort of one's home or community centers with minimal intervention from a specialist professional.

In addition, a well-designed digitized cognitive test allows for the reduction of "learning bias" by using pools of tasks to avoid repetition. In addition, it has the potential to be adapted to various contexts and people with different characteristics, thereby improving personalization and accessibility. However, these characteristics require further investigation.

Regarding the characteristics provided by the different digital tests, this work extends past reviews by providing details about the instructions and how they are delivered to the patient, the environment or place where the test is performed, and the sequence of tasks that the participants must perform on each of the tests. In addition, to enhance comparability among tests covering similar cognitive domains, we included their sample size, the details of how many participants had MCI or dementia, and the psychometric parameters used for each test in their comparative tables.

With regard to test validation, there is usually sufficient correlation between digital tests and their traditional paper-and-pencil counterparts, which validates their effectiveness in most cases. However, in many studies, the sample size of adults was small, and the psychometric parameters used were too heterogeneous, making comparisons difficult. It is also necessary to conduct further studies that consider differences in education level, gender, ethnicity, and culture.

Finally, most studies assume that the older adults use technology. However, an older person may not be familiar with the use of computers, tablets, or smartphones and may even need assistance in some cases. Therefore, it is necessary to incorporate other factors in future studies on digital tests, such as the degree of technology acceptance (Senior Technology Acceptance Model) as well as the emotional aspects of the participants, such as motivation and effort, along with usability evaluations.

Acknowledgments

We thank the School of Computer Engineering at the Pontificia Universidad Católica de Valparaíso and the Research Center for the Development of Cognition and Language at the Universidad de Valparaíso, Chile, for providing the facilities for the development of this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement parameters of the psychometric validation process.
[PDF File (Adobe PDF File), 68 KB - mental_v10i1e47487_app1.pdf]

Multimedia Appendix 2
PRISMA Checklist.
[PDF File (Adobe PDF File), 68 KB - mental_v10i1e47487_app2.pdf]

References

1. Cancino M, Rehbein L. Factores de riesgo y precursores del Deterioro Cognitivo Leve (DCL): Una mirada sinóptica. Ter Psicol 2016 Dec;34(3):183-189 [FREE Full text] [doi: 10.4067/s0718-48082016000300002]
2. Mathers CD, Stevens GA, Boerma T, White RA, Tobias MI. Causes of international increases in older age life expectancy. Lancet 2015 Feb 07;385(9967):540-548. [doi: 10.1016/S0140-6736(14)60569-9] [Medline: 25468166]
3. Staff PAHO. Crecimiento acelerado de la población adulta de 60 años y más de edad: Reto para la salud pública. Pan American Health Organization. 2012 Sep 28. URL: https://www.paho.org/en/news/28-9-2012-crecimiento-acelerado-poblacion-adulta-60-anos-mas-edad-reto-para-salud-publica [accessed 2023-08-21]
4. Fuentes P, Albala C. An update on aging and dementia in Chile. Dement Neuropsychol 2014;8(4):317-322 [FREE Full text] [doi: 10.1590/S1980-57642014DN84000003] [Medline: 29213920]
5. Villalobos Dintrans P. Panorama de la dependencia en Chile. Rev Med Chil 2019;147(1):83-90 [FREE Full text] [doi: 10.4067/S0034-98872019000100083] [Medline: 30848769]
6. Organización Panamericana de la Salud, Comisión Económica para América Latina y el Caribe de las Naciones Unidas. Perspectivas demográficas del envejecimiento poblacional en la Región de las Américas. Washington, DC: Pan American Health Organization; 2023.
7. Albala C. El envejecimiento de la población chilena y los desafíos para la salud y el bienestar de las personas mayores. Revista Médica Clínica Las Condes 2020 Jan;31(1):7-12 [FREE Full text] [doi: 10.1016/j.rmclc.2019.12.001]
8. Staff MINSAL. Programa nacional de salud integral para personas mayores. Ministerio de Salud de Chile. 2021. URL: https://www.minsal.cl/wp-content/uploads/2021/11/Plan-Nacional-de-Salud-Integral-para-Personas-Mayores.pdf [accessed 2023-08-25]
9. Gaete M, Jorquera S, Bello-Lepe S, Mendoza Y, Véliz M, Alonso-Sanchez M, et al. Resultados estandarizados del Montreal Cognitive Assessment (MoCA) para cribado neurocognitivo en población chilena. Neurologia (Engl Ed) 2020 Nov 05;38(4):246-255 [FREE Full text] [doi: 10.1016/j.nrl.2020.08.017] [Medline: 33160725]
10. Prince MJ, Wu F, Guo Y, Gutierrez Robledo LM, O'Donnell M, Sullivan R, et al. The burden of disease in older people and implications for health policy and practice. Lancet 2015 Feb 07;385(9967):549-562. [doi: 10.1016/S0140-6736(14)61347-7] [Medline: 25468153]
11. Prince MJ. World Alzheimer report 2015: the global impact of dementia. Alzheimer Disease International. 2015. URL: https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf [accessed 2023-11-24]
12. Mogollón GE. Una propuesta para el mejoramiento cognitivo en el adulto mayor: una alternativa al entrenamiento cerebral. Rev Electron Educ 2014 Apr 02;18(2):1-17 [FREE Full text] [doi: 10.15359/ree.18-2.1]
13. Conway AR, Getz S. Cognitive ability: does working memory training enhance intelligence? Curr Biol 2010 Apr 27;20(8):R362-R364 [FREE Full text] [doi: 10.1016/j.cub.2010.03.001] [Medline: 21749957]
14. Morrison AB, Chein JM. Does working memory training work? The promise and challenges of enhancing cognition by training working memory. Psychon Bull Rev 2011 Feb 17;18(1):46-60 [FREE Full text] [doi: 10.3758/s13423-010-0034-0] [Medline: 21327348]
15. Shipstead Z, Redick TS, Engle RW. Is working memory training effective? Psychol Bull 2012 Jul;138(4):628-654. [doi: 10.1037/a0027473] [Medline: 22409508]
16. Romero Á, Fernandez EF, Argüelles JL, Piña JD, Jiménez JD. Updating the Diagnosis and Management of Dementia. Practical Contributions in our Context. Rev. Revista Finlay 2012 Jan;2(1):76 [FREE Full text]
17. Tong T, Chignell M. Developing a serious game for cognitive assessment: choosing settings and measuring performance. In: Proceedings of the 2nd International Symposium of Chinese CHI. 2014 Presented at: Chinese CHI '14; April 26-27, 2014; Toronto, ON p. 70-79 URL: https://dl.acm.org/doi/10.1145/2592235.2592246 [doi: 10.1145/2592235.2592246]
18. Woodford H, George J. Cognitive assessment in the elderly: a review of clinical methods. QJM Mon J Assoc Physicians 2007 Aug 02;100(8):469-484. [doi: 10.1093/qjmed/hcm051] [Medline: 17566006]
19. Mikulic I. Construccion y adaptacion de pruebas psicologicas; Test de Exploración y Diagnóstico, Cátedra I - Módulo I. Facultad de Psicología; Universidad de Buenos Aires. URL: https://pdf4pro.com/amp/view/construccion-y-adaptacion-de-pruebas-17911d.html [accessed 2023-11-24]
20. Zeng Z, Wang D. Towards long-term tracking and detection of early dementia: a computerized cognitive test battery with gamification. In: Proceedings of the 3rd International Conference on Crowd Science and Engineering. 2018 Presented at: ICCSE '18; July 28-31, 2018; Singapore, Singapore p. 1-10 URL: https://dl.acm.org/doi/10.1145/3265689.3265719 [doi: 10.1145/3265689.3265719]
21. Petersen RC. Mild cognitive impairment clinical trials. Nat Rev Drug Discov 2003 Aug;2(8):646-653. [doi: 10.1038/nrd1155] [Medline: 12904814]

22. Doniger G, Zucker D, Schweiger A, Dwolatzky T, Chertkow H, Crystal H, et al. Towards practical cognitive assessment for detection of early dementia: a 30-minute computerized battery discriminates as well as longer testing. Curr Alzheimer Res 2005 Apr 01;2(2):117-124. [doi: 10.2174/1567205053585792] [Medline: 15974907]
23. Darby D, Maruff P, Collie A, McStephen M. Mild cognitive impairment can be detected by multiple assessments in a single day. Neurology 2002 Oct 08;59(7):1042-1046. [doi: 10.1212/wnl.59.7.1042] [Medline: 12370459]
24. Dwolatzky T, Whitehead V, Doniger GM, Simon ES, Schweiger A, Jaffe D, et al. Validity of a novel computerized cognitive battery for mild cognitive impairment. BMC Geriatr 2003 Nov 02;3(1):4 [FREE Full text] [doi: 10.1186/1471-2318-3-4] [Medline: 14594456]
25. López M. Cómo Entrenar tu Cerebro. Barcelona, España: Ediciones B; 2015.
26. Funciones cognitivas: Plataforma web de Neurorehabilitación. NeuroUP. 2019. URL: https://www.neuronup.com/es/areas/functions [accessed 2023-11-24]
27. Manes F, Niro F. Usar el Cerebro, Conocer Nuestra Mente Para Vivir Mejor. Buenos Aires, Argentina: Planeta; 2017.
28. Atención Proceso Cognitivo. CogniFit Inc. 2021. URL: https://www.cognifit.com/cl/atencion [accessed 2023-11-24]
29. Marco Teórico: Conceptos Generales. NeuroUP. 2019. URL: https://app.neuronup.com/assets/pdf/theoretical_framework/theoretical_framework_es_ES.pdf [accessed 2023-11-24]
30. Soto K. La memoria Humana. Libro sobre Memoria. 2015. URL: https://www.mheducation.es/bcv/guide/capitulo/8448180607.pdf [accessed 2023-11-24]
31. Muñoz-Neira C, Henríquez Ch F, Ihnen J J, Sánchez C M, Flores M P, Slachevsky Ch A. Propiedades psicométricas y utilidad diagnóstica del Addenbrooke's Cognitive Examination-Revised (ACE-R) en una muestra de ancianos chilenos. Rev Méd Chile 2012 Aug;140(8):1006-1013. [doi: 10.4067/s0034-98872012000800006]
32. Burga A. Applications of Technology to Psychometric Evaluation. Propós Represent 2019;8(1):1-11 [FREE Full text] [doi: 10.20511/pyr2019.v7nSPE.318]
33. Roebuck-Spencer TM, Glen T, Puente AE, Denney RL, Ruff RM, Hostetter G, et al. Cognitive screening tests versus comprehensive neuropsychological test batteries: a national academy of neuropsychology education paper. Arch Clin Neuropsychol 2017 Jun 01;32(4):491-498. [doi: 10.1093/arclin/acx021] [Medline: 28334244]
34. Silva H, Cerejeira J. Cow milking game: evaluating a serious game for cognitive stimulation with an elderly population. In: Proceedings of the International Symposium on Interactive Technology and Ageing Populations. 2016 Presented at: ITAP '16; October 20-22, 2016; Kochi, Japan p. 44-53 URL: https://dl.acm.org/doi/10.1145/2996267.2996272 [doi: 10.1145/2996267.2996272]
35. Screening for cognitive impairment in older adults. Mini-Cog©. URL: https://mini-cog.com [accessed 2023-11-24]
36. Bello-Lepe S, Alonso-Sánchez M, Ortega A, Gaete M, Veliz M, Lira J, et al. Montreal cognitive assessment as screening measure for mild and major neurocognitive disorder in a Chilean population. Dement Geriatr Cogn Disord Extra 2020 Oct 14;10(3):105-114. [doi: 10.1159/000506280] [Medline: 33250917]
37. Reyes M. Valoración Geriátrica en APS Enfermera: María del Carmen Reyes. Ministerio de Salud de Chile. 2015. URL: https://www.minsal.cl/wp-content/uploads/2015/09/Enfermera-Mar%C3%ADa-del-Carmen-Reyes-VGI-APS.pdf [accessed 2023-11-24]
38. Abbreviated Mental Test (AMT). Patient. URL: https://patient.info/doctor/abbreviated-mental-test-amt [accessed 2023-11-24]
39. Qué es el test de Stroop y su utilidad en la evaluación del TDAH. Fundación CADAH. URL: https://www.fundacioncadah.org/web/articulo/que-es-el-test-de-stroop-y-su-utilidad-en-la-evaluacion-del-tdah.html [accessed 2023-11-24]
40. Guevara M, Hernández M, Hevia J. Memoria de trabajo visoespacial evaluada a través de los Cubos de Corsi. Rev Esp Neuropsicol 2014;14(1):208-222 [FREE Full text]
41. El trail making test. Teoría y Técnica de la Evaluación Psicológica. URL: https://evaluacionpsico.wordpress.com/2012/11/21/trail-making-test/ [accessed 2023-11-24]
42. Test de la Figura Compleja de Rey: ¿Qué evalúa y cómo se aplica? Psicólogos. URL: https://psicologoscordoba.org/test-de-la-figura-compleja-de-rey-como-se-aplica/ [accessed 2023-11-24]
43. Ruiz L. Qué es la psicometría? PSYCIENCIA. 2022. URL: https://www.psyciencia.com/que-es-la-psicometria/ [accessed 2023-11-24]
44. Ardila A, Ostrosky F. GUÍA PARA EL DIAGNÓSTICO NEUROPSICOLÓGICO, Material de Estudio, American Board of Professional Neuropsychology Estado de la Florida. 2012. URL: https://www.researchgate.net/publication/266558385_guia_para_el_diagnostico_neuropsicologico [accessed 2023-11-24]
45. Chiner E. Tema 6: La Validez; material de estudio, Repositorio Institucional de la Universidad de Alicante; España. Institutional Repository of the University of Alicante. URL: https://rua.ua.es/dspace/bitstream/10045/19380/25/Tema%206-Validez.pdf [accessed 2023-11-24]
46. Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. Acta Paediatrica 2007 Feb 23;96(3):338-341. [doi: 10.1111/j.1651-2227.2006.00180.x]
47. Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. Acta Paediatrica 2007 Mar 21;96(5):644-647. [doi: 10.1111/j.1651-2227.2006.00178.x]
48. Sharma T, Bilder R. Standardisation Study of Cogtest ; Neurocognitive Battery for Clinical Trials. academia.edu. 2005. URL: https://www.academia.edu/2894704/standardisation_Study_of_Cogtest [accessed 2023-03-14]

49. Robbins T, James M, Owen A, Sahakian B, McInnes L, Rabbitt P. Cambridge neuropsychological test automated battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. Dement Geriatr Cogn Disord 2010 May 26;5(5):266-281. [doi: 10.1159/000106735]
50. Mide tu Salud Mental. Cognifit Inc. URL: https://www.cognifit.com/es [accessed 2023-11-24]
51. Home page. NeuronUP. URL: https://www.neuronup.com/es [accessed 2023-11-24]
52. APP de estimulación cognitiva. Stimulus. URL: https://stimuluspro.com [accessed 2023-11-24]
53. Zygouris S, Tsolaki M. Computerized cognitive testing for older adults: a review. Am J Alzheimers Dis Other Demen 2015 Feb 13;30(1):13-28 [FREE Full text] [doi: 10.1177/1533317514522852] [Medline: 24526761]
54. Aslam RW, Bates V, Dundar Y, Hounsome J, Richardson M, Krishan A, et al. A systematic review of the diagnostic accuracy of automated tests for cognitive impairment. Int J Geriatr Psychiatry 2018 Apr 22;33(4):561-575 [FREE Full text] [doi: 10.1002/gps.4852] [Medline: 29356098]
55. Marques-Costa C, Almiro P, Simões M. Computerized Cognitive Tests (CCT) in elderly: a psychometric review. Eur Rev Appl Psychol 2018 Mar;68(2):61-68 [FREE Full text] [doi: 10.1016/j.erap.2018.04.002]
56. Tsoy E, Zygouris S, Possin K. Current state of self-administered brief computerized cognitive assessments for detection of cognitive disorders in older adults: a systematic review. J Prev Alzheimers Dis 2021;8(3):267-276 [FREE Full text] [doi: 10.14283/jpad.2021.11] [Medline: 34101783]
57. Köstering L, Schmidt CS, Weiller C, Kaller CP. Analyses of rule breaks and errors during planning in computerized tower tasks: insights from neurological patients. Arch Clin Neuropsychol 2016 Nov 22;31(7):738-753. [doi: 10.1093/arclin/acw059] [Medline: 27590302]
58. Scharre DW, Chang SI, Nagaraja HN, Vrettos NE, Bornstein RA. Digitally translated Self-Administered Gerocognitive Examination (eSAGE): relationship with its validated paper version, neuropsychological evaluations, and clinical assessments. Alzheimers Res Ther 2017 Jun 27;9(1):44 [FREE Full text] [doi: 10.1186/s13195-017-0269-3] [Medline: 28655351]
59. Wong A, Fong C, Mok VC, Leung K, Tong RK. Computerized cognitive screen (CoCoSc): a self-administered computerized test for screening for cognitive impairment in community social centers. J Alzheimers Dis 2017;59(4):1299-1306. [doi: 10.3233/JAD-170196] [Medline: 28731437]
60. Valladares-Rodriguez S, Perez-Rodriguez R, Facal D, Fernandez-Iglesias MJ, Anido-Rifon L, Mouriño-Garcia M. Design process and preliminary psychometric study of a video game to detect cognitive impairment in senior adults. PeerJ 2017;5:e3508 [FREE Full text] [doi: 10.7717/peerj.3508] [Medline: 28674661]
61. Rapp SR, Barnard RT, Sink KM, Chamberlain DG, Wilson V, Lu L, et al. Computer simulations for assessing cognitively intensive instrumental activities of daily living in older adults. Alzheimers Dement (Amst) 2018 Feb 22;10(1):237-244 [FREE Full text] [doi: 10.1016/j.dadm.2018.01.008] [Medline: 29560411]
62. Possin KL, Moskowitz T, Erlhoff SJ, Rogers KM, Johnson ET, Steele NZ, et al. The brain health assessment for detecting and diagnosing neurocognitive disorders. J Am Geriatr Soc 2018 Jan 22;66(1):150-156 [FREE Full text] [doi: 10.1111/jgs.15208] [Medline: 29355911]
63. Groppell S, Soto-Ruiz KM, Flores B, Dawkins W, Smith I, Eagleman DM, et al. A rapid, mobile neurocognitive screening test to aid in identifying cognitive impairment and dementia (BrainCheck): cohort study. JMIR Aging 2019 Mar 21;2(1):e12615 [FREE Full text] [doi: 10.2196/12615] [Medline: 31518280]
64. Khaligh-Razavi S, Habibi S, Sadeghi M, Marefat H, Khanbagi M, Nabavi SM, et al. Integrated cognitive assessment: speed and accuracy of visual processing as a reliable proxy to cognitive performance. Sci Rep 2019 Jan 31;9(1):1102 [FREE Full text] [doi: 10.1038/s41598-018-37709-x] [Medline: 30705371]
65. Eraslan Boz H, Limoncu H, Zygouris S, Tsolaki M, Giakoumis D, Votis K, et al. A new tool to assess amnestic mild cognitive impairment in Turkish older adults: virtual supermarket (VSM). Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2020 Sep;27(5):639-653 [FREE Full text] [doi: 10.1080/13825585.2019.1663146] [Medline: 31482749]
66. Takahashi J, Kawai H, Suzuki H, Fujiwara Y, Watanabe Y, Hirano H, et al. Development and validity of the Computer-Based Cognitive Assessment Tool for intervention in community-dwelling older individuals. Geriatr Gerontol Int 2020 Mar;20(3):171-175 [FREE Full text] [doi: 10.1111/ggi.13836] [Medline: 31916344]
67. Ichii S, Nakamura T, Kawarabayashi T, Takatama M, Ohgami T, Ihara K, et al. CogEvo, a cognitive function balancer, is a sensitive and easy psychiatric test battery for age-related cognitive decline. Geriatr Gerontol Int 2020 Mar;20(3):248-255 [FREE Full text] [doi: 10.1111/ggi.13847] [Medline: 31851431]
68. Schulz-Heik RJ, Fahimi A, Durazzo TC, Friedman M, Bayley PJ. Evaluation of adding the CANTAB computerized neuropsychological assessment battery to a traditional battery in a tertiary care center for veterans. Appl Neuropsychol Adult 2020 Jan 11;27(3):256-266 [FREE Full text] [doi: 10.1080/23279095.2018.1534735] [Medline: 30633552]
69. Cahn-Hidalgo D, Estes PW, Benabou R. Validity, reliability, and psychometric properties of a computerized, cognitive assessment test (Cognivue). World J Psychiatry 2020 Jan 19;10(1):1-11 [FREE Full text] [doi: 10.5498/wjp.v10.i1.1] [Medline: 31956523]
70. Tsoy E, Erlhoff SJ, Goode CA, Dorsman KA, Kanjanapong S, Lindbergh CA, et al. BHA-CS: a novel cognitive composite for Alzheimer's disease and related disorders. Alzheimers Dement (Amst) 2020 Jun 21;12(1):e12042 [FREE Full text] [doi: 10.1002/dad2.12042] [Medline: 32582835]

71. Chin J, Kim DE, Lee H, Yun J, Lee BH, Park J, et al. A validation study of the inbrain CST: a tablet computer-based cognitive screening test for elderly people with cognitive impairment. J Korean Med Sci 2020 Aug 31;35(34):e292 [FREE Full text] [doi: 10.3346/jkms.2020.35.e292] [Medline: 32864906]
72. Lunardini F, Luperto M, Romeo M, Basilico N, Daniele K, Azzolino D, et al. Supervised digital neuropsychological tests for cognitive decline in older adults: usability and clinical validity study. JMIR Mhealth Uhealth 2020 Sep 21;8(9):e17963 [FREE Full text] [doi: 10.2196/17963] [Medline: 32955442]
73. Noguchi-Shinohara M, Domoto C, Yoshida T, Niwa K, Yuki-Nozaki S, Samuraki-Yokohama M, et al. A new computerized assessment battery for cognition (C-ABC) to detect mild cognitive impairment and dementia around 5 min. PLoS One 2020 Dec 11;15(12):e0243469 [FREE Full text] [doi: 10.1371/journal.pone.0243469] [Medline: 33306697]
74. Chan JY, Wong A, Yiu B, Mok H, Lam P, Kwan P, et al. Electronic cognitive screen technology for screening older adults with dementia and mild cognitive impairment in a community setting: development and validation study. J Med Internet Res 2020 Dec 18;22(12):e17332 [FREE Full text] [doi: 10.2196/17332] [Medline: 33337341]
75. Rodríguez-Salgado AM, Llibre-Guerra JJ, Tsoy E, Peñalver-Guia AI, Bringas G, Erlhoff SJ, et al. A brief digital cognitive assessment for detection of cognitive impairment in Cuban older adults. J Alzheimers Dis 2021;79(1):85-94 [FREE Full text] [doi: 10.3233/JAD-200985] [Medline: 33216033]
76. Bottiroli S, Bernini S, Cavallini E, Sinforiani E, Zucchella C, Pazzi S, et al. The smart aging platform for assessing early phases of cognitive impairment in patients with neurodegenerative diseases. Front Psychol 2021 Mar 15;12:635410 [FREE Full text] [doi: 10.3389/fpsyg.2021.635410] [Medline: 33790839]
77. Pilatti A, Godoy J, Brussino S. Adaptación de instrumentos entre culturas: ejemplos de procedimientos seguidos para medir las expectativas hacia el alcohol en el ámbito argentino. Trastor Adict 2012 Apr;14(2):58-64 [FREE Full text] [doi: 10.1016/s1575-0973(12)70045-4]
78. Bartram D, Hambleton R. ITC guidelines for translating and adapting tests (Second Edition). Int J Test 2028;18(2):101-134 [FREE Full text]
79. Hernández A, Hidalgo M, Hambleton RK, Gómez-Benito J. International Test Commission guidelines for test adaptation: a criterion checklist. Psicothema 2020 Aug;32(3):390-398 [FREE Full text] [doi: 10.7334/psicothema2019.306] [Medline: 32711675]
80. Nieto A, Hernández-Rodríguez E, Hernández-Torres A, Velasco Rodríguez-Solís P, Hess-Medler S, Machado-Fernández A, et al. Versión paralela del test de aprendizaje verbal España-Complutense (TAVEC). Rev Neurol 2014 May 01;58(9):389-395 [FREE Full text] [Medline: 24777766]
81. Chen K, Chan AH. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014 Mar 24;57(5):635-652 [FREE Full text] [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
82. Chen K, Lou V. Measuring senior technology acceptance: development of a brief, 14-item scale. Innov Aging 2020 May;4(3):igaa016 [FREE Full text] [doi: 10.1093/geroni/igaa016] [Medline: 32617418]
83. Chan JY, Yau ST, Kwok TC, Tsoi KK. Diagnostic performance of digital cognitive tests for the identification of MCI and dementia: a systematic review. Ageing Res Rev 2021 Dec;72:101506. [doi: 10.1016/j.arr.2021.101506] [Medline: 34744026]

Abbreviations

AI: artificial intelligence
BHA: Brain Health Assessment
CVLT: California Verbal Learning Test
eSAGE: Self-Administered Gerocognitive Examination
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROC: receiver operating characteristic
STM: short-term memory
TMT: Trail Making Test
VR: virtual reality
VSM: Virtual Supermarket

Edited by J Torous; submitted 21.03.23; peer-reviewed by E Siegmund-Schultze, A Kuerbis; comments to author 11.07.23; revised version received 03.09.23; accepted 24.10.23; published 08.12.23.

Please cite as:
Cubillos C, Rienzo A
Digital Cognitive Assessment Tests for Older Adults: Systematic Literature Review
JMIR Ment Health 2023;10:e47487
URL: https://mental.jmir.org/2023/1/e47487
doi:10.2196/47487
PMID:38064247

Review

Assessing the Impact of Evidence-Based Mental Health Guidance During the COVID-19 Pandemic: Systematic Review and Qualitative Evaluation

Katharine A Smith[1,2,3], MD, PhD; Edoardo G Ostinelli[1,3], MD; Roger Ede[2]; Lisa Allard[4], CertClinPharm, GradDipClinPharm; Michaela Thomson[5], MA; Kiran Hewitt[6], MSc; Petra Brown[7,8]; Caroline Zangani[1,2,3], MD; Matthew Jenkins[9], PhD; Verena Hinze[1,3], PhD; George Ma[10], GradDipClinPharm; Prajnesh Pothulu[11], GCertPharmPrac; Catherine Henshall[2,12,13], MA, PhD; Gin S Malhi[1,14,15], MSt, MD; Susanna Every-Palmer[9], MD, PhD; Andrea Cipriani[1,2,3], MD, PhD

[1]Department of Psychiatry, University of Oxford, Oxford, United Kingdom
[2]Oxford Health NHS Foundation Trust, Oxford, United Kingdom
[3]Oxford Precision Psychiatry Lab, NIHR Oxford Health Biomedical Research Centre, Oxford, United Kingdom
[4]Pharmacy Department, Royal Brisbane and Women's Hospital, Brisbane, Australia
[5]Mersey Care NHS Foundation Trust, Liverpool, United Kingdom
[6]Lincolnshire Partnership NHS Foundation Trust, Lincoln, United Kingdom
[7]Pennine Care NHS Foundation Trust, Manchester, United Kingdom
[8]Department of Pharmacy and Optometry, University of Manchester, Manchester, United Kingdom
[9]Department of Psychological Medicine, University of Otago, Wellington, New Zealand
[10]Pharmacy Department, The Prince Charles Hospital, Metro North Health, Brisbane, Australia
[11]Pharmacy Department, Princess Alexandra Hospital, Brisbane, Australia
[12]Oxford Institute of Applied Health Research, Oxford Brookes University, Oxford, United Kingdom
[13]Nursing and Midwifery Office, National Institute for Health and Care Research, London, United Kingdom
[14]Academic Department of Psychiatry, Faculty of Medicine and Health, Kolling Institute, Northern Clinical School, The University of Sydney, Sydney, Australia
[15]CADE Clinic and Mood-T, Royal North Shore Hospital, Northern Sydney Local Health District, St. Leonards, New South Wales, Australia

Corresponding Author:
Katharine A Smith, MD, PhD
Department of Psychiatry
University of Oxford
Warneford Hospital
Warneford Lane, Headington
Oxford, OX3 7JX
United Kingdom
Phone: 44 1865902135
Email: katharine.smith@psych.ox.ac.uk

Abstract

Background: During the COVID-19 pandemic, the Oxford Precision Psychiatry Lab (OxPPL) developed open-access web-based summaries of mental health care guidelines (OxPPL guidance) in key areas such as digital approaches and telepsychiatry, suicide and self-harm, domestic violence and abuse, perinatal care, and vaccine hesitancy and prioritization in the context of mental illness, to inform timely clinical decision-making.

Objective: This study aimed to evaluate the practice of creating evidence-based health guidelines during health emergencies using the OxPPL guidance as an example. An international network of clinical sites and colleagues (in Australia, New Zealand, and the United Kingdom) including clinicians, researchers, and experts by experience aimed to (1) evaluate the clinical impact of the OxPPL guidance, as an example of an evidence-based summary of guidelines; (2) review the literature for other evidence-based summaries of COVID-19 guidelines regarding mental health care; and (3) produce a framework for response to future global health emergencies.

Methods: The impact and clinical utility of the OxPPL guidance were assessed using clinicians' feedback via an international survey and focus groups. A systematic review (protocol registered on Open Science Framework) identified summaries or syntheses of guidelines for mental health care during and after the COVID-19 pandemic and assessed the accuracy of the methods used in the OxPPL guidance by identifying any resources that the guidance had not included.

Results: Overall, 80.2% (146/182) of the clinicians agreed or strongly agreed that the OxPPL guidance answered important clinical questions, 73.1% (133/182) stated that the guidance was relevant to their service, 59.3% (108/182) said that the guidelines had or would have a positive impact on their clinical practice, 42.9% (78/182) that they had shared or would share the guidance, and 80.2% (146/182) stated that the methodology could be used during future health crises. The focus groups found that the combination of evidence-based knowledge, clinical viewpoint, and visibility was crucial for clinical implementation. The systematic review identified 2543 records, of which 2 syntheses of guidelines met all the inclusion criteria, but only 1 (the OxPPL guidance) used evidence-based methodology. The review showed that the OxPPL guidance had included the majority of eligible guidelines, but 6 were identified that had not been included.

Conclusions: The study identified an unmet need for web-based, evidence-based mental health care guidance during the COVID-19 pandemic. The OxPPL guidance was evaluated by clinicians as having a real-world clinical impact. Robust evidence-based methodology and expertise in mental health are necessary, but easy accessibility is also needed, and digital technology can materially help. Further health emergencies are inevitable and now is the ideal time to prepare, including addressing the training needs of clinicians, patients, and carers, especially in areas such as telepsychiatry and digital mental health. For future planning, guidance should be widely disseminated on an international platform, with allocated resources to support adaptive updates.

(JMIR Ment Health 2023;10:e52901) doi:10.2196/52901

KEYWORDS

evidence synthesis; guidelines; mental health; systematic review; focus group; survey; COVID-19; pandemic; digital health; eHealth; mobile phone

Introduction

Background

Globally, the needs of those with mental illness were particularly acute during the COVID-19 pandemic, especially during lockdown periods. COVID-19 highlighted preexisting disparities in health care and the increased risks for those with mental disorders [1,2]. Guidelines specifically for mental health disorders were often limited or difficult to find, particularly in the early phases of the pandemic [3]. As it was primarily a respiratory infection, early COVID-19 guidance often focused on the physical management of patients, whereas guidance for those with mental disorders was largely absent or hidden within the wider recommendations [3]. However, the needs of people with serious mental illness and their difficulties in complying with isolation and distancing regulations (often because of the structural design and layout of psychiatric facilities), in addition to the associated physical, social, and economic disadvantages associated with long-term serious mental health conditions, made this group especially vulnerable to the adverse effects of the pandemic [4].

At the same time, the COVID-19 pandemic also created an opportunity to examine the practice of developing evidence-based health guidelines during a health emergency [5,6]. Substantial progress was made in the rapid generation of new evidence, particularly in physical health care (eg, the RECOVERY and COVID-19 vaccine trials [7]), but there were also examples of multiple research studies and systematic reviews that overlapped or were too small to produce significant findings [8]. Ideally, guidelines, including those for mental disorders, should be an easily accessible resource for clinicians and patients to enable them to improve clinical outcomes [9]. Before the pandemic, there were already established methods for developing guidelines (eg, the GRADE Evidence to Decision frameworks [10]), but in practice, these were often not followed [11]. There are also well-defined methods for rapid review and guideline development, applicable in the context of health emergencies [12-14]. However, the large number of clinical guidelines and consensus statements produced in the early phase of the COVID-19 pandemic were often country or service specific, were focused on physical care, were of variable methodology and quality, or were not regularly updated, often leaving clinicians confused about where to turn [12].

During the COVID-19 pandemic, we (the Oxford Precision Psychiatry Lab [OxPPL]; [15]) developed evidence-based summaries of guidelines regarding mental disorders using a pragmatic, evidence-based approach ("OxPPL guidance") [3]. This study aimed to complete a formal assessment of the real-world clinical impact of these syntheses of guidelines. We recruited an international network of clinical sites and colleagues (in Australia, New Zealand, and the United Kingdom) to provide feedback from countries and regions that had experienced different case rates and lockdown restrictions throughout the COVID-19 pandemic.

Study Objectives

Our study objectives were as follows:

1. To assess the impact and clinical utility of the OxPPL guidance, by collecting clinicians' feedback via an international survey and focus groups
2. To conduct a systematic review of the literature to identify and compare the OxPPL guidance with other syntheses of mental health care guidelines in the context of COVID-19

and to assess the accuracy of the OxPPL guidance methodology
3. To produce a framework for evidence syntheses to support guideline development in mental health care for future global health emergencies

Methods

This study aimed to assess the real-world impact of the resources developed by OxPPL [15]. The methods used to develop the OxPPL guidance are outlined in detail elsewhere [3] and are summarized briefly for context in this paper.

Development of Guidance Resources

Starting in March 2020, a multidisciplinary team at OxPPL [15] developed evidence-based summaries of guidelines for managing mental health disorders in the context of COVID-19.

The OxPPL guidance team consisted of mental health researchers (physicians, methodologists, nurses, and pharmacists) who systematically searched English-language websites for guidelines about a range of topics in managing mental illness in the context of the COVID-19 pandemic and synthesized these into summaries of guidance using a validated, evidence-based approach [3]. A "bottom up" as well as "top-down" approach [16] was used: the choice of topics was driven by clinical needs, in consultation with clinician colleagues within local UK National Health Service (NHS) mental health services and with international collaborators. The topics included a range of approaches to the care of patients with mental illness during the COVID-19 pandemic including medications, psychological treatment, organization of services, and modes of delivery. Initially we focused on immediate priority areas. The rapid transition to telepsychiatry services [17,18] prompted the need for guidance syntheses regarding digital approaches and telepsychiatry [19], followed by other areas of mental health including inpatient care; use of clozapine, lithium, and antipsychotics; suicide and self-harm; domestic violence and abuse; substance use disorders [20]; perinatal care [21]; and vaccine prioritization [22] and hesitancy [23] in the context of mental illness. The team updated the guidance regularly and collaborated with experts in each area to keep the guidance focused, comprehensive, and globally representative. The OxPPL guidance [24] was free to access and advertised via NHS websites, social media, and academic and clinical contacts. The guidance team also collaborated with other sites to adapt and translate the guidance for use in 6 non–English-speaking countries (China, Italy, Bulgaria, France, Japan, and Turkey).

Recruitment of Sites to Collaborate in Assessing the Impact of the OxPPL Guidance

We recruited a multidisciplinary international network of clinical sites and colleagues (including clinicians, researchers, and experts by experience), with 4 sites across the United Kingdom, 2 sites in Australia, and 1 site in New Zealand. Collaborators from the sites participated in the systematic review and the development of the survey, including providing adapted versions for clinicians in Australia and New Zealand. They also identified and facilitated the routes for dissemination of the survey to patient-facing clinicians in their mental health care services and identified potential participants for the focus groups.

Development of the Survey

A multidisciplinary group involving clinicians from medicine, nursing, psychology, and pharmacy and a Patient and Public Involvement representative developed the survey using an iterative process. The focus of the survey was to collect multisite, multidisciplinary, international feedback about the OxPPL guidance in its current format, its usefulness during the acute phase and immediate aftermath of the pandemic, and any potential uses and adaptations for future use.

Different versions of the survey were directed to respondents if they had or had not seen the OxPPL guidance before completing the survey. Those who had not seen the guidance previously were invited to do so before answering the survey questions. The main survey was created by the multidisciplinary group (based in the United Kingdom), and collaborators from the Australia and New Zealand sites created the adapted versions, which were consistent with local practices. Key changes for Australia and New Zealand were to use country-specific ethnicity categories and descriptions of mental health service backgrounds.

The survey was hosted on Microsoft Forms, and a link was sent with a covering email to all the clinicians in mental health care within the participating sites inviting a response, and follow-up reminder emails were sent after this. This process was coordinated by the lead at each site using existing email databases, which included (but were not necessarily limited to) staff members within relevant mental health services. The survey and email specified that completion was restricted to patient-facing mental health care staff. In addition, any respondents who answered *no* to the initial screening question within the survey about this were redirected and thanked for their time, and no further survey completion was allowed. The survey was open for 3 months (from September 13, 2022, to December 25, 2022).

The survey (copy available on request from the authors) was anonymous but collected demographic data (eg, age, gender, ethnicity, place of work, and professional background) for descriptive purposes. All participants were asked which topics they had looked at or used, whether they thought that the guidance answered important clinical questions, whether the methods used were appropriate, whether the layout was easy to access, and whether the extra features (eg, downloadable summaries) were useful. Respondents were also asked about the real-world impact of the OxPPL guidance: whether it had made an impact on their clinical practice; whether it was relevant and applicable in their work setting and patient population; and whether they had shared the OxPPL guidance with other clinicians, patients, or carers. Answers were scored on a 5-point Likert scale ranging from strongly disagree to strongly agree.

Ethical Considerations

The participating sites obtained ethics and locality approval for the study, as required under local governance. The 4 UK sites gained approvals as service evaluations from their local UK NHS Trusts. Ethics approval in Australia was obtained from

the Gold Coast Hospital and Health Service (ethics EX/2022/QGC/87527) and in New Zealand from the University of Otago Ethics Committee (reference 22/103), with Whatu Ora locality approval and Māori Consultation with the Ngāi Tahu Research Committee. Participants in the focus groups received recompense for their time (£15, approximately US $19, or an equivalent voucher) after completion of the group.

Focus Groups

The aim of the focus groups was to provide more detailed information about clinicians' views regarding the OxPPL guidance. Further details are available in Multimedia Appendix 1, but the broad topic prompts included a short opening section with prompts to discuss participants' work settings, experiences during the COVID-19 pandemic, and general routes of information seeking for the management of mental illness. Most of the group's time focused on the use of the OxPPL guidance. The facilitator followed the topic guide and there were broad prompts encouraging feedback about the OxPPL guidance resources, including clinical relevance, ease of use, reliability, and areas for improvement. Finally, there was a discussion about its applicability to future health crises or pandemics.

In total, 2 focus groups were conducted at different times to allow for international participation across different time zones and work patterns. Participants from a range of professional backgrounds within the participating mental health services were invited, and recruitment was through convenience sampling. Participants received a participant information sheet and completed written informed consent before the group, with the opportunity to ask further questions if needed. The focus groups lasted for an hour and were facilitated by 2 researchers (CH and KAS) using a semistructured topic guide (Multimedia Appendix 1). The groups were conducted remotely, recorded, and transcribed. The data were analyzed thematically and managed using the Framework method [25], with double coding of transcripts by 2 researchers (KAS and CH) to ensure consistency. Once the researchers had coded and categorized the data within the Framework matrix, the team discussed any emerging findings, to aid interpretation and explore and develop themes relating to participants' views and experiences. The COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [26] were used to report the qualitative results.

Systematic Review

The primary aim of the systematic review was to assess whether there were other reported syntheses of guidance in the management of mental illness during the COVID-19 pandemic. A secondary aim was to assess the accuracy of the methodology used in the OxPPL guidance to include all the available individual resources—this was completed by also extracting all the papers about individual resources, which met the criteria for inclusion in the OxPPL guidance, and comparing this with the group of resources actually included.

The protocol for the systematic review with full details was published on Open Science Framework [27], and we have reported the results following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [28]. The completed PRISMA checklist is available in Multimedia Appendix 2. The search strategy included broad terms relating to mental health, the COVID-19 pandemic, and guidelines or guidance, to capture all the available records (for full details, refer to the protocol in Open Science Framework [27]) from database inception until the search date (March 22, 2023).

At least 2 members of the review team (KAS, SE-P, and GSM) independently screened the title and abstract of the retrieved records. Full texts of the potentially eligible records were then assessed against the eligibility criteria by 2 researchers (KAS and EGO). Any disagreement was discussed with another member of the research team (AC). Overall, three main groups of papers were considered to be eligible: those describing (1) syntheses or (2) collections of guidelines regarding mental illness and COVID-19, and (3) papers reporting individual guidelines that met the criteria for inclusion in the OxPPL guidance were also retrieved, and their content was compared against the OxPPL guidance. We included only reports that were relevant to a diagnosis of mental illness (therefore, we did not include reports solely related to the prevention of mental health symptoms or the well-being of health care staff or the general public). Guidelines related to mental disorders following the acute pandemic (eg, post–COVID-19 condition [long COVID], the neuropsychiatric consequences of COVID-19) were not included. Data extraction from the included reports was performed and double checked by 2 researchers (KAS and EGO), including the countries covered by the guidelines, date, methodology, type (synthesis, collection, or individual guideline), and key findings.

Results

Characteristics of the Sites Participating in the Survey and Focus Groups

Characteristics and locations of the 6 participating sites are shown in Table 1 and Figure 1. There were 4 sites in the United Kingdom (in England, across different areas of the country); 2 sites in Australia (Sydney [focus groups only] and Brisbane [survey only]); and 1 in Wellington, New Zealand.

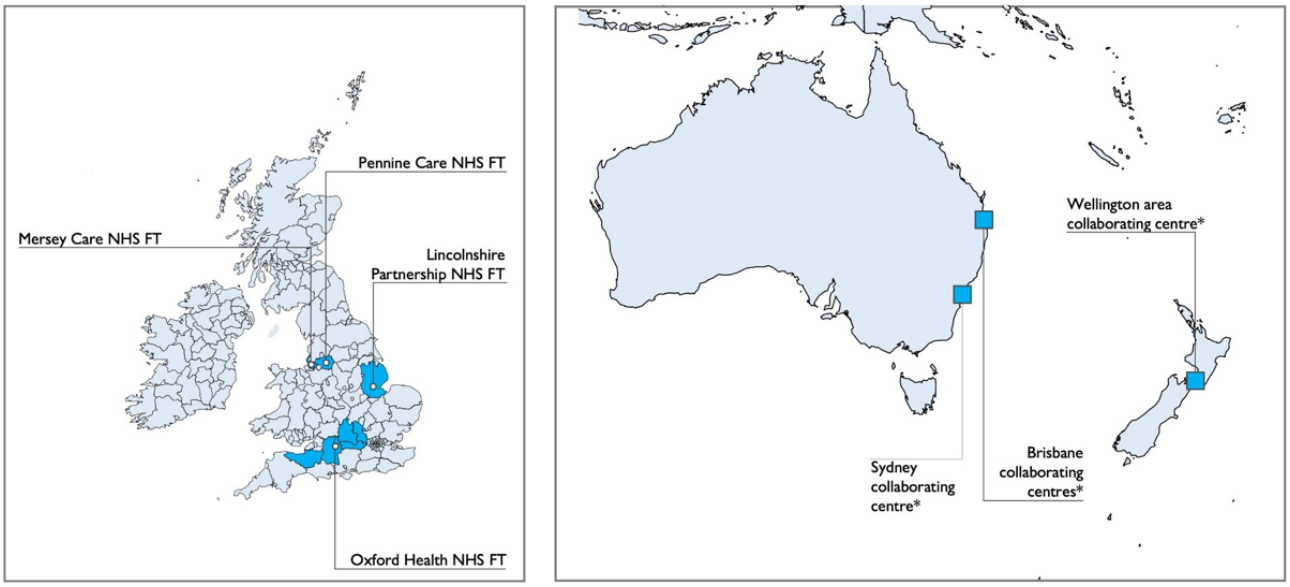
Table 1. Characteristics of the participating sites.

Site[a]	Staff who were sent the survey link (N=17,473), n (%)	Further information
Mersey Care NHS[b] Foundation Trust, United Kingdom	9028 (51.67) patient-facing mental health staff	• This NHS Trust provides physical health and mental health services in the North West of England, serving >1.4 million people. It is also commissioned for services that cover the North West, North Wales, and the Midlands. Core centers are in Liverpool, Sefton, Knowsley, St Helens, Halton, and Warrington. • It provides specialist inpatient and community services to support physical and mental health and specialist inpatient mental health, learning disability, addiction, and brain injury services. • It is 1 of 3 NHS Trusts in the United Kingdom that offer high-security mental health facilities.
Pennine Care NHS Foundation Trust, United Kingdom	3083 (17.64) mental health and learning disabilities staff	• This is an NHS Trust in the North of England, serving a population of 1.3 million in 6 boroughs: Bury, Glossop, Oldham, Rochdale, Stockport, and Tameside.
Lincolnshire Partnership NHS Foundation Trust, United Kingdom	1770 (10.13) patient-facing mental health staff	• This NHS Trust in the East of England provides services to a population of 766,000 in Lincolnshire and 160,000 in North East Lincolnshire. Main sites are Lincoln, Grantham, and Boston. • It includes community mental health teams and several other specialist, crisis, and home treatment services and inpatient beds.
Oxford Health NHS Foundation Trust, United Kingdom	2113 (12.09) mental health staff in the Oxfordshire and Buckinghamshire Mental Health directorates	• This NHS Trust provides community health, mental health, and specialized health services to approximately 2 million people in an area in the Southeast of England across the counties of Oxfordshire, Buckinghamshire, Berkshire, Wiltshire, Swindon, Bath, and North East Somerset. • It also provides a range of specialized health services including forensic mental health and eating disorder services across a wide geographic area including support for patients in Berkshire and Wales.
Brisbane, Australia	830 (4.75) patient-facing mental health staff	• Staff approached were from mental health services within Metro South Health and Metro North Health, Brisbane, Queensland, including the Royal Brisbane and Women's Hospital, Prince Charles Hospital, and Princess Alexandra Hospital. • Staff worked within a variety of services including acute adult inpatients, adult community mental health (outpatients), older persons mental health, mobile intensive and long-term rehabilitation, early psychosis mental health, adolescent mental health, homeless outreach team, psychiatric emergency care, and alcohol and drug services.
Wellington, New Zealand	649 (3.71) patient-facing mental health clinicians	• Staff were surveyed from the Mental Health, Addiction, and Intellectual Disability Service, which serves the lower North Island of New Zealand and includes local, regional, and national services. • Services are provided from multiple sites within greater Wellington, Hutt Valley, and Wairarapa. • The service covers a range of specialties in mental health including adult, community, crisis, consult liaison, child and adolescent, older adult, forensic, eating disorders, and substance misuse.

[a]Sydney, Australia, site participated in the focus group but not in the survey.

[b]NHS: National Health Service.

Figure 1. Participating sites and their locations. NHS FT: National Health Service Foundation Trust; *Brisbane collaborating centers: Royal Brisbane and Women's Hospital; Prince Charles Hospital; and Princess Alexandra Hospital, Brisbane, Queensland (survey only); *Sydney collaborating center: Royal North Shore Hospital, Sydney, New South Wales (focus group only); *Wellington area collaborating center: Mental Health, Addiction and Intellectual Disability Service, Wellington and surrounding areas, North Island. This figure was created using MAPSVG [29] which is licensed under Creative Commons Attribution 4.0 International License [30].



Survey

Response rates to the survey were low. From a total of 17,473 staff who were emailed across all the sites, only 184 (1.05%) survey responses were received. In country-specific comparisons, response rates were higher in New Zealand (48/649, 7.4%), but the absolute numbers were small.

Characteristics of the Survey Participants

Characteristics of the 184 survey participants are shown in Table 2 and Multimedia Appendix 3. The clinical staff who completed the survey had a range of professional backgrounds including nursing (74/184, 40.2%), allied health professions (58/184, 31.5%), and medical (37/184, 20.1%). Within mental health, different services were represented, with 52.2% (96/184) working in adult mental health. Of the 184 participants, 120 (65.2%) were women and 164 (89.1%) were aged between 25 and 64 years. Most participants in the United Kingdom identified their ethnic group as White (British, Irish, or other; 97/120, 80.8%), most of those in New Zealand identified as New Zealand European (25/48, 52%) or White (2/48, 4%), and most of those in Australia identified as Oceanian (7/16, 44%) or White (2/16, 13%). Other ethnic backgrounds were represented but in much smaller numbers. For example, in the United Kingdom, 3.3% (4/120) identified as Black or Black British, and in New Zealand, 15% (7/48) identified as Māori or Māori-New Zealand European.

Table 2. Characteristics of the survey participants.

Characteristics	United Kingdom (n=120), n (%)	New Zealand (n=48), n (%)	Australia (n=16), n (%)	Total (n=184), n (%)
Professional background				
Medical	28 (23.3)	8 (17)	1 (6)	37 (20.1)
Nursing	46 (38.3)	24 (50)	4 (25)	74 (40.2)
Occupational therapy	1 (0.8)	3 (6)	1 (6)	5 (2.7)
Pharmacy	6 (5)	0 (0)	3 (19)	9 (4.9)
Physiotherapy	2 (1.7)	0 (0)	0 (0)	2 (1.1)
Psychology	14 (11.7)	9 (19)	2 (13)	25 (13.6)
Social work	10 (8.3)	2 (4)	5 (31)	17 (9.2)
Other	13 (10.8)	2 (4)	0 (0)	15 (8.2)
Main service (>50% of the time)				
Adult mental health	65 (54.2)	18 (38)	13 (81)	96 (52.2)
Child and adolescent mental health	9 (7.5)	12 (25)	1 (6)	22 (12.0)
Forensic mental health	12 (10)	13 (27)	1 (6)	26 (14.1)
Learning disability	6 (5)	0 (0)	0 (0)	6 (3.3)
Memory assessment clinic	3 (2.5)	0 (0)	0 (0)	3 (1.6)
Older adult mental health	7 (5.8)	2 (4)	1 (6)	10 (5.4)
Not related to mental health	2 (1.7)	0 (0)	0 (0)	2 (1.1)
Other	16 (13.3)	3 (6)	0 (0)	19 (10.3)
Age group (years)				
18-24	9 (7.5)	0 (0)	0 (0)	9 (4.9)
25-34	29 (24.2)	14 (29)	3 (19)	46 (25)
35-44	24 (20)	11 (23)	6 (38)	41 (22.3)
45-54	34 (28.3)	10 (21)	3 (19)	47 (25.5)
55-64	16 (13.3)	10 (21)	4 (25)	30 (16.3)
65-74	5 (4.2)	0 (0)	0 (0)	5 (2.7)
≥75	0 (0)	2 (4)	0 (0)	2 (1.1)
Prefer not to say or N/A[a]	3 (2.5)	1 (2)	0 (0)	2 (1.1)
Gender				
Female	78 (65)	31 (65)	11 (69)	120 (65.2)
Male	39 (32.5)	16 (33)	5 (31)	60 (32.6)
Nonbinary	1 (0.8)	0 (0)	0 (0)	1 (0.5)
Prefer not to say or N/A	2 (1.7)	1 (2)	0 (0)	3 (1.6)

[a]N/A: not applicable.

Survey Responses

Survey responses are described in Multimedia Appendix 4. Of the 184 who started the survey, 2 (1.3%) responded that they did not work within mental health, and therefore could not complete the subsequent mental health guidance questions. Overall, 56.6% (103/182) reported not having seen the OxPPL guidance before the survey, with higher rates in both Australia and New Zealand than in the United Kingdom. Of those who were already aware of the resources (ie, had seen them), the more frequently accessed topics were suicide and self-harm (accessed by 38/79, 48%) and vaccine uptake and hesitancy (accessed by 34/79, 43%). Overall, 35% (28/79) had looked at telepsychiatry and digital approaches. Of those who looked at this guidance for the first time within the study, more frequently accessed topics were also suicide and self-harm (38/103, 36.9%) and clozapine treatment (35/103, 34.0%). Furthermore, 23.3% (24/103) had looked at telepsychiatry and digital approaches.

Of the survey respondents, 80.2% (146/182) agreed or strongly agreed that the OxPPL guidance answered important clinical

questions, 76.4% (139/182) stated that the methods used were appropriate (ie, sufficiently trustworthy to be used in clinical practice), 78% (142/182) said that the web-based layout was easy to access, and 79.7% (145/182) stated that the extra web-based features (eg, downloadable summaries) were useful. Overall, 73.1% (133/182) reported that the guidance was relevant and applicable for their service, and 72% (131/182) that it was relevant and applicable for their patient population. Overall, 59.3% (108/182) reported that the guidelines had or would have a positive impact on their clinical practice, and 42.9% (78/182) reported that they had shared or would share the guidance with others. For those who had already shared the guidance, 90% (70/78) had shared with coworkers, 33% (26/78) with other professionals, 21% (16/78) with patients, and 12% (9/78) with carers.

Furthermore, 73.1% (133/182) of the survey respondents reported having also used other resources for information including specialty-based, profession-based, and governmental websites. However, 26.9% (49/182) reported having used no other resource for guidance in mental health diagnosis and treatment during the COVID-19 pandemic. Overall, 80.2% (146/182) agreed or strongly agreed that the methodology used by the OxPPL guidance could be used in future pandemics and health crises, and 64.3% (117/182) stated that other topics could then be added to meet these future needs.

Focus Groups

The 2 focus groups included a total of 18 participants (n=8, 44% women), including 9 (50%) physicians, 2 (11%) pharmacists, 3 (17%) nurses, 1 (6%) occupational therapist, 2 (11%) health care assistants, and 1 (6%) social worker. The participants worked across different areas of mental health including adult, child and adolescent, specialist bipolar disorder, forensic, and eating disorders (Multimedia Appendix 5).

The main themes, with illustrative quotes arising from the framework analysis of the content of the focus groups are summarized in Table 3, and the full Framework analysis is presented in Multimedia Appendix 6.

In total, 4 main themes emerged from the analysis. These were (1) challenges and uncertainty during the pandemic, (2) need for trustworthy information, (3) feedback about the OxPPL guidance, and (4) use of the OxPPL guidance in the future. In theme 1, many participants commented about the challenges and uncertainties faced by mental health clinicians and patients during the pandemic. Participants identified the constantly changing nature of the challenges, high workload, staff shortages, changing staff roles, and the pressure and urgency to implement changes. The rapid transition to telepsychiatry was highlighted as a key change which required adaptation. It was also noted that mental health often seemed to be forgotten, especially in the early stages, but that mental health care had particular challenges in the context of the pandemic, including in inpatient care. Participants from New Zealand and Australia reflected about their different experiences compared with those in the United Kingdom and other countries. Theme 2 explored the need for formal guidance in mental health care: participating clinicians commented that they felt their own uncertainty had affected patients. They also noted that patients often felt more uncertain because of their mental health issues, which led to seeking information from unreliable sources. Participants reported uncertainty about decision-making and noted that there had been many complex decisions in mental health settings. They also described their view that there was an unmet need for reliable and trustworthy guidance to support their decision-making. Theme 3 explored a discussion about the OxPPL guidance: most participants had not been aware of this before the group but reported that they wished they had seen it sooner. Areas such as layout, the web-based open-access format, range of topics, ease of sharing with colleagues, and suitability for different professionals received positive feedback. The combination of evidence-based methods and clinical relevance was felt to be essential. Theme 4 focused on future uses: increased visibility was a key area, new topics were suggested, and possible modifications for patient use were discussed. The participants felt that COVID-19 would continue to be an issue and that the OxPPL guidance would continue to be relevant. The resources were felt to be relevant and easily adaptable for future pandemics or health crises.

Table 3. Quotes relating to themes arising from the focus groups with clinicians.

Themes	Example quotes
Challenges and uncertainty during the COVID-19 pandemic and beyond	• "It's absolutely decimated us from the start; we've had outbreak after outbreak with our patients...it's been horrific and continues to be as bad as the onset." [Social worker; male; UK[a] 1] • "In the early days, we did have to really fight the corner for mental health within...the different services and settings that we have. That, kind of, continued I guess throughout the whole vaccine rollout as well." [Pharmacist; female; UK 3] • "I have seen a lot of people [patients] staying stagnant with us through the COVID period and without the opportunity to go through the rehabilitation phases...I think it had a big impact because people felt like they were stuck there for a very long time." [Consultant psychiatrist; male; UK 1] • "The anatomy of our pandemic...was very different...we initially followed an elimination strategy which was quite successful with the first variants up until Omicron...[we] benefitted from the experience you'd had before." [Consultant psychiatrist; female; NZ[b]]
The need for reliable and trustworthy information	• "It was interesting there were a number of places they [patients and the public] were getting misinformation from." [Consultant psychiatrist; male; AUS[c]] • "Dealing with many patients who are sceptical of many things such as worldwide pandemics was very challenging." [Nurse; male; UK 1] • "Our wards and our medical teams but also our patients needed to understand what was a really ever-changing picture." [Pharmacist; female; UK 3] • "To be honest, at the very beginning, I think we went back to almost first principles...At the very, very early stages, that's what it likely came down to, was independent clinical decision making." [Pharmacist; male; UK 2] • "There were definitely areas that we were grappling with, you know, depots, clozapine, benzos, rapid tranq [sic], and also that whole, sort of, vaccine hesitancy and the confusion that came really with different groupings of who went when." [Pharmacist; female; UK 3] • "A lot of the patients we see do have complicated problems with thyroid, renal function complications, combination strategies, etc., and lithium's always a difficult molecule to manage in that environment." [Consultant psychiatrist; male; AUS] • "A lot of this evidence is around supporting clinicians to make decisions, not about making them for them, so that if you're going to take a risk, you're supporting that risk." [Pharmacist; male; UK 2]
Feedback about the current OxPPL[d] guidance	• "So, I think it's really fantastic, and I wish, like, we'd probably had sight of some of this before to help with some issues that we'll have had in all sorts of our cases, that we've had to manage during the whole of the pandemic." [Nurse matron; female; UK 2] • "I think that just the dissemination and reach is where it could be improved. I just think it's only useful if people are actually reading it and using it." [Mental health care assistant; female; UK 4] • "I like the way that it's laid out by clinical questions and it feels that it's had a lot of clinical input, that you've thought carefully about what is going to matter to patients, families and clinicians. So, the questions are really good." [Consultant psychiatrist; female; NZ] • "Many colleagues that I've forwarded it to recently have found it immensely relevant." [Social worker; male; UK 1] • "So I would see it as...evidence-based, this is authoritative." [Consultant psychiatrist; male; UK 4] • "The clinical relevance is the bit that I loved the most. I had questions and I looked and there they were." [Consultant psychiatrist; female; NZ] • "I did notice there were sections that were relevant to specific professions, like there was nursing and...I always appreciate that, to find things that are specific for what we're doing." [Nurse; female; NZ] • "The first thing that really struck me was the immense number of topics...and the thoroughness with which they've been covered...Also, some of the evidence and data that you've drawn on to substantiate various claims, again, it was very nice to see that." [Consultant psychiatrist; male; AUS] • "I remember when I wanted to join the clozapine clinic...I got loads of documents to go through, it was so much. I felt the information we have on the [OxPPL] guidelines, is compact, something that one can go through within a few minutes...before you go in for the clinic. So I think it's much relevant to this present time that we are in." [Mental health care assistant; male; UK 2] • "So, it's useful in times of great uncertainty where people don't know what the right thing to do is and they're worried about what the risks might be to themselves, for example, about clozapine or lithium treatment." [Consultant psychiatrist; female; NZ] • "We're not in [a] pandemic now, but they're still useful." [Psychiatry trainee; female; UK 2]

Themes	Example quotes
Ideas about future use and relevance to future pandemics	• "It reinforces what you've said and they [patients] do remember it better because when they're reading it later on and it says the same stuff you've just read out with them, it makes a lot more impact, or they share it with their partner or someone else..." [Consultant psychiatrist; male; AUS] • "I think personally, over the next two to three years...COVID [will] probably remain a topical issue and a pressure in the winter." [Pharmacist; male; UK 2] • "As a model, I think it's highly relevant to future pandemics or health crises and you could follow that model quite tightly and come out with something excellent much faster than you could the first time." [Consultant psychiatrist; female; NZ] • "So, I think there is a significant role for this kind of guidance going forward...I think if we are susceptible to one pandemic, maybe we might be susceptible to others, as well." [Consultant psychiatrist; male; UK 1] • "So I think that what you'd need to do is have a critical mass of people who could quickly mobilise to generate those questions and check the specific evidence relating to those questions quickly, because there is quite a bit of intensive resourcing that needs to go in, particularly at the beginning, and then keeping something up to date to make it relevant to the front-of-mind questions that people have." [Consultant psychiatrist; female; NZ]

[a]UK: United Kingdom; sites are numbered from 1 to 4.

[b]NZ: New Zealand.

[c]AUS: Australia.

[d]OxPPL: Oxford Precision Psychiatry Lab.

Systematic Review

The PRISMA flowchart is shown in Figure 2 [28]. We identified a total of 3444 records, which reduced to 2543 records after the removal of duplicates. After screening the abstracts and full texts, a total 46 papers, representing 41 individual guidelines or collections and syntheses of guidelines, were included in the review.

The extracted papers are listed in Tables 4 and 5 [3, 19-21, 31-40, 41, 42] and Multimedia Appendix 7 [43-74]. The papers were divided into three main types:

1. Papers in which the authors or group had aimed to collect and then summarize the currently available guidance on mental health care and COVID-19 (syntheses of guidance; Table 4).
2. Papers in which the authors or groups had collected the available guidance on mental health care and COVID-19 without summary or synthesis ("collections" of guidance; Table 5).
3. Individual guidelines that would have been eligible for inclusion but had not been included within the OxPPL guidance (Multimedia Appendix 7).

Of the 2 syntheses specifically focusing on mental health disorders, the systematic review identified only 1 record in addition to the OxPPL guidance, which reported the summary guidance from the UK Royal College of Psychiatrists [42]. An additional 9 papers were identified describing collections of guidance. These were narrative reviews of guidelines either for specific areas within mental health care, such as tele–mental health or electroconvulsive therapy, or collections of individual guidelines within a specific geographical area or country (Tables 4 and 5).

In total, 30 individual guidelines were retrieved (Multimedia Appendix 7), which had not been included in the OxPPL guidance. Most (24/30) could be excluded because our guidance did not cover those specific areas or because they were not available on the web, and therefore, our methodology could not have captured them. However, our systematic search revealed 6 guidelines that were eligible and had not been included. Of these 6 guidelines, 4 related to specialist areas within telepsychiatry, 1 to behavioral emergencies, and 1 to inpatient care of adults with cognitive impairment, each within the context of COVID-19.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. OxPPL: Oxford Precision Psychiatry Lab; *Single reason is given for exclusion as per the protocol, but many fulfilled multiple exclusion criteria.

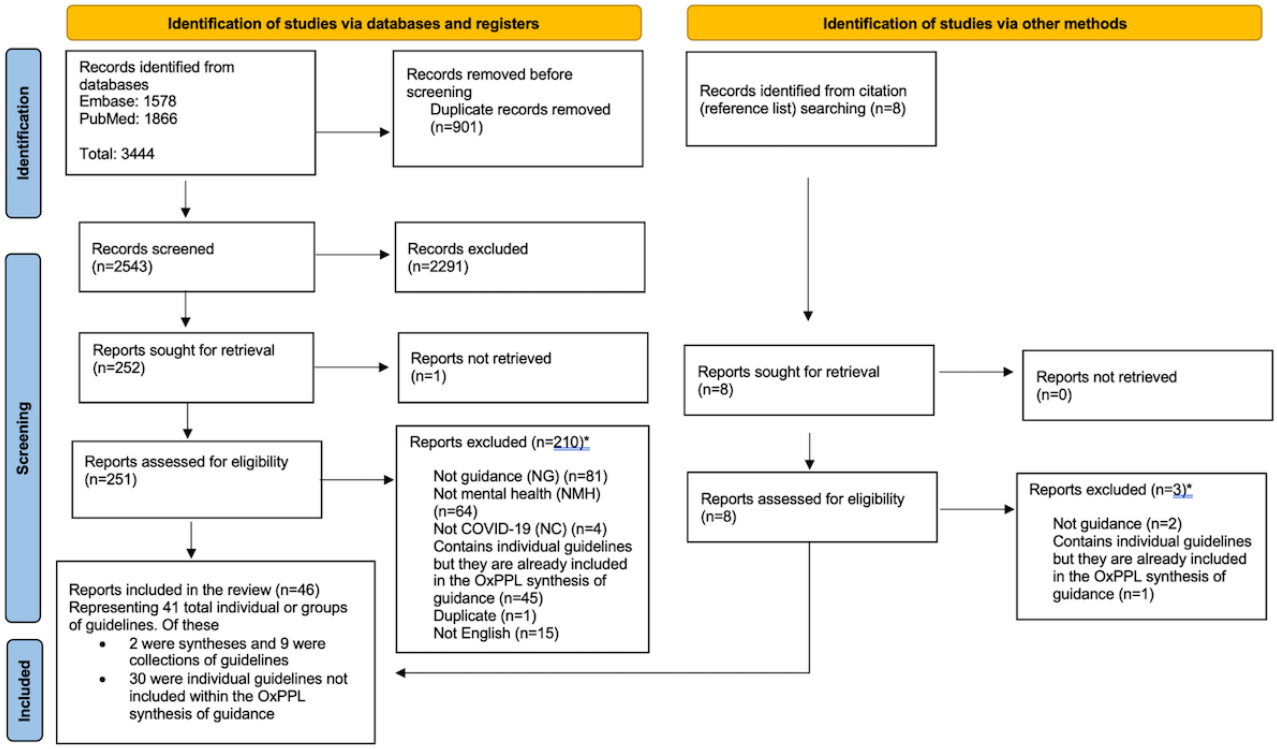


Table 4. Syntheses of guidance about mental health and COVID-19.

Type	Group or organization	Publication	Dates	Methodology	Key aims and web link (if applicable)	Country or language
Synthesis	• Royal College of Psychiatrists [41]	Burn and Mudholkar [33] 2020	2020 to June 2022	• Expert consensus • Developed with input from the COVID-19 college advisory group and approved by the college registrar	• Synthesis of guidance regarding mental health in the context of COVID-19 in the United Kingdom [42]	United Kingdom
Synthesis	• Oxford Precision Psychiatry Lab [15]	Ostinelli et al [20], 2022; Smith et al [19], 2020; Smith et al [3], 2020; and Smith et al [21], 2023	2020 to 2023	• Evidence-based approach • Systematic search, reference checking, and extraction	• Synthesis of available clinical guidelines regarding mental health in the context of the COVID-19 pandemic • Evidence-based approach covering existing guidelines in the English language (covering the United Kingdom, United States, Canada, Australia, New Zealand, and Singapore) and translated for use in 6 other countries (China, Italy, Bulgaria, France, Japan, and Turkey) [24]	United Kingdom, United States, Singapore, Canada, Australia, and New Zealand

Table 5. Collections of guidance about mental health and COVID-19.

Type	Group or organization	Publication	Dates	Methodology	Key aims and web link (if applicable)	Country or language
Collection	__[a]	Sugarman and Busch [40] 2023	Up to February 23, 2022	• Systematic search of reviews/meta-analyses • Narrative review of guidelines	• Systematic review of tele–mental health or telepsychiatry reviews, systematic reviews, and meta-analyses • Narrative review of the guidelines regarding tele–mental health from several relevant major mental health organizations	English-language resources
Collection	—	Bojdani et al [32] 2020	Published in May 2020	• Narrative review combined with informal survey and discussions with colleagues	• Narrative review and PubMed search of the guidelines available in the United States for mental health issues related to COVID-19	United States
Collection	—	Rangaswamy et al [37] 2022	Published in October 2022	• Narrative review combined with expert opinions from informal sessions, talks, and interviews	• Narrative review of Indian mental health care services provided during the COVID-19 pandemic, including the guidelines issued	India
Collection	—	Molebatsi et al [35] 2021	Published in May 2021	• Web-based search for relevant health guidelines, with emails to request guidelines from mental health practitioners in various sub-Saharan African countries, and expert opinions from interviews	• Narrative summary of sub-Saharan African guidelines regarding mental health and psychological support in response to COVID-19	Sub-Saharan African countries
Collection	—	Choi et al [34] 2020	Published in September 2020	• Narrative review of guidelines	• Narrative review of recommendations or guidelines for pregnancy and childbirth during the COVID-19 pandemic, supplemented by the authors' recommendations for clinical practice	English-language resources
Collection	—	Alqahtani et al [31] 2021	2020	• Narrative review of guidelines	• Narrative review of existing guidelines for telepsychology services to inform and adapt to Saudi Arabia and other Arabic communities	Saudi Arabia
Collection	—	Samy et al [38] 2021	2020 and 2021	• Web-based search of PubMed, Web of Science, and Google Scholar	• Narrative review of primary studies and guidelines regarding mental health and COVID-19 in the Asia-Pacific region	Asia-Pacific region

Type	Group or organization	Publication	Dates	Methodology	Key aims and web link (if applicable)	Country or language
Collection	—	Purushothaman et al [36] 2020	2020	• Narrative review of research databases and existing guidelines	• Narrative review of guidelines regarding anesthesia during ECT[b] and its status within aerosol-generating procedures, supplemented by the authors' recommendations about ECT and use of PPEc	English-language resources
Collection	Individual members of the Network of Early Career Professionals working in Addiction Medicine	Scheibein et al [39] 2020	6 months following the COVID-19 pandemic	• Narrative review of guidelines	• Narrative review of country-level guidelines regarding addiction medicine developed in the 6 months following the COVID-19 pandemic	Worldwide

[a]Not reported.

[b]ECT: electroconvulsive therapy.

[c]PPE: personal protective equipment.

Discussion

Principal Findings

In this study, we established a multicenter, international network of sites to assess the real-world clinical implementation of web-based, evidence-based, mental health guidance resources in the context of COVID-19. We reviewed the available evidence about approaches to providing guidance for clinicians in mental health during global health emergencies, using the COVID-19 pandemic as an example. We evaluated clinicians' attitudes toward the need for guidance in general and the OxPPL guidance in particular. We assessed their views about the usefulness of such an approach in future health crises or seasonal demands on health care services and sought to identify the elements needed in preparation for future challenges.

Through the survey and focus groups, clinicians reported that (1) there was a clinical, unmet need for easy-to-access summaries of evidence-based guidance in mental health care during the pandemic, and this need was likely to continue afterward; (2) the web-based, evidence-based summaries of guidance (OxPPL guidance) were clinically useful; (3) they would have or had an impact on their clinical practice; (4) the combination of evidence-based knowledge and a clinical viewpoint was crucial for clinical implementation; and (5) the methodology and web-based format were relevant to future use, including for seasonal surges in illness, future pandemics, and patient adaptation. However, the evaluation also showed that a significant number of the clinicians were not aware of the OxPPL guidance.

Comparison With Previous Studies

Our results are broadly consistent with those of a previous study. Millard et al [75] used surveys and a small number of individual interviews with Australian clinicians to assess attitudes toward the Australian National COVID-19 Clinical Evidence Taskforce living guidelines [76]. Consistent with our findings, they reported that the guidelines were assessed as being relevant to their practice and trustworthy. More than 50% of the respondents had used the guidelines to support their own clinical decision-making (however, they were not explicitly asked whether the guidelines had made a significant impact). They also found that frequent updates and an evidence-based and clinician-led approach were key qualities. However, in contrast to the OxPPL guidance, this focused on 1 country and on the general medical rather than mental health care context.

During a fast-moving pandemic, frequent and responsive updates are important. Guidelines are usually based on evidence from systematic reviews, but these take time and resources to produce, and in the context of an evolving pandemic, they can quickly become out of date. Living systematic reviews can be used to inform clinical guidelines in a more adaptive manner. Examples include the Australian National COVID-19 Clinical Evidence Taskforce living guidelines [76,77]; the World Health Organization's Therapeutics and COVID-19: Living Guidelines [78]; and the Global Alliance for Living Evidence on Anxiety, Depression, and Psychosis (GALENOS) [79,80]. However, these guidelines also require a significant investment of time and resources to maintain, and therefore, a pragmatic, evidence-based approach, such as that used in the OxPPL guidance, may be an alternative.

The need for adaptability and responsiveness to a changing environment during the COVID-19 pandemic were key themes raised by clinicians during the focus groups and will be important aspects of preparing the response to future health emergencies. The COVID-19 pandemic resulted in a rapid transition in clinical practice from in-person consultation to telepsychiatry [17], and clinicians in the focus groups described how they were required to quickly adapt their skills to remote consultations. However, a study of mental health professionals during the COVID-19 pandemic suggests that not all clinicians

were able to adapt easily and that lower self-ratings of digital competence were associated with higher rates of stress [81]. Realizing the full potential of digital interventions to increase the access to and quality of mental health care both in the aftermath of COVID-19 and in planning for other subsequent health emergencies will require clinicians to feel confident and competent in integrating both synchronous approaches (such as telephone or video consultations) and asynchronous techniques (such as the use of apps or smartphones for monitoring and delivering treatments) into the clinical setting [82]. Teaching and training in telepsychiatry and in digital mental health will be essential elements in planning for the next crises [83]. Although there are many potential challenges in implementing training in digital mental health, the first step will be providing easily accessible summaries of evidence-based knowledge, such as the OxPPL guidance, to extend the skills and competencies of clinicians. Training for patients and carers will also be equally important to allow them to access the best combination of evidence and treatments that are available [82].

For successful real-world implementation, the information provided in guidelines also needs to be effectively integrated into clinical decision-making [9]. Guidelines should be freely accessible, with the methodology, potential conflicts of interest, and dates of the updates clearly defined, so that clinicians can assess the reliability of the recommendations. In addition, uptake in the clinic also depends on the willingness of both clinicians and patients to change and the capacity of clinicians to keep up with new recommendations and provide the additional clinician time needed [84]. However, current frameworks for producing guidelines do not yet formally assess patient or clinician burden [84,85] or context [7] in their implementation. Nonadherence to guidelines has important consequences and can contribute to increased adverse outcomes such as hospitalizations, mortality rates, and health care spending [86]. COVID-19 was a recent example where the real-world impact of rapid, evidence-based guidelines could be assessed. This study provides an assessment of the impact and clinical utility of the OxPPL guidance: guidelines will only make an impact on patient care if they are acted upon by clinical teams, and therefore, any approach to guideline development needs to be formally assessed in terms of its impact in real-world clinical care.

Visibility of the resources was a challenge. The results of the survey and focus groups emphasized the need to combine key elements: evidence-based methodology; clinical relevance; and providing the resources on a recognized platform, which would be quickly accessible and adaptable for future health crises. The systematic review identified only 1 other similar resource. Although this was on a well-recognized platform provided by the UK Royal College of Psychiatrists [42], the guidance was not synthesized using evidence-based methods and was predominantly focused on the United Kingdom.

Strengths and Limitations

We recognize that this study has some limitations. Survey response rates were very low, despite active reminders. These are perhaps explained by a level of COVID-19 fatigue or burnout [87,88] at the time of the survey. In addition, each site used their existing email databases for staff. At least some of these databases had significant overlaps with non–mental health staff and non–patient-facing staff (such as administration, or technology support staff). Although the email and survey specified that the survey applied only to patient-facing mental health care staff and that staff could not proceed further if they were not involved in mental health care and were not patient facing, these factors mean that the response rate may be inaccurate and appear to be significantly lower because of an overestimate of the denominator.

Although the survey was launched across multiple sites, all 4 sites in the United Kingdom were in England, and of the Australian sites, Sydney participated only in the focus group and Brisbane participated only in the survey. In addition, our evaluation was undertaken only in English-speaking countries, as adding other languages would have complicated the analysis of the survey and focus groups. However, the survey provided a wide context with a multidisciplinary and international perspective across sites in England, Australia, and New Zealand. Both the survey and focus group feedback was subjective and reflected only the views of mental health care staff about the resources that had been developed for use primarily by clinicians. Several clinicians highlighted how they had also used the guidance usefully in their clinical interactions with patients, and there were many suggestions regarding modifications for patient use. However, further studies would be needed to explore the possible adaptations of the resources using feedback from patients and carers.

The survey and focus group findings highlighted how strongly clinicians felt that reliable guidance in mental health care was needed during health crises. Overall, 80.2% (146/182) of those surveyed felt that the methods and approaches used in the OxPPL guidance would be helpful in addressing the need for mental health care information during future pandemics or health crises. The advantage of our international focus was that it included perspectives from different countries, with focus group participants from Australia and New Zealand highlighting their different pandemic-related experiences. The COVID-19 elimination strategy in New Zealand and stringent lockdowns in Australia meant that case rates were often lower, whereas rates were higher in other countries. During these times, participants reported looking for guidance from countries such as the United Kingdom, the United States, and Canada that had higher rates of COVID-19 but struggled to find reliable, easily accessible sources of guidance in mental health care, particularly during the early part of the pandemic.

We actively updated the resources according to the stage of the pandemic but used a pragmatic, evidence-based approach: rather than a systematic review of the published literature, we used a modified method in which we rapidly and systematically searched web-based sources of guidance [3]. When we performed a systematic review, across all the mental health areas covered in the OxPPL guidance we found only 6 individual guidelines that would have been eligible but were not included. Although this number is low, it may indicate some potential limitations to our approach; the advantages of speed and responsiveness mean that guidelines, particularly those hosted on specialized or region-wide rather than country-wide websites, could be missed. When reviewing the omitted guidelines, we

found that 5 of the 6 guidelines were in more specialized resources that we had not searched, which supports this hypothesis. Given the huge array of available guidelines produced during the COVID-19 pandemic, it is also possible that our systematic review could also have missed some reports.

Conclusions

In summary, we have suggested a successful, evidence-based approach during the pandemic, and our formal assessment of its impact supports the usefulness and relevance of resources such as these in real-world, clinical implementation in mental health care. Now is the time to prepare for the next challenge. COVID-19 is the most recent and striking example of a global health pandemic, but it is by no means the only one. There have been numerous other outbreaks of infectious diseases in the past century, including influenza pandemics or epidemics and severe outbreaks of Ebola and of other coronaviruses [89], and there will be further health crises in the future. The need for timely evidence and rapid and effective communication of recommendations to clinicians to respond to global health emergencies such as COVID-19 is currently at the front of our minds, and this is the ideal time to review the quality of our response and identify learning points [90,91]. Preparing for the next health crisis will also include addressing the training needs of clinicians, patients, and carers, especially in areas such as telepsychiatry and digital mental health [83,92].

The combination of evidence-based knowledge and clinical expertise in mental health care is needed, but this can only be truly successful if hosted on an easily accessible, widely disseminated, web-based platform, and digital technology can materially help [92]. Creating a living document with responsive and frequent updates demands significant resources and a dedicated approach from an international and multidisciplinary team. Establishing all these things now would ensure effective and prompt synthesis of guidance when it is needed in the future [93]. Governments and funding agencies across the world should be aware of this and start preparing immediately to be ready when the next pandemic or international health emergency arises [94].

Acknowledgments

The authors thank the following for their help and support with this project: Orla Macdonald, Simon Kaye, Simon Kelshaw, Joshua Confue, Anna Pearson, Karen Bruce, Alison Walshe, Joe Rafferty, Anna Amiry, Francesca Waddington, Valerie Paulley, Eleanor Paris, Rachel Delahay, Wendy Hurst, Balaji Rajendran, Jessica Powell, Paul Leeks, Tanya Smith, James Hong, Orla McBride, and Rania Elgarf.

KAS, CZ, CH, and AC are supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility. AC is also supported by NIHR Research Professorship (grant RP-2017-08-ST2-006), NIHR Oxford and Thames Valley Applied Research Collaboration, the NIHR Oxford Health Biomedical Research Centre (grant NIHR203316), and Wellcome. VH is supported by the Stiftung Oskar-Helene-Heim. GSM has received grant or research support from National Health and Medical Research Council, Australian Rotary Health, NSW Health, American Foundation for Suicide Prevention, Ramsay Research and Teaching Fund, Elsevier, AstraZeneca, Janssen-Cilag, Lundbeck, Otsuka, and Servier. The views expressed are those of the authors and not necessarily those of the UK National Health Service, NIHR, or UK Department of Health.

Authors' Contributions

KAS and AC devised the project. KAS coordinated all sites, led the development of the survey, cofacilitated the focus groups, led the focus group analysis and the systematic review, and wrote the first draft of the paper. CH cofacilitated the focus groups and supported the focus group analysis. AC, SE-P, GSM, and EGO were coinvestigators in the systematic review. All coauthors were involved in the surveys or focus groups, revised the drafts of the paper, and approved the final version.

Conflicts of Interest

AC has received research and consultancy fees from INCiPiT (Italian Network for Pediatric Trials), CARIPLO Foundation, Lundbeck, and Angelini Pharma, outside this study. KAS, SE-P, and MJ have no conflicts to declare. GSM has been a consultant for AstraZeneca, Janssen-Cilag, Lundbeck, Otsuka, and Servier.

Multimedia Appendix 1
Focus group topic guide.
[DOCX File , 15 KB - mental_v10i1e52901_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 20 KB - mental_v10i1e52901_app2.docx]

Multimedia Appendix 3
Characteristics of respondents to the survey.

[DOCX File , 23 KB - mental_v10i1e52901_app3.docx]

Multimedia Appendix 4
Survey responses.
[DOCX File , 24 KB - mental_v10i1e52901_app4.docx]

Multimedia Appendix 5
Characteristics of focus group participants.
[DOCX File , 15 KB - mental_v10i1e52901_app5.docx]

Multimedia Appendix 6
Framework analysis of the focus group transcripts.
[DOCX File , 26 KB - mental_v10i1e52901_app6.docx]

Multimedia Appendix 7
Individual guidelines about mental health and COVID-19, which were not originally included in the Oxford Precision Psychiatry Lab guidance.
[DOCX File , 25 KB - mental_v10i1e52901_app7.docx]

References

1. Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, et al. How mental health care should change as a consequence of the COVID-19 pandemic. Lancet Psychiatry 2020 Sep;7(9):813-824 [FREE Full text] [doi: 10.1016/S2215-0366(20)30307-2] [Medline: 32682460]
2. Kola L, Kohrt BA, Hanlon C, Naslund JA, Sikander S, Balaji M, et al. COVID-19 mental health impact and responses in low-income and middle-income countries: reimagining global mental health. Lancet Psychiatry 2021 Jun;8(6):535-550 [FREE Full text] [doi: 10.1016/S2215-0366(21)00025-0] [Medline: 33639109]
3. Smith K, Ostinelli E, Cipriani A. COVID-19 and mental health: a transformational opportunity to apply an evidence-based approach to clinical practice and research. Evid Based Ment Health 2020 May;23(2):45-46 [FREE Full text] [doi: 10.1136/ebmental-2020-300155] [Medline: 32317345]
4. Yao H, Chen JH, Xu YF. Patients with mental health disorders in the COVID-19 epidemic. Lancet Psychiatry 2020 Apr;7(4):e21 [FREE Full text] [doi: 10.1016/S2215-0366(20)30090-0] [Medline: 32199510]
5. Carley S, Horner D, Body R, Mackway-Jones K. Evidence-based medicine and COVID-19: what to believe and when to change. Emerg Med J 2020 Sep 10;37(9):572-575. [doi: 10.1136/emermed-2020-210098] [Medline: 32651176]
6. Chloros GD, Prodromidis AD, Giannoudis PV. Has anything changed in Evidence-Based Medicine? Injury 2023 May;54 Suppl 3:S20-S25 [FREE Full text] [doi: 10.1016/j.injury.2022.04.012] [Medline: 35525704]
7. Pearson H. How COVID broke the evidence pipeline. Nature 2021 May 12;593(7858):182-185. [doi: 10.1038/d41586-021-01246-x] [Medline: 33981057]
8. Horton R. Offline: COVID-19-the lessons that science forgot. Lancet 2022 Nov 19;400(10365):1753 [FREE Full text] [doi: 10.1016/S0140-6736(22)02358-3] [Medline: 36403609]
9. Pilbeam C, Tonkin-Crine S, Martindale AM, Atkinson P, Mableson H, Lant S, et al. How do healthcare workers 'do' guidelines? Exploring how policy decisions impacted UK healthcare workers during the first phase of the COVID-19 pandemic. Qual Health Res 2022 Apr 29;32(5):729-743 [FREE Full text] [doi: 10.1177/10497323211067772] [Medline: 35094621]
10. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: introduction. BMJ 2016 Jun 28;353:i2016 [FREE Full text] [doi: 10.1136/bmj.i2016] [Medline: 27353417]
11. Yao L, Guyatt GH, Djulbegovic B. Can we trust strong recommendations based on low quality evidence? BMJ 2021 Nov 25;375:n2833 [FREE Full text] [doi: 10.1136/bmj.n2833] [Medline: 34824089]
12. Elliott JH, Jeppesen BT. Rapid and living guidance for COVID-19. Ann Intern Med 2021 Aug;174(8):1171-1172 [FREE Full text] [doi: 10.7326/M21-2245] [Medline: 34029486]
13. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane rapid reviews methods group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol 2021 Feb;130:13-22 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.10.007] [Medline: 33068715]
14. Morgan RL, Florez I, Falavigna M, Kowalski S, Akl EA, Thayer KA, et al. Development of rapid guidelines: 3. GIN-McMaster guideline development checklist extension for rapid recommendations. Health Res Policy Syst 2018 Jul 13;16(1):63 [FREE Full text] [doi: 10.1186/s12961-018-0330-0] [Medline: 30005679]
15. Welcome to the Oxford Precision Psychiatry Lab. Oxford Precision Psychiatry Lab. URL: https://oxfordhealthbrc.nihr.ac.uk/our-work/oxppl/ [accessed 2023-09-18]

16. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
17. Zangani C, Ostinelli EG, Smith KA, Hong JS, Macdonald O, Reen G, et al. Impact of the COVID-19 pandemic on the global delivery of mental health services and telemental health: systematic review. JMIR Ment Health 2022 Aug 22;9(8):e38600 [FREE Full text] [doi: 10.2196/38600] [Medline: 35994310]
18. Reilly SE, Zane KL, McCuddy WT, Soulliard ZA, Scarisbrick DM, Miller LE, et al. Mental health practitioners' immediate practical response during the COVID-19 pandemic: observational questionnaire study. JMIR Ment Health 2020 Oct 01;7(9):e21237. [doi: 10.2196/21237] [Medline: 32931440]
19. Smith K, Ostinelli E, Macdonald O, Cipriani A. COVID-19 and telepsychiatry: development of evidence-based guidance for clinicians. JMIR Ment Health 2020 Aug 28;7(8):e21108 [FREE Full text] [doi: 10.2196/21108] [Medline: 32658857]
20. Ostinelli EG, Smith K, Zangani C, Ostacher MJ, Lingford-Hughes AR, Hong JS, et al. COVID-19 and substance use disorders: a review of international guidelines for frontline healthcare workers of addiction services. BMC Psychiatry 2022 Mar 31;22(1):228 [FREE Full text] [doi: 10.1186/s12888-022-03804-7] [Medline: 35361184]
21. Smith KA, Howard LM, Vigod SN, D'Agostino A, Cipriani A. Perinatal mental health and COVID-19: navigating a way forward. Aust N Z J Psychiatry 2023 Jul 28;57(7):937-943. [doi: 10.1177/00048674221137819] [Medline: 36440619]
22. Cipriani A, Smith K, Macdonald O. Covid-19 vaccination programme: where do people with mental health difficulties lie within the order of priority. BMJ Opinion. URL: http://tinyurl.com/2s43epxc [accessed 2023-11-24]
23. Smith K, Lambe S, Freeman D, Cipriani A. COVID-19 vaccines, hesitancy and mental health. Evid Based Ment Health 2021 May 13;24(2):47-48 [FREE Full text] [doi: 10.1136/ebmental-2021-300266] [Medline: 33849994]
24. COVID-19 and clinical management of mental health issues. The Oxford Precision Psychiatry Lab. URL: https://oxfordhealthbrc.nihr.ac.uk/our-work/oxppl/covid-19-and-mental-health-guidance/) [accessed 2023-09-18]
25. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
26. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
27. Protocol for: systematic review of syntheses (summaries) of guidelines for mental health during and after COVID-19. Open Science Framework. URL: https://osf.io/amsbj/ [accessed 2023-09-18]
28. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
29. mapsvg. URL: https://mapsvg.com/maps/world [accessed 2023-12-20]
30. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2023-12-20]
31. Alqahtani MM, Alkhamees HA, Alkhalaf AM, Alarjan SS, Alzahrani HS, AlSaad GF, et al. Toward establishing telepsychology guideline. Turning the challenges of COVID-19 into opportunity. Ethics Med Public Health 2021 Mar;16:100612. [doi: 10.1016/j.jemep.2020.100612] [Medline: 33283033]
32. Bojdani E, Rajagopalan A, Chen A, Gearin P, Olcott W, Shankar V, et al. COVID-19 pandemic: impact on psychiatric care in the United States. Psychiatry Res 2020 Jul;289:113069 [FREE Full text] [doi: 10.1016/j.psychres.2020.113069] [Medline: 32413707]
33. Burn W, Mudholkar S. Impact of COVID-19 on mental health: update from the United Kingdom. Indian J Psychiatry 2020 Sep;62(Suppl 3):S365-S372 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_937_20] [Medline: 33227076]
34. Choi KR, Records K, Low LK, Alhusen JL, Kenner C, Bloch JR, et al. Promotion of maternal-infant mental health and trauma-informed care during the COVID-19 pandemic. J Obstet Gynecol Neonatal Nurs 2020 Sep;49(5):409-415 [FREE Full text] [doi: 10.1016/j.jogn.2020.07.004] [Medline: 32795425]
35. Molebatsi K, Musindo O, Ntlantsana V, Wambua GN. Mental health and psychosocial support during COVID-19: a review of health guidelines in sub-Saharan Africa. Front Psychiatry 2021 May 20;12:571342 [FREE Full text] [doi: 10.3389/fpsyt.2021.571342] [Medline: 34093251]
36. Purushothaman S, Fung D, Reinders J, Garrett-Walcott S, Buda M, Moudgil V, et al. Electroconvulsive therapy, personal protective equipment and aerosol generating procedures: a review to guide practice during Coronavirus Disease 2019 (COVID-19) pandemic. Australas Psychiatry 2020 Dec;28(6):632-635. [doi: 10.1177/1039856220953699] [Medline: 32910692]
37. Rangaswamy T, Grover S, Tyagi V, Bhan A. How did the mental health care system in India respond to COVID 19 pandemic? Schizophr Bull Open 2022 Jan;3(1):sgac043 [FREE Full text] [doi: 10.1093/schizbullopen/sgac043] [Medline: 36348643]
38. Samy AL, Awang Bono S, Tan SL, Low WY. Mental health and COVID-19: policies, guidelines, and initiatives from the Asia-pacific region. Asia Pac J Public Health 2021 Nov;33(8):839-846. [doi: 10.1177/10105395211025901] [Medline: 34308673]

39. Scheibein F, Stowe MJ, Arya S, Morgan N, Shirasaka T, Grandinetti P, et al. Responding to COVID-19: emerging practices in addiction medicine in 17 countries. Front Psychiatry 2021 Mar 12;12:634309 [FREE Full text] [doi: 10.3389/fpsyt.2021.634309] [Medline: 33796034]
40. Sugarman DE, Busch AB. Telemental health for clinical assessment and treatment. BMJ 2023 Jan 16;380:e072398. [doi: 10.1136/bmj-2022-072398] [Medline: 36646462]
41. Royal College of Psychiatrists. URL: https://www.rcpsych.ac.uk/ [accessed 2023-12-06]
42. Royal College of Psychiatrists. URL: https://web.archive.org/web/20230202122900/https://www.rcpsych.ac.uk/about-us/responding-to-COVID-19/responding-to-COVID-19-guidance-for-clinicians [accessed 2023-12-11]
43. Alexander R, Ravi A, Barclay H, Sawhney I, Chester V, Malcolm V, et al. Guidance for the treatment and management of COVID-19 among people with intellectual disabilities. J Policy Pract Intellect Disabil 2020 Sep;17(3):256-269 [FREE Full text] [doi: 10.1111/jppi.12352] [Medline: 32837529]
44. Bikson M, Hanlon CA, Woods AJ, Gillick BT, Charvet L, Lamm C, et al. Guidelines for TMS/tES clinical services and research through the COVID-19 pandemic. Brain Stimul 2020 Jul;13(4):1124-1149 [FREE Full text] [doi: 10.1016/j.brs.2020.05.010] [Medline: 32413554]
45. Bilder RM, Postal KS, Barisa M, Aase DM, Cullum CM, Gillaspy SR, et al. Inter organizational practice committee recommendations/guidance for teleneuropsychology in response to the COVID-19 pandemic†. Arch Clin Neuropsychol 2020 Aug 28;35(6):647-659 [FREE Full text] [doi: 10.1093/arclin/acaa046] [Medline: 32666093]
46. Bilder RM, Postal KS, Barisa M, Aase DM, Cullum CM, Gillaspy SR, et al. InterOrganizational practice committee recommendations/guidance for teleneuropsychology (TeleNP) in response to the COVID-19 pandemic. Clin Neuropsychol 2020 Oct;34(7-8):1314-1334 [FREE Full text] [doi: 10.1080/13854046.2020.1767214] [Medline: 32673163]
47. Branjerdporn G, Sarma S, McCosker L, Dong V, Martin D, Loo C. "ECT should never stop": exploring the experiences and recommendations of ECT clinical directors and anesthetists about ECT during the COVID-19 pandemic. Front Psychiatry 2022 Sep 14;13:946748 [FREE Full text] [doi: 10.3389/fpsyt.2022.946748] [Medline: 36186870]
48. Brunson McClain M, Roanhorse TT, Harris B, Heyborne M, Zemantic PK, Azad G. School-based autism evaluations in the COVID-19 era. Sch Psychol 2021 Sep;36(5):377-387. [doi: 10.1037/spq0000447] [Medline: 34591588]
49. Cortese S, Asherson P, Sonuga-Barke E, Banaschewski T, Brandeis D, Buitelaar J, European ADHD Guidelines Group. ADHD management during the COVID-19 pandemic: guidance from the European ADHD Guidelines Group. Lancet Child Adolesc Health 2020 Jun;4(6):412-414 [FREE Full text] [doi: 10.1016/S2352-4642(20)30110-3] [Medline: 32311314]
50. Cortese S, Coghill D, Santosh P, Hollis C, Simonoff E, European ADHD Guidelines Group. Starting ADHD medications during the COVID-19 pandemic: recommendations from the European ADHD Guidelines Group. Lancet Child Adolesc Health 2020 Jun;4(6):e15 [FREE Full text] [doi: 10.1016/S2352-4642(20)30144-9] [Medline: 32405517]
51. Couturier J, Pellegrini D, Miller C, Bhatnagar N, Boachie A, Bourret K, et al. The COVID-19 pandemic and eating disorders in children, adolescents, and emerging adults: virtual care recommendations from the Canadian consensus panel during COVID-19 and beyond. J Eat Disord 2021 Apr 16;9(1):46 [FREE Full text] [doi: 10.1186/s40337-021-00394-9] [Medline: 33863388]
52. de Siqueira Rotenberg L, Nascimento C, Cohab Khafif T, Silva Dias R, Lafer B. Psychological therapies and psychoeducational recommendations for bipolar disorder treatment during COVID-19 pandemic. Bipolar Disord 2020 Sep;22(6):644-646 [FREE Full text] [doi: 10.1111/bdi.12981] [Medline: 32780503]
53. Fineberg NA, Van Ameringen M, Drummond L, Hollander E, Stein DJ, Geller D, et al. How to manage obsessive-compulsive disorder (OCD) under COVID-19: a clinician's guide from the International College of Obsessive Compulsive Spectrum Disorders (ICOCS) and the Obsessive-Compulsive and Related Disorders Research Network (OCRN) of the European College of Neuropsychopharmacology. Compr Psychiatry 2020 Jul;100:152174 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152174] [Medline: 32388123]
54. Geddes MR, O'Connell ME, Fisk JD, Gauthier S, Camicioli R, Ismail Z, Alzheimer Society of Canada Task Force on Dementia Care Best Practices for COVID‐19. Remote cognitive and behavioral assessment: report of the Alzheimer Society of Canada Task Force on dementia care best practices for COVID-19. Alzheimers Dement (Amst) 2020 Sep 22;12(1):e12111 [FREE Full text] [doi: 10.1002/dad2.12111] [Medline: 32999916]
55. Jauhar S, Lai S, Bonoldi I, Salazar de Pablo G, di Forti M, Alameda L, et al. Early intervention in psychosis during the COVID-19 pandemic: maudsley recommendations. Eur Neuropsychopharmacol 2021 Jun;47:130-135 [FREE Full text] [doi: 10.1016/j.euroneuro.2021.02.005] [Medline: 33642195]
56. Kaltenbach E, McGrath PJ, Schauer M, Kaiser E, Crombach A, Robjant K. Practical guidelines for online Narrative Exposure Therapy (e-NET) - a short-term treatment for posttraumatic stress disorder adapted for remote delivery. Eur J Psychotraumatol 2021 Mar 01;12(1):1881728 [FREE Full text] [doi: 10.1080/20008198.2021.1881728] [Medline: 34025923]
57. Lapid MI, Seiner S, Heintz H, Hermida AP, Nykamp L, Sanghani SN, et al. Electroconvulsive therapy practice changes in older individuals due to COVID-19: expert consensus statement. Am J Geriatr Psychiatry 2020 Nov;28(11):1133-1145 [FREE Full text] [doi: 10.1016/j.jagp.2020.08.001] [Medline: 32863137]
58. Martin-Khan M, Bail K, Yates MW, Thompson J, Graham F, Cognitive ImpairmentCOVID-19, Hospital Care Guidance Committee. Interim guidance for health-care professionals and administrators providing hospital care to adult patients with

cognitive impairment, in the context of COVID-19 pandemic. Australas J Ageing 2020 Sep;39(3):283-286 [FREE Full text] [doi: 10.1111/ajag.12831] [Medline: 33051999]
59. McGrath J. ADHD and COVID-19: current roadblocks and future opportunities. Ir J Psychol Med 2020 Sep;37(3):204-211 [FREE Full text] [doi: 10.1017/ipm.2020.53] [Medline: 32434606]
60. Palmese CA, Wyman-Chick KA, Racine C, Pollak LE, Lin G, Farace E, Neuropsychology Focus Group of the Functional Neurosurgical Working Group (FNSWG) of the Parkinson Study Group (PSG). Assessment of deep brain stimulation candidacy during the COVID-19 pandemic: lessons learned and future directions for neuropsychologists. Clin Neuropsychol 2022 Jan;36(1):72-84. [doi: 10.1080/13854046.2021.1929496] [Medline: 34030595]
61. Peros OM, Webb L, Fox S, Bernstein A, Hoffman L. Conducting exposure-based groups via telehealth for adolescents and young adults with social anxiety disorder. Cogn Behav Pract 2021 Nov;28(4):679-689 [FREE Full text] [doi: 10.1016/j.cbpra.2021.04.001] [Medline: 34690482]
62. Pinzón-Espinosa J, Valdés-Florido MJ, Riboldi I, Baysak E, Vieta E, EFPT Psychiatry Across Borders Working Group. The COVID-19 pandemic and mental health of refugees, asylum seekers, and migrants. J Affect Disord 2021 Feb 01;280(Pt A):407-408. [doi: 10.1016/j.jad.2020.11.029] [Medline: 33227670]
63. Postal KS, Bilder RM, Lanca M, Aase DM, Barisa M, Holland AA, et al. InterOrganizational practice committee guidance/recommendation for models of care during the novel coronavirus pandemic. Clin Neuropsychol 2021 Jan;35(1):81-98 [FREE Full text] [doi: 10.1080/13854046.2020.1801847] [Medline: 32996823]
64. Richmond JS, Dragatsi D, Stiebel V, Rozel JS, Rasimas JJ. American association for emergency psychiatry recommendations to address psychiatric staff shortages in emergency settings. Psychiatr Serv 2021 Apr 01;72(4):437-443. [doi: 10.1176/appi.ps.201900501] [Medline: 33467869]
65. Sabe M, Dorsaz O, Huguelet P, Kaiser S. Toxicity of psychotropic drugs in patients with COVID-19: a systematic review. Gen Hosp Psychiatry 2021 May;70:1-9 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.02.006] [Medline: 33631694]
66. Santosh P, Cortese S, Hollis C, Bölte S, Daley D, Coghill D, et al. Remote assessment of ADHD in children and adolescents: recommendations from the European ADHD Guidelines Group following the clinical experience during the COVID-19 pandemic. Eur Child Adolesc Psychiatry 2023 Jun;32(6):921-935 [FREE Full text] [doi: 10.1007/s00787-023-02148-1] [Medline: 36764972]
67. Schlief M, Saunders KR, Appleton R, Barnett P, Vera San Juan N, Foye U, et al. Synthesis of the evidence on what works for whom in telemental health: rapid realist review. Interact J Med Res 2022 Sep 29;11(2):e38239 [FREE Full text] [doi: 10.2196/38239] [Medline: 35767691]
68. Scott TM, Marton KM, Madore MR. A detailed analysis of ethical considerations for three specific models of teleneuropsychology during and beyond the COVID-19 pandemic. Clin Neuropsychol 2022 Jan;36(1):24-44. [doi: 10.1080/13854046.2021.1889678] [Medline: 33761834]
69. Starace F, Ferrara M. COVID-19 disease emergency operational instructions for mental health departments issued by the Italian Society of Epidemiological Psychiatry. Epidemiol Psychiatr Sci 2020 Mar 31;29:e116 [FREE Full text] [doi: 10.1017/S2045796020000372] [Medline: 32228737]
70. Thompson AD, Berkman ER, Simmons SW, Porter KM, Kroon L, Goldman R, et al. Ethical considerations in balancing use of seclusion and restraint with risk of COVID-19 exposure: recommendations for youth inpatient psychiatry units. J Am Acad Child Adolesc Psychiatry 2022 Nov;61(11):1319-1321 [FREE Full text] [doi: 10.1016/j.jaac.2022.04.016] [Medline: 35513190]
71. Tiamson-Kassab M. ACLP response to the recommendations of the ACLP task force on lessons learned from COVID-19 pandemic. J Acad Consult Liaison Psychiatry 2021;62(4):375-376 [FREE Full text] [doi: 10.1016/j.jaclp.2021.04.005] [Medline: 34039531]
72. van Daele T, Karekla M, Kassianos AP, Compare A, Haddouk L, Salgado J, et al. Recommendations for policy and practice of telepsychotherapy and e-mental health in Europe and beyond. J Psychother Integr 2020 Jun;30(2):160-173 [FREE Full text] [doi: 10.1037/int0000218]
73. Waller G, Pugh M, Mulkens S, Moore E, Mountford VA, Carter J, et al. Cognitive-behavioral therapy in the time of coronavirus: clinician tips for working with eating disorders via telehealth when face-to-face meetings are not possible. Int J Eat Disord 2020 Jul;53(7):1132-1141 [FREE Full text] [doi: 10.1002/eat.23289] [Medline: 32383530]
74. Wells SY, Morland LA, Wilhite ER, Grubbs KM, Rauch SA, Acierno R, et al. Delivering prolonged exposure therapy via videoconferencing during the COVID-19 pandemic: an overview of the research and special considerations for providers. J Trauma Stress 2020 Aug;33(4):380-390 [FREE Full text] [doi: 10.1002/jts.22573] [Medline: 32881116]
75. Millard T, Elliott JH, Green S, Tendal B, Vogel JP, Norris S, National COVID-19 Clinical Evidence Taskforce. Awareness, value and use of the Australian living guidelines for the clinical care of people with COVID-19: an impact evaluation. J Clin Epidemiol 2022 Mar;143:11-21 [FREE Full text] [doi: 10.1016/j.jclinepi.2021.11.035] [Medline: 34852274]
76. Caring for people with COVID-19. The National Clinical Evidence Taskforce. URL: https://clinicalevidence.net.au/covid-19/ [accessed 2023-09-01]
77. Elliott J, Lawrence R, Minx JC, Oladapo OT, Ravaud P, Tendal Jeppesen B, et al. Decision makers need constantly updated evidence synthesis. Nature 2021 Dec;600(7889):383-385 [FREE Full text] [doi: 10.1038/d41586-021-03690-1] [Medline: 34912079]

78. Therapeutics and COVID-19: living guideline, 13 January 2023. World Health Organization. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-therapeutics-2023.1 [accessed 2023-09-18]
79. Home page. Global Alliance for Living Evidence on aNxiety, depressiOn and pSychosis. URL: https://galenos.org.uk/ [accessed 2023-09-18]
80. Cipriani A, Seedat S, Milligan L, Salanti G, Macleod M, Hastings J, et al. New living evidence resource of human and non-human studies for early intervention and research prioritisation in anxiety, depression and psychosis. BMJ Ment Health 2023 Jun 08;26(1):e300759 [FREE Full text] [doi: 10.1136/bmjment-2023-300759] [Medline: 37290906]
81. Golz C, Peter KA, Müller TJ, Mutschler J, Zwakhalen SM, Hahn S. Technostress and digital competence among health professionals in Swiss psychiatric hospitals: cross-sectional study. JMIR Ment Health 2021 Nov 04;8(11):e31408 [FREE Full text] [doi: 10.2196/31408] [Medline: 34734840]
82. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
83. Smith K, Torous J, Cipriani A. Teaching telepsychiatry skills: building on the lessons of the COVID-19 pandemic to enhance mental health care in the future. JMIR Ment Health 2022 Oct 14;9(10):e37939 [FREE Full text] [doi: 10.2196/37939] [Medline: 35358948]
84. Johansson M, Guyatt G, Montori V. Guidelines should consider clinicians' time needed to treat. BMJ 2023 Jan 03;380:e072953 [FREE Full text] [doi: 10.1136/bmj-2022-072953] [Medline: 36596571]
85. Dobler CC, Harb N, Maguire CA, Armour CL, Coleman C, Murad MH. Treatment burden should be included in clinical practice guidelines. BMJ 2018 Oct 12;363:k4065 [FREE Full text] [doi: 10.1136/bmj.k4065] [Medline: 30314981]
86. Finkelstein A, Persson P, Polyakova M, Shapiro JM. A taste of their own medicine: guideline adherence and access to expertise. Am Econ Rev Insights 2022 Dec;4(4):507-526 [FREE Full text] [doi: 10.1257/aeri.20210591]
87. Smallwood N, Bismark M, Willis K. Burn-out in the health workforce during the COVID-19 pandemic: opportunities for workplace and leadership approaches to improve well-being. BMJ Lead 2023 Sep 10;7(3):178-181 [FREE Full text] [doi: 10.1136/leader-2022-000687] [Medline: 37192091]
88. Vera San Juan N, Martin S, Badley A, Maio L, Gronholm PC, Buck C, et al. Frontline health care workers' mental health and well-being during the first year of the COVID-19 pandemic: analysis of interviews and social media data. J Med Internet Res 2023 Aug 14;25:e43000 [FREE Full text] [doi: 10.2196/43000] [Medline: 37402283]
89. Maxmen A. Has COVID taught us anything about pandemic preparedness? Nature 2021 Aug;596(7872):332-335 [FREE Full text] [doi: 10.1038/d41586-021-02217-y] [Medline: 34389832]
90. COVID-19: make it the last pandemic. The Independent Panel. URL: https://theindependentpanel.org/wp-content/uploads/2021/05/COVID-19-Make-it-the-Last-Pandemic_final.pdf, [accessed 2023-11-24]
91. Moeti M, Gao GF, Herrman H. Global pandemic perspectives: public health, mental health, and lessons for the future. Lancet 2022 Aug;400(10353):e3-e7 [FREE Full text] [doi: 10.1016/s0140-6736(22)01328-9]
92. Smith KA, Blease C, Faurholt-Jepsen M, Firth J, Van Daele T, Moreno C, et al. Digital mental health: challenges and next steps. BMJ Ment Health 2023 Feb 17;26(1):e300670 [FREE Full text] [doi: 10.1136/bmjment-2023-300670] [Medline: 37197797]
93. Torous J, Benson NM, Myrick K, Eysenbach G. Focusing on digital research priorities for advancing the access and quality of mental health. JMIR Ment Health 2023 Apr 24;10:e47898 [FREE Full text] [doi: 10.2196/47898] [Medline: 37093624]
94. Cohen A, Naslund JA, Chang S, Nagendra S, Bhan A, Rozatkar A, et al. Relapse prediction in schizophrenia with smartphone digital phenotyping during COVID-19: a prospective, three-site, two-country, longitudinal study. Schizophrenia (Heidelb) 2023 Jan 27;9(1):6 [FREE Full text] [doi: 10.1038/s41537-023-00332-5] [Medline: 36707524]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
GALENOS: Global Alliance for Living Evidence on Anxiety, Depression, and Psychosis
NHS: National Health Service
OxPPL: Oxford Precision Psychiatry Lab
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by J Torous; submitted 20.09.23; peer-reviewed by U Mehta, N Benson; comments to author 21.10.23; revised version received 30.10.23; accepted 03.11.23; published 22.12.23.

Please cite as:
Smith KA, Ostinelli EG, Ede R, Allard L, Thomson M, Hewitt K, Brown P, Zangani C, Jenkins M, Hinze V, Ma G, Pothulu P, Henshall C, Malhi GS, Every-Palmer S, Cipriani A
Assessing the Impact of Evidence-Based Mental Health Guidance During the COVID-19 Pandemic: Systematic Review and Qualitative Evaluation
JMIR Ment Health 2023;10:e52901
URL: https://mental.jmir.org/2023/1/e52901
doi:10.2196/52901
PMID:38133912

Review

Feasibility, Adherence, and Effectiveness of Blended Psychotherapy for Severe Mental Illnesses: Scoping Review

Yamina Ehrt-Schäfer[1,2], MSc; Milan Rusmir[1], BSc; Johannes Vetter[1,2], PhD; Erich Seifritz[2], Prof Dr Med; Mario Müller[1,2], PhD; Birgit Kleim[1,2], Prof Dr

[1]Department of Psychology, University of Zurich, Zurich, Switzerland

[2]Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital, University of Zurich, Zurich, Switzerland

Corresponding Author:
Yamina Ehrt-Schäfer, MSc
Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital
University of Zurich
Lenggstrasse 31
Zurich, 8032
Switzerland
Phone: 41 +41583842817
Email: yamina.ehrt@pukzh.ch

Abstract

Background: Blended psychotherapy (bPT) combines face-to-face psychotherapy with digital interventions to enhance the effectiveness of mental health treatment. The feasibility and effectiveness of bPT have been demonstrated for various mental health issues, although primarily for patients with higher levels of functioning.

Objective: This scoping review aims to investigate the feasibility, adherence, and effectiveness of bPT for the treatment of patients with severe mental illnesses (SMIs).

Methods: Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, we conducted searches in PubMed, MEDLINE, Embase, PsycINFO, and PsycArticles for studies published until March 23, 2023.

Results: Out of 587 screened papers, we incorporated 25 studies encompassing 23 bPT interventions, involving a total of 2554 patients with SMI. The intervention formats and research designs exhibited significant variation. Our findings offer preliminary evidence supporting the feasibility of bPT for SMI, although there is limited research on adherence. Nevertheless, the summarized studies indicated promising attrition rates, spanning from 0% to 37%, implying a potential beneficial impact of bPT on adherence to SMI treatment. The quantity of evidence on the effects of bPT for SMI was limited and challenging to generalize. Among the 15 controlled trials, 4 concluded that bPT interventions were effective compared with controls. However, it is noteworthy that 2 of these studies used the same study population, and the control groups exhibited significant variations.

Conclusions: Overall, our review suggests that while bPT appears promising as a treatment method, further research is necessary to establish its effectiveness for SMI. We discuss considerations for clinical implementation, directions, and future research.

(JMIR Ment Health 2023;10:e43882) doi:10.2196/43882

KEYWORDS

blended psychotherapy; severe mental illnesses; digital health intervention; e-mental health; scoping review

Introduction

Mental illnesses are widespread globally, with estimated lifetime prevalence rates ranging from 12% to 47% [1,2]. There is a potential increase in these rates [3,4], placing a substantial burden on individuals [5,6], their families [7], and both economic and public health systems [8]. Psychotherapy stands as an evidence-based primary treatment for the majority of mental disorders [9], as advocated by both national and international guidelines, such as those from the American Psychological Association (APA) and the National Institute for Health and Care Excellence (NICE). Blended psychotherapy (bPT), alternatively termed “blended therapy” or “blended psychological treatment,” involves combining conventional face-to-face therapy with digital intervention components within a unified treatment protocol. This integration aims to amplify

the efficacy and cost-effectiveness of mental health care [9-11]. In recent years, the literature has detailed various modes of bPT, encompassing integrated use (where face-to-face and digital components are concurrently administered) and sequential use (where the digital intervention takes place either before or after the face-to-face component) [11]. From a conceptual standpoint, Bielinski and colleagues [10] make a further distinction between transformational blends (using digital and face-to-face components in an interconnected manner) and additional blends (involving more independent combinations of digital and face-to-face therapy components).

Recent research offers evidence supporting the feasibility and effectiveness of bPT when compared with control conditions [11]. This holds true across a spectrum of mental health conditions, including depression, anxiety disorders, and substance abuse, and is applicable to both adult and adolescent patients [11-13]. bPT offers several advantages and potentials in comparison to conventional face-to-face psychotherapy. These include heightened exposure to treatment within and beyond the therapy session, fostering patient empowerment in self-management [14], improved consolidation of therapy content [15], enhanced transfer of in-session outcomes into everyday life [16], and the sustenance of therapeutic changes beyond the acute treatment phase. Moreover, bPT has the potential to contribute to the development and stabilization of the therapeutic relationship over time [11,17]. For therapists, the benefits of using bPT are potential time savings [11] and the capacity to customize treatments more flexibly according to patients' needs. Furthermore, the incorporation of digital data into the bPT intervention, whether actively (eg, through direct symptom monitoring, ecological momentary assessment) or passively (eg, via wearable devices and device metrics), grants therapists access to detailed patient data. These data can be instrumental in enhancing treatment through feedback mechanisms and early detection of crises or emergency situations [18-20].

Numerous studies have concentrated on patients with relatively high levels of functioning [13,21]. Nevertheless, there is limited research on the feasibility and effects of bPT for patients with severe mental illnesses (SMIs) [12,13,22,23], even though there is mounting evidence regarding the effectiveness of psychotherapy for this population [24-28]. Individuals with SMI, marked by considerable limitations in intrapersonal and social functioning [24,29], frequently necessitate long-term, intensive, multidisciplinary treatment and care, which can be financially burdensome [30,31]. These patients, with an estimated annual prevalence rate of 2.33 per 1000 for all disorders [32], frequently make up the majority of patients in psychiatric hospitals and long-term psychiatric care. Consequently, our objective is to consolidate the current evidence on the feasibility and effectiveness of bPT for SMI. In particular, we aim to scrutinize the characteristics of bPT interventions implemented for SMI patient groups, examining aspects such as feasibility, adherence, and effectiveness.

Methods

Study Design

We performed a scoping review, following the framework stages proposed by Arksey and O'Malley [33] and further refined by Levac and colleagues [34,35]. The reporting adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [36] (Multimedia Appendix 1). This review approach is designed to address an exploratory research question with the aim of mapping the existing literature and key concepts in a defined field. It serves to identify gaps and types of evidence in research [35].

Search Strategy

We performed a systematic search of the PubMed, MEDLINE, Embase (via Elsevier), PsycINFO, and PsycArticles databases. The search terms included the following keywords: (1) treatment ((blended) AND (therapy)); (2) broader digital health ((internet) OR (mobile) OR (online) OR (digital)); and (3) population ("severe mental illness" OR "serious mental illness" OR "schizophrenia" OR "psychosis" OR "psychotic" OR "psychotic disorder*" OR "schizoaffective" OR "schizo-affective" OR "bipolar disorder*" OR "mania" OR "manic" OR "bipolar" OR "depression" OR "major depressive disorder" OR "antipsychotic" OR "personality disorder" OR "inpatient"). The search strings were converted for each database. We also manually searched the bibliographies of relevant studies and reviews using the snowball principle to identify further eligible articles. We included publications from the date of inception up to March 23, 2023, in the search.

Study Selection

Eligibility Criteria

Studies were included if they met the following criteria: (1) peer-reviewed articles written in English; (2) had a focus on blended psychotherapeutic study interventions (ie, a combination of face-to-face psychotherapy treatment with online, mobile, or other digital content [10]); (3) inclusion of a study population with severe psychiatric disorders. SMI could be defined by the study author; a psychiatric diagnosis reflecting common definitions of SMI [29,32], such as schizophrenia spectrum disorders or other psychotic disorders, bipolar disorders, severe depression, or severe personality disorders; severity of a diagnosis, indicated by psychometric measures (eg, Beck's Depression Inventory Second Edition [BDI-II]≥30; 9-item Patient Health Questionnaire [PHQ-9]≥20; Generalized Anxiety Disorder [GAD]≥15); or inpatient or previous inpatient treatment setting (indicating greater functional impairment and severity). Articles reporting on (4) feasibility or adherence or efficacy or effectiveness of the respective bPT intervention were included.

Because of the novelty of the field, there was no restriction on the variety of study designs. However, we excluded articles on digital stand-alone interventions, as well as those on digital therapist-assisted (guided) interventions that lacked face-to-face therapist-patient contact. Additionally, e-mental health applications solely focused on symptom monitoring (ambulatory

assessment) without a psychological intervention component were excluded.

Study Selection Process

The relevant research papers that were identified underwent screening based on information available from the study title and abstract. One reviewer (MR) piloted a study screening manual, which was subsequently discussed and further refined by the study team through group consensus. Abstracts and full-text articles were independently reviewed by 3 reviewers (MR, YE-S, and Lilian Fraefel). Any disagreements were resolved through discussions with the senior researcher (BK), who made the final decision regarding inclusion in cases of uncertainty.

Quality Assessment

In accordance with the PRISMA-ScR guidelines for conducting scoping reviews [36,37], formal quality appraisal was not incorporated into the study selection process. Nevertheless, we evaluated the study quality of the selected articles using a quality assessment checklist developed by Ellis and colleagues [38]. This was done to ensure that the included studies adhered to standard research norms and to provide insights for future research. The scores of the studies ranged from 9 to 11 (out of a maximum of 12), signifying a high level of quality for the selected articles. Further details can be found in Multimedia Appendix 2 [39-61].

Data Charting

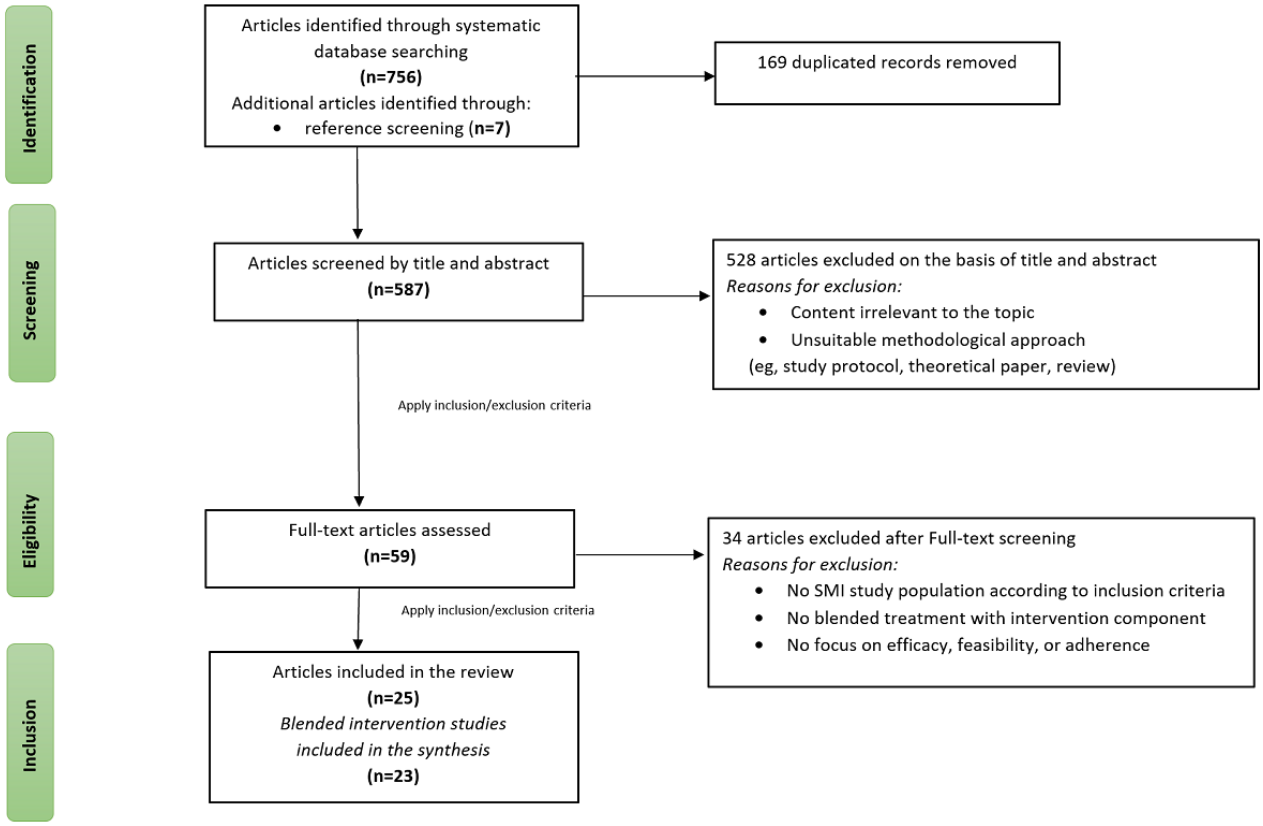
Data charting was undertaken as a collaborative and iterative process. A customized data extraction worksheet was created and consistently adjusted to capture all data that could be pertinent to addressing the outlined research questions. The ultimate data chart underwent scrutiny for validity and data accuracy by 6 members of the research team. The comprehensive data extraction form is available in this publication (refer to Multimedia Appendix 3 [30,39-62]).

Results

Study Selection

Following the removal of duplicates, a total of 587 articles underwent screening, and 59 papers were assessed in full for eligibility. We identified 25 articles that were relevant to the scope of the review. Among these, 23 papers [39-61] reported as main studies on aspects of feasibility, adherence, and effectiveness of bPT interventions, while 2 [30,62] were substudies analyzing aspects of adherence only. For the quantitative analysis of the review, these 2 substudies will not be included. We specifically utilized the information for the analysis of adherence. The outcomes of the article selection process are illustrated in Figure 1.

Figure 1. PRISMA flow diagram. The overall process of article selection following PRISMA-ScR guidelines. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA-ScR: PRISMA Extension for Scoping Reviews; SMI: severe mental illness.



Study Characteristics

Tables 1 and 2 outline the characteristics of the included studies (n=23) and patients (n=2554). The papers incorporated into the review were predominantly from research groups in Germany (8/23, 35%) [43,44,46-48,53,57,58], followed by the United Kingdom (5/23, 22%) [39,40,51,54,59], the Netherlands (4/23, 17%) [49,50,55,61], and Australia (3/23, 13%) [42,45,56]; 1 study was conducted in the United States [60], and 2 were structured as transnational collaborative research [41,52]. We identified articles published between 2012 and 2021.

Among the bPT studies, the majority (12/23, 52%) utilized randomized controlled trial designs [43-49,51,52,57,59,61]; 2/23 studies (9%) were nonrandomized controlled trials [42,55], and 1 used a quasi-experimental 4-group design [39]. The remaining studies comprised 6/23 (26%) observational pilots [50,53,54,56,58,60], 1 observational study [40], and 1 qualitative evaluation [41].

Among the bPT trials with control groups (15/23; 65%) [39,40,42-49,51,52,57,59,61], 11 utilized treatment as usual (TAU) groups [40,45-49,51,52,57,59,61], 3 used an active control group [39,43,44], and 1 study had a waiting list control group [42]. The bPT interventions covered a range of SMI study populations, involving a total of 2554 patients with SMI. Notably, for the trials conducted by Zwerenz and colleagues [43,44], the same study population was utilized to generate different data sets—1 on inpatient treatment and another for follow-up on the same population after inpatient treatment.

Table 1. The main features of the blended psychotherapy studies included (n=23).

Article characteristics	Value, n
Country of origin (n=23)	
Australia/United Kingdom	1
Germany/Denmark	1
United States	1
Australia	3
The Netherlands	4
United Kingdom	5
Germany	8
Publication year (n=23)	
2012	1
2013	2
2016	3
2017	2
2018	4
2019	2
2020	3
2021	6
Research design (n=23)	
Randomized controlled trial	12
Nonrandomized controlled studies	2
Quasi-experimental	1
Observational pilot	6
Observational	1
Qualitative evaluation	1
Control group designs (n=15)	
Treatment as usual	11
Active control group	3
Waiting list	1

Table 2. Study populations.[a]

Patient demographics	Value (N=2554), n
Disorder specific (n=14)	
Hoarding disorder	1
Bulimia nervosa	1
Anorexia nervosa	2
Borderline personality disorder	2
Major depression[a]	3
Psychotic disorder	5
Transdiagnostic/symptom oriented (n=8)	
Affective symptoms	3
Psychotic symptoms	2
Severe mental illness, various disorders[b]	3

[a]The study population of Zwerenz et al [43,44] counts as 1 population. The total is accordingly adjusted.

[b]Patients fulfilling SMI criteria according to [29].

We identified 14/22 (64%) disorder-specific study populations, comprising patients with psychotic disorders (n=5), severe major depression (n=3), borderline personality disorder (n=2), anorexia nervosa (n=2), bulimia nervosa (n=1), and hoarding disorder (n=1). Additionally, we identified 8/23 (35%) transdiagnostic or symptom-oriented bPT interventions: 3 targeting affective symptoms, 2 addressing psychotic symptoms, and 2 designed for patients meeting the criteria of SMI as defined by the Dutch consensus group on SMI [29].

Characteristics of bPT Interventions

We identified a variety of bPT intervention formats (Tables 3-6), primarily focusing on outpatient treatment (21/23, 91%). The duration of the analyzed interventions ranged from 2 sessions to 18 months, varying based on the specific objectives of each intervention. Among the 23 interventions, 17 (74%) used an integrated blending approach [39,43-45,49-61], with 12 of them adopting a concept of transformational blending, and 5 [42,45-48] utilizing an additional blend. Additionally, 6/23 (26%) studies designed the bPT interventions in a sequential manner [40,42,45-48]. One study incorporated the digital component before intensive face-to-face psychotherapy [40], while 5 studies [42,45-48] included the digital component as part of an after-care concept following intensive face-to-face treatment. Among the sequentially designed interventions, 2 used a transformational blend [40,42], and 4 utilized an additional blend [45-48].

Table 3. Characteristics of bPT[a] study interventions: Integrated use of digital and face-to-face components with complementary contents of both components (transformational blend).

Study	Project/digital component duration	Study population	Treatment setting	Psychotherapy treatment	Digital component and face-to-face component	Peer group
Beentjes et al [61]	e-IMR/12 months	SMI[b]; various disorders	OP[c]	IMR[d]/recovery	• Digital: An online platform providing IMR content with added illustrative videos showing peer testimonials to encourage participants to talk more freely about themselves and to take steps in their recovery process; additionally, problem-solving forms at the end of each module, registration of successful coping strategies, and a symptom-monitoring page. • Face-to-face: IMR-based care with standardized, curriculum-based group therapy providing psychoeducation on skills for illness management and recovery; additionally, extensive inpatient or outpatient psychiatric treatment including case management.	Yes
Bell et al [52]	SAVVy/3 weeks	Psychotic disorder	OP	CBT[e]/CSE[f]	• Digital: usage of an existing smartphone app with EMA[g] and EMI[h]; EMA included 10 surveys per day for 6 days; EMI included 5 coping reminders per day and evening survey with feedback for every day; EMI use was divided into 2 periods of 10 days each. • Face-to-face: 4 sessions between EMA and EMI usage, discussion on EMA/EMI usage, app training and app coding, analyzing and further developing of coping strategies, and providing positive reinforcement.	No
Bendig et al [53]	Blended SST/4 weeks	SMI; various disorders	IP[i]	CBT/social skills	• Digital: an online program with modules on social skills training, including psychoeducation and exercises on social competence/social skills. • Face-to-face: group training on social competence/social skills.	No
Blankers et al [55]	Blended FACT[j]/Arkin mental health care/3 months	SMI; various disorders	OP	FACT/recovery	• Digital: an online platform for i-FACT (flexible assertive community treatment), using the internet portal “myMentrum”; psychoeducation videos, leisure activity bulletin board, agenda for scheduling appointments with the psychiatric nurse; a web forum to establish contact with other patients; additional use of Skype for contact with psychiatric nurses (either scheduled or during office hours). • Face-to-face: FACT.	Yes
Craig et al [59]	AVATAR/12 weeks	Psychotic disorder; auditory verbal; symptoms	OP	RA[k]/IPT[l]	• Digital: computer-based AVATAR program with therapist facilitation of discussion between patient and avatar. • Face-to-face: standard psychiatric care, including progress discussion after each session.	No
Garety et al [51]	SlowMo/12 weeks	Psychotic disorder; paranoia symptoms	OP	CBTp[m]	• Digital: a web-based and mobile app with i-CBTp[n]; interactive features including psychoeducation, animated vignettes, games, and personalized content on coping with distressing paranoia symptoms. • Face-to-face: digitally supported CBTp consisting of 8 individual, face-to-face sessions (60-90 minutes) in accordance with a clinical manual.	No
Granholm et al [60]	CBT2go/24 weeks	Psychotic disorder	OP	CBT/cognitive skills training	• Digital: smartphone mobile app with cognitive skills content, personalized activity prompts, pleasure-savoring prompts (photos and journal), and reminder alerts for activities. • Face-to-face: modified 12-session version of the Cognitive Skills Module of CBSST[o], weekly group therapy sessions.	Yes
Jacob et al [58]	priovi/12 months	BPD[p]	OP	ST[q]	• Digital: an online program with 8 ST-based modules; simulated dialogs via chatbot, daily SMS text messages, exercises; and a hotline for technical support. • Face-to-face: weekly PT[r] sessions.	No

Study	Project/digital component duration	Study population	Treatment setting	Psychotherapy treatment	Digital component and face-to-face component	Peer group
Kooistra et al [49]	dCBT[s] program/9 months	Severe depression	OP	CBT	• Digital: an online program with modular CBT-based content, including psychoeducation, homework tasks/exercises, and an open-ended session evaluation for feedback on sessions and reflective thinking; and an additional email reminder to encourage patients to access the online platform. • Face-to-face: weekly semistructured CBT-PT session.	No
Kooistra et al [50]	dCBT program/10 weeks	Severe depression	OP	CBT	• Digital: an online program with modular CBT-based content, including psychoeducation, homework tasks/exercises, and an open-ended session evaluation for feedback on sessions and reflective thinking; and an additional email reminder to encourage patients to access online platform. • Face-to-face: weekly semistructured CBT-PT sessions.	No
Sedgwick et al [54]	GRASP/4 weeks	Psychotic disorder	OP	CBT/social cognition training	• Digital: a smartphone app with social cognition training content, including homework modules covering different topics and interactive features such as videos, games, and tasks. • Face-to-face: social cognition training group therapy with an introduction to the homework app.	No
Thomas et al [56]	SMART/3 months	Psychotic symptoms	OP	Recovery	• Digital: a website with a series of peer video interviews of people with lived experience of psychosis discussing how they had navigated issues within their own recovery and recovery modules including psychoeducation, additional reflective exercises, and self-monitoring. • Face-to-face: 8 sessions with a mental health care worker, discussing contents of the website material.	No

[a]bPT: blended psychotherapy.

[b]SMI: severe mental illness.

[c]OP: outpatient.

[d]IMR: Illness Management and Recovery Program.

[e]CBT: cognitive behavioral therapy.

[f]CSE: coping strategy enhancement.

[g]EMA: ecological momentary assessment.

[h]EMI: ecological momentary intervention.

[i]IP: inpatient.

[j]FACT: flexible assertive community treatment.

[k]RA: relational approach.

[l]IPT: interpersonal therapy.

[m]CBTp: cognitive behavioral therapy for psychosis.

[n]i-CBTp: internet-based cognitive behavioral therapy for psychosis.

[o]CBSST: cognitive-behavioral social skills training.

[p]BPD: borderline personality disorder.

[q]ST: schema therapy.

[r]PT: psychotherapy.

[s]dCBT: digital cognitive behavioral therapy.

Table 4. Characteristics of bPT[a] study interventions: integrated use of digital and face-to-face components with independent contents of both components (additional blend).

Study	Project/digital component duration	Study population	Treatment setting	Psychotherapy treatment	Digital component and face-to-face component	Peer group
Austin et al [41]	IMPACHS (Monsenso and Time 4U)/6 months	Psychotic disorder	OP[b]	CBTp[c]	• Digital: 2 integrated mHealth components: (1) the Monsenso smartphone app for self-monitoring of mood and psychotic symptoms including a tool for self-reflection, a library function for psychoeducation, an action-plan function for customizable plans on coping strategies, tools to identify and describe factors that lead to symptom exacerbation; (2) Time 4U, an internet-based training and learning management system, comprising 8 interactive CBTp modules (texts, videos, tasks, and follow-up questions); additionally, an interface that enables clinicians to monitor and support both components. • Face-to-face: CBTp-based assertive outreach; psychoeducation with a focus on symptom coping, recovery, and functioning facilitation.	No
Cardi et al [39]	VodCast/2 sessions	Anorexia nervosa	IP[d]/OP	CBT	• Digital: vodcasts (a short video clip to encourage reappraisal and acceptance of food), auditory and visual imagery with background relaxation and soothing music, psychoeducational content, and motivational interviewing. • Face-to-face: standard inpatient treatment (CBT)/standard outpatient treatment (unspecified).	No
Klein et al [57]	REVISIT-BPD/12 months	BPD[e]	OP	ST[f]	• Digital: an online program comprising 8 modules with contents of scheme therapy, simulated dialogs via chatbot, daily SMS text messages, and exercises; and a hotline for technical support. • Face-to-face: standard outpatient treatment (unspecified).	No
Zwerenz et al [44]	Deprexis/12 weeks	Severe depression	IP	CBT/psychodynamic psychotherapy	• Digital: an online program comprising a web-based self-help program (CBT-based modules), conducted 2 h per week (scheduled). • Face-to-face: standard inpatient treatment (psychodynamic).	No
Zwerenz et al [43]	Deprexis/6 months	Severe depression	OP	CBT	• Digital: an online program comprising a web-based self-help program (CBT-based modules). • Face-to-face: standard outpatient treatment (unspecified).	No

[a]bPT: blended psychotherapy.

[b]OP: outpatient.

[c]CBTp: cognitive behavioral therapy for psychosis.

[d]IP: inpatient.

[e]BPD: borderline personality disorder.

[f]ST: schema therapy.

Table 5. Characteristics of bPT[a] study interventions: sequential use of digital and face-to-face components with complementary contents of both components (transformational blend).

Study	Project/digital component duration	Study population	Treatment setting	Psychotherapy treatment	Digital component and face-to-face component	Peer group
Duffy et al [40]	Silver Cloud iCBT[b]/undefined, depending on the waiting list	Severe affective symptoms	OP[c]/before face-to-face PT[d]	CBT/pretherapy support	• Digital: an online platform with CBT-based modules, including tools for self-monitoring, behavioral activation, cognitive restructuring, and challenging core beliefs. • Face-to-face: group therapy, face-to-face counseling, face-to-face CBT, or CBT delivered by a clinician via the internet after the online intervention.	No
Fitzpatrick et al [42]	HoPE/12 weeks of group therapy and an 8-week online program	Hoarding disorder	OP/after face-to-face PT	CBT/MT[e]/RP[f]	• Digital: an online platform with 8 CBT-based modules, including worksheets and interactive content, homework for each week, and optional therapist–assisted email support (a maximum of 2 emails per week). • Face-to-face: group therapy before the digital intervention.	No

[a]bPT: blended psychotherapy.

[b]iCBT: internet-based cognitive behavioral therapy.

[c]OP: outpatient.

[d]PT: psychotherapy.

[e]MT: maintenance therapy.

[f]RP: relapse prevention.

Concerning psychotherapeutic approaches, the majority of bPT interventions (16/23, 70%) used principles and methods derived from cognitive behavioral therapy (CBT) with various focuses. These ranged from disorder-specific CBT for depression [43,44,49,50], eating disorders [47,48,51], or psychosis [41,51,52] to competence training for social skills [53], social cognition skills [54], or voice coping [52]. Some interventions (4/23, 17%) centered on recovery [30,45,55,56] or maintenance therapy and relapse prevention [42,46-48]. Klein et al [57] and Jacob and colleagues [58] based their bPT interventions on the principles of schema therapy, and Craig and colleagues [59] used an interpersonal approach as the foundation of an AVATAR-based bPT program for the treatment of patients with SMI with psychotic symptoms [59]. Zwerenz and colleagues [43] integrated face-to-face psychodynamic inpatient treatment with a CBT-based online application (deprexis).

Among the 23 included bPT studies, a total of 20 different digital components were used. Some of the trials used the same digital component for different studies. Specifically, Jacob and colleagues [58] and Klein et al [57] used the priovi digital component, Kooistra and colleagues [49,50] used a self-developed i-CBT program, and Zwerenz et al [43,44] used the deprexis digital component.

The majority of the digital components (13/20, 65%) were based on online programs, while 5 (25%) used smartphone-based apps. One intervention incorporated an app that was compatible for both online and mobile use [51]. Two (10%) non-online–based digital components were included in the review. One study used vodcasts, which are short video clips aimed at encouraging reappraisal and acceptance of food in the blended treatment of patients with anorexia nervosa [39]. Another study combined an AVATAR program with face-to-face psychotherapy and psychosocial care to improve coping with distressing auditory verbal symptoms in psychosis [59].

A total of 7/35 (35%) digital applications provided opportunities for peer contact, either with known peers from the same treatment group [47,48,55,60] or with a broader peer community for mutual support and inspiration [30,45,46]. Therapist assistance for the digital component was implemented in 11/20 (55%) studies [41,42,45-50,55,58,61], either through individual feedback on patients' completed assignments or by providing the option for patients to contact therapists via email or chat. Six applications offered electronic platforms or message boards for contacting and communicating with program peers [45-48,55,60], sometimes facilitated by a therapist (2/20, 10%) [47,48]. Tables 3-6 present a detailed account of the bPT intervention details.

Table 6. Characteristics of bPT[a] study interventions: sequential use of digital and face-to-face components with independent contents of both components (additional blend).

Study	Project/digital component duration	Study population	Treatment setting	Psychotherapy treatment	Digital component and face-to-face component	Peer group
Alvarez-Jimenez et al [45]	Horyzons/18 months	Psychotic symptoms	OP[b]	Recovery/social functioning training	• Digital: online platform based on the Moderated Online Social Therapy Model, interactive online therapy ("pathways" and "steps"), peer-to-peer online social networking, peer moderation, and expert support. • Face-to-face: specialized care for early psychosis before the intervention and generic medical and mental health care according to individual needs.	Yes
Ebert et al [46]	TIMT/12 weeks	Affective symptoms	OP/after IP[c]	CBT[d]/MT[e]	• Digital: an online platform for peer support on the individual development plan, publishing weekly diary according to the individual development plan, and online peer support group. • Face-to-face: standard inpatient treatment during hospital stay; independent treatment after discharge (unspecified).	Yes
Fichter et al [47]	VIA/9 months	Anorexia nervosa	OP/after IP	CBT/MT/RP[f]	• Digital: an internet platform with (1) modular psychoeducational content on anorexia nervosa, (2) a diary for self-monitoring, (3) a tool for structured functional behavioral analysis (SORC scheme), (4) an electronic message board for therapist-moderated peer support, and (5) regular email contact with the therapist. Patients' platform usage is tracked to inform therapists. If the program is not used, then the therapist contacts the patient via phone or informs emergency contacts. • Face-to-face: inpatient treatment during hospital stay and independent therapy after discharge (unspecified).	Yes
Jacobi et al [48]	In-at-program/9 months	Bulimia nervosa	OP/after IP	CBT/MT/RP	• Digital: an online platform including CBT-based content on bulimia nervosa with a moderated asynchronous online patient discussion group; personal mailbox; the opportunity to write down thoughts, to protocol own behavior including potential relapse trigger; CBT-based knowledge provision on prevention and self-management; 11 sessions; first 2 months a new session forth weekly, then 1 session per week on self-monitoring, psychoeducation, and peer-peer + patient-therapist communication. • Face-to-face: standard inpatient treatment during hospital stay and independent treatment after discharge (unspecified).	Yes

[a]bPT: blended psychotherapy.

[b]OP: outpatient.

[c]IP: inpatient.

[d]CBT: cognitive behavioral therapy

[e]MT: maintenance therapy

[f]RP: relapse prevention.

Feasibility and Adherence

We defined feasibility as whether the bPT study intervention was suitable for the given setting and the respective study population. Adherence addresses participants' engagement and compliance with the bPT treatment [62,63]. We found that 22/23 (96%) studies concluded that their bPT interventions were feasible. One study reported nonfeasibility due to participants' lack of computer skills, trainers' hesitation toward the program, and the program's inflexibility [30]. Feasibility aspects focused on acceptability, patient and therapist satisfaction, and overall ratings, which were reported to be good to excellent (see Multimedia Appendix 3). Safety considerations were addressed in 7 studies [41,45,48,54,56,57,60], with 1 study [58] reporting adverse emotions triggered by self-use of the digital application.

Regarding treatment adherence, all but 1 study (22/23, 96%) reported on it, mainly focusing on patients' adherence to the digital component. Various indicators were used, such as completed tasks, frequency and duration of usage, log-ins, self-reported usage, and participation in live chats. Four studies analyzed the impact of treatment adherence on intervention effects [47,48,58,60]. Two studies found no correlation [48,58], while 1 reported a positive association between app engagement and symptom reduction [60], and another identified a significant effect of treatment adherence on participants' body weight [47].

Adherence to the face-to-face component was examined in 1 study [48], which found that patients without regular face-to-face treatment were more likely to not complete the online program. The overall treatment adherence rates varied between 0% and 37%, and in controlled trials, lower dropout rates were generally observed in the intervention treatment compared with control groups. In controlled trials, 12/15 (80%) reported lower dropout rates in the intervention treatment compared with control groups, although differences overall appeared small and heterogeneous, ranging between 0.5% [46] and 15% [55].

Efficacy and Effectiveness of bPT for SMI

To provide a concise overview of the treatment effects of bPT interventions for SMI, the main outcomes of the controlled studies are summarized in Table 7.

Among the 15/23 (65%) studies that used control group designs to examine the efficacy or effectiveness of bPT interventions, 4 (17%) reported significant effects in favor of the intervention [43,44,46,59], with 2 of them using the same study population [43,44]. Three of these studies focused on bPT effects in patients with affective disorders [43,44,46], while Craig et al [59] investigated the treatment of distressing auditory verbal hallucinations as part of psychosis. Two trials compared bPT with TAU groups without a digital component [46,59], and Zwerenz et al [43,44] compared the intervention with an active control group with face-to-face treatment and different digital interventions.

Among the trials without significant main intervention effects, 5 reported significant effects on secondary outcomes [45,47,48,51,52], indicating potential efficacy. Klein et al [57], in a study involving patients with borderline personality disorder, found a significant intervention effect in a per-protocol analysis, specifically for participants from the intervention group who regularly used the digital component (priovi). Additionally, Cardi et al [39], in their quasi-experimental bPT study on the positive impacts of listening to vodcasts on food appraisal in patients with anorexia nervosa, found an interaction effect with the context: outpatients benefited more from the vodcast intervention, while inpatients reported more benefits from the control condition, which involved listening to classical music.

As many as 3/15 (20%) of the controlled studies [42,49,55] concluded that there were no group differences between intervention and control groups. Kooistra et al [49] in their investigation of the efficacy and cost-effectiveness of a manualized bPT intervention for patients with severe depression found no cost-effectiveness from a societal perspective but an acceptable probability of being cost-effective from the health care provider perspective.

Beentjes et al [61] reported significant effects in favor of the intervention for self-management, recovery, and general health perception. However, due to low treatment and study adherence, the authors deemed the results inconclusive.

Table 7. Main outcomes of controlled studies in the sample.

Study	Study design/mental health symptoms	Sample size, n	Control group	Efficacy/effectiveness
Ebert et al [46]	RCT[a]/affective symptoms	400	TAU[b] included standard outpatient treatment without any digital component	The IG[c] group was superior to the CG[d]; lower differences were observed in the change of general psychopathological symptom severity from discharge to 3- and 12-month follow-up in the IG. The IG exhibited less frequent symptom deteriorations and a higher frequency of remission at the follow-up assessments.
Zwerenz et al[e] [44]	RCT/depression	229	The active CG received inpatient PT[f] and psychoeducational information	The IG was superior to the CG; the IG demonstrated lower levels of depression and anxiety, along with higher quality of life and self-esteem, when compared with the CG.
Zwerenz et al[e] [43]	RCT (follow-up)/depression	215	The active CG received outpatient PT and psychoeducational information	The IG was superior to the CG; the IG demonstrated lower levels of depression and anxiety, along with higher quality of life and self-esteem, when compared with the CG.
Craig et al [59]	RCT/psychotic disorder	150	TAU included supportive counseling without any digital component	The IG was superior to the CG; there was a reduction in the severity of auditory verbal hallucinations, but no effect on the malevolence of voices. At the 24-week follow-up, no significant differences were observed between the IG and the CG.
Alvarez-Jimenez et al [45]	RCT/psychotic symptoms	170	TAU included standard care and community mental health services	There were no group differences observed in the primary outcome of social functioning. However, in the IG, there were 5.5 times greater odds of finding employment or participating in an education program. By contrast, in the CG, there was a 2-fold increase in hospital admissions due to psychosis.
Bell et al [52]	RCT (pilot)/psychotic disorder	34	TAU included standard psychiatric care	There were no significant effects observed on the primary outcome of psychotic symptoms measured by PSYRATS[g]. However, the IG did show a significant effect, particularly in the Visual Analog Scale items related to coping with voices and awareness of patterns in voices.
Garety et al [51]	RCT/psychotic disorder	361	TAU included standard outpatient psychiatric care	There was no significant effect on the primary outcome of self-reported paranoia at 24 weeks compared with TAU. However, secondary beneficial effects on this measure were observed at 12 weeks. Both self-reported persecution and observer-rated paranoia showed improvement at both assessment points.
Fichter et al [47]	RCT/anorexia nervosa	258	TAU included standard outpatient treatment without any digital component	No group difference for the primary outcome general psychopathology; patients in the IG reported more weight gain than those in the CG.
Jacobi et al [48]	RCT/bulimia nervosa	253	TAU included standard outpatient treatment without any digital component	No group difference in abstinence rates; patients in the IG reported a lower frequency of vomiting than those in the CG.
Klein et al [57]	RCT/BPD[h]	204	TAU included standard outpatient treatment without any digital component	There was no significant effect in favor of the intervention in the overall analysis. The primary outcome, which measured the change in BPD symptoms (BPDSI[i]), decreased similarly in both groups. However, in the prespecified per-protocol analysis, which included only participants from the intervention group who used the intervention for at least 3 hours, a statistically significant intervention effect was found.
Cardi et al [39]	Quasi-experimental design/anorexia nervosa	38	The active CG received 4 pieces of classical music (20 minutes)	There was an interaction effect with the context. Specifically, outpatients derived more benefits from the vodcast, whereas inpatients experienced greater benefits from music. Interestingly, the vodcast did not lead to a reduction in distress and vigilance to food in the inpatient setting. Overall, there were no significant group differences between vodcast and music in the primary outcomes.
Fitzpatrick et al [42]	Pilot study; 2-group design/hoarding disorder	16	Waiting list	There were no group differences in the primary outcomes related to psychopathology. However, in the IG, there was a trend toward continued improvement in overall hoarding scores.
Blankers et al [55]	Pilot study; 2-group design/various disorders and SMI[j]	47	TAU included FACT[k] without any digital component	There were no group differences observed, with both the IG and the care-as-usual group showing comparable improvements in quality of life and self-efficacy beliefs regarding their mental health problems. Additionally, HONOS[l] scores did not change over time.

Study	Study design/mental health symptoms	Sample size, n	Control group	Efficacy/effectiveness
Kooistra et al [49]	RCT/depression	102	TAU included standard CBT[m] without any digital component	There were no group differences observed for depressive episodes, response to treatment, and quality-adjusted life years. The intervention did not demonstrate cost-effectiveness from a societal perspective. However, there was an acceptable probability of being cost-effective from the health care provider's perspective.
Beentjes et al [61]	RCT/SMI and various disorders	60	TAU included IMR[n] without any digital component	There were significant effects in favor of the intervention for self-management, recovery, and general health perception. However, the results on effectiveness are inconclusive due to confounding factors and interaction modifications attributed to low study adherence.

[a]RCT: randomized controlled trial.

[b]TAU: treatment as usual

[c]IG: intervention group.

[d]CG: control group.

[e]bPT interventions with the same study population but different data.

[f]PT: psychotherapy.

[g]PSYRATS: Psychotic Symptom Rating Scales.

[h]BPD: borderline personality disorder.

[i]BPDSI: Borderline Personality Disorder Severity Index.

[j]SMI: severe mental illness.

[k]FACT: flexible assertive community treatment.

[l]HONOS: Health of the Nation Outcome Scale.

[m]CBT: cognitive behavioral therapy.

[n]IMR: Illness Management Recovery Program.

Discussion

Principal Findings

In this scoping review, we provided a comprehensive summary of the current evidence on bPT for SMI, with a specific emphasis on feasibility, adherence, and effectiveness. Our review encompassed 25 research papers, consolidating data from 23 bPT interventions that involved a total of 2554 patients with SMI. Our findings offer preliminary evidence supporting the feasibility of bPT treatment for patients with SMI. In examining adherence, our analysis identified it as a crucial factor in exploring and advancing bPT treatment for SMI. However, it is important to note that empirical research in this area is currently limited. The potential positive impact of bPT on adherence in SMI treatment is indicated by the attrition rates reported in the summarized studies. These rates ranged from 0% to 37%, demonstrating favorable outcomes when compared with both guided and unguided technology-based interventions (TBIs). For instance, Musiat et al [64] reported average completion rates of 64% for guided TBIs and 52% for unguided TBIs. When considering attrition rates for mobile health (mHealth) interventions, with a recent meta-analysis reporting an average of 24% [65], the attrition rates found in our studies appear comparable. Nevertheless, the accumulated evidence on the effects of bPT for SMI remains sparse and challenging to extrapolate. Among the 15 studies with a control group design, 4 concluded that bPT interventions were effective compared with control groups. However, 2 of these studies used the same study population, with significantly varied control groups. Additionally, 3 studies investigated patients with affective disorders.

Our feasibility results are in line with earlier systematic evidence on the feasibility of eHealth applications for SMI [66,67], as well as with 2 recent reviews on mHealth for patients with SMI, with a focus on monitoring [18] and disorder management applications [68]. These reviews also established the general acceptance and feasibility of digital interventions in patients with SMI. Our efficacy conclusion is more reflective of the quality of the reviewed research than of the actual subject matter. It primarily focuses on 2 noticeable aspects: First, bPT interventions under scrutiny in the analyzed studies varied greatly in their therapeutic approaches, objectives, digital and face-to-face component features, target groups, and treatment settings, making it challenging to generalize across studies. Second, the studies were heterogeneous regarding research designs, control types, and outcome measures. While these aspects are an outcome of the review's overall broad scope, they also mirror the current heterogeneity of this field of research.

Despite these conclusions, we remain optimistic about the potential of bPT for the treatment of SMI based on several considerations. On the one side, substantial and increasing evidence supports the effectiveness of face-to-face psychotherapy and psychosocial care for SMI [69-72], although the effects are low to moderate [73,74], and access is often limited due to systemic [75,76] and individual factors [77]. On the other side, the utilization of digital technology and applications is becoming increasingly common among individuals with SMI, and there is ongoing development of digital health interventions [78], including recent advancements in artificial intelligence for SMI treatment [18,79]. However, research also indicates a negative correlation between adherence

and the effectiveness of stand-alone TBIs and disorder severity [23,65,80], indicating a fundamental need for human facilitation of digital interventions for SMI [66]. Therefore, bPT, with its concept of enhancing empirically supported psychological treatment with digital content, appears to hold promise for effective, potentially cost-effective, and more accessible treatment for this complex and severely affected patient group. We conclude our review by emphasizing the need for further research in this field.

A Future Research Agenda on bPT for SMI

We propose several considerations for further research and clinical directions. First, the design of bPT interventions should take into account various contexts of bPT for SMI and consider the advantages and opportunities of different bPT concepts (Multimedia Appendix 4). Using participatory research approaches, as demonstrated by Austin et al [41], can help ensure the creation of appropriate and sustainable treatment solutions based on patients' needs. Second, it would be advantageous for bPT research to investigate the effects of bPT, giving equal attention to both digital and face-to-face components. This should include the exploration of how these components interact with each other and with additional contexts, such as treatment settings, therapists, and mental health care systems. This would necessitate careful deliberation of suitable control groups. Third, adherence is a critical factor in the effectiveness of TBIs in general [23,65,80], and particularly in the treatment of SMI [12,31,66,81]. Therefore, more research on this topic, including the development of analytical concepts, is required. Fourth, given the multidisciplinary and often long-term nature of SMI treatment, implementation research on bPT for SMI could foster the development of protocols for integrating face-to-face and digital treatment components across varied treatment settings. Lastly, considering the chronic nature of SMI and the increased potential for relapse or decompensation, research on bPT for SMI should encompass the investigation of short- and long-term treatment effects, as well as aspects of treatment safety.

Limitations

In line with our study design, our literature review was broad in scope, encompassing a wide range of inclusion criteria in terms of study populations, bPT treatment approaches, and study designs. Given the nascent state of the field, this broad approach seemed fitting and allowed for a preliminary overview of the literature. However, the resulting heterogeneity in the findings may reflect this broad approach. Future systematic investigations should consider a more narrow focus to validate our conclusions and build upon the systematic evidence concerning subgroups and categories of bPT treatments for patients with SMI. Additionally, it would be beneficial to include stakeholder and patient perspectives directly in future research endeavors.

Similarly, the inclusion of both controlled and uncontrolled studies may limit the findings of this review. Moreover, due to the absence of quality assessment standards for digital mental health interventions [82], we did not exclude studies based on the potentially inadequate quality of the digital applications utilized. In our review, we only identified 5 studies using mHealth interventions for bPT, which was surprising given the significant advancements and research dynamic in this field. This might suggest that our findings may not generalize to bPT with mHealth.

Conclusions

The treatment of SMI, which is both prevalent and burdensome for individuals and society, remains challenging [69,75]. In recent decades, psychotherapeutic interventions have emerged as evidence-based first-line treatments, even for patients within this group, yielding mild to moderate effects. Concurrently, the past 25 years have witnessed the development of a broad array of digital mental health interventions. These digital mental health interventions are designed to address treatment gaps, offering affordable, accessible, and scalable alternatives to traditional mental health treatments, and contributing to the overall improvement of mental health care [20,23,80]. In this review, we have summarized the literature on the feasibility, adherence, and effectiveness or efficacy of bPT for patients with SMI. Our findings confirm the feasibility of such treatments for SMI and underline the need for further research to establish the effectiveness of this treatment delivery mode.

Considering that SMI treatment often entails a long history of patient suffering and cost-intensive multidisciplinary care [29,83,84], the further development and implementation of bPT for SMI treatment presents a significant opportunity. This opportunity not only has the potential to further improve the effectiveness and cost-effectiveness of mental health care but is also particularly relevant for this population.

Acknowledgments

We thank Lilian Fraefel (Department of Psychology, University of Zurich, Zurich, Switzerland) for conducting an additional literature search as part of the revision process.

Authors' Contributions

BK led the development and implementation of the protocol. MR, YE-S, and BK designed the search strategy. MR and YE-S retrieved articles, designed the data extraction process, screened the search results, and led on writing the manuscript. JV and MM contributed to the development of the research design and took part in writing and revising the manuscript. ES took part in the final revision of the manuscript. All authors read and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 108 KB - mental_v10i1e43882_app1.docx]

Multimedia Appendix 2
Quality assessment.
[XLSX File (Microsoft Excel File), 12 KB - mental_v10i1e43882_app2.xlsx]

Multimedia Appendix 3
Data extraction form.
[XLSX File (Microsoft Excel File), 61 KB - mental_v10i1e43882_app3.xlsx]

Multimedia Appendix 4
Mapping the field of research on bPT for SMI. bPT: blended psychotherapy; SMI: severe mental illness.
[PNG File , 581 KB - mental_v10i1e43882_app4.png]

References

1. Kessler RC, Angermeyer M, Anthony JC, DE Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
2. Steel Z, Marnane C, Iranpour C, Chey T, Jackson J, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
3. Richter D, Wall A, Bruen A, Whittington R. Is the global prevalence rate of adult mental illness increasing? Systematic review and meta-analysis. Acta Psychiatr Scand 2019 Nov 03;140(5):393-407. [doi: 10.1111/acps.13083] [Medline: 31393996]
4. Santomauro DF, Mantilla Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. The Lancet 2021 Nov;398(10312):1700-1712. [doi: 10.1016/s0140-6736(21)02143-7]
5. Bergner RM, Bunford N. Mental disorder is a disability concept, not a behavioral one. Philosophy, Psychiatry, & Psychology 2017;24(1):25-40. [doi: 10.1353/ppp.2017.0004]
6. Wakefield J. The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis. World Psychiatry 2007 Oct;6(3):149-156 [FREE Full text] [Medline: 18188432]
7. Souza A, Guimarães RA, de Araújo Vilela D, de Assis RM, de Almeida Cavalcante Oliveira LM, Souza M, et al. Factors associated with the burden of family caregivers of patients with mental disorders: a cross-sectional study. BMC Psychiatry 2017 Oct 25;17(1):353 [FREE Full text] [doi: 10.1186/s12888-017-1501-1] [Medline: 29070012]
8. Doran CM, Kinchin I. A review of the economic impact of mental illness. Aust Health Review 2019;43(1):43. [doi: 10.1071/ah16115]
9. Lungu A, Jun JJ, Azarmanesh O, Leykin Y, Chen CEJ. Blended care-cognitive behavioral therapy for depression and anxiety in real-world settings: pragmatic retrospective study. J Med Internet Res 2020 Jul 06;22(7):e18723 [FREE Full text] [doi: 10.2196/18723] [Medline: 32628120]
10. Bielinski L, Trimpop L, Berger T. [All in the mix? Blended psychotherapy as an example of digitalization in psychotherapy]. Psychotherapeut (Berl) 2021;66(5):447-454 [FREE Full text] [doi: 10.1007/s00278-021-00524-3] [Medline: 34257478]
11. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
12. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
13. Rasing S. Blended Treatment for Depressive Disorders in Youth: a Narrative Review. J Cogn Ther 2020 Sep 25;14(1):47-58. [doi: 10.1007/s41811-020-00088-1]
14. Sander J, Bolinski F, Diekmann S, Gaebel W, Günther K, Hauth I, et al. Online therapy: an added value for inpatient routine care? Perspectives from mental health care professionals. Eur Arch Psychiatry Clin Neurosci 2022 Feb 16;272(1):107-118 [FREE Full text] [doi: 10.1007/s00406-021-01251-1] [Medline: 33725165]
15. Lindhiem O, Bennett C, Rosen D, Silk J. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. Behav Modif 2015 Nov;39(6):785-804 [FREE Full text] [doi: 10.1177/0145445515595198] [Medline: 26187164]

16. Schuster R, Fichtenbauer I, Sparr V, Berger T, Laireiter AR. Feasibility of a blended group treatment (bGT) for major depression: uncontrolled interventional study in a university setting. BMJ Open 2018 Mar 12;8(3):e018412 [FREE Full text] [doi: 10.1136/bmjopen-2017-018412] [Medline: 29530905]
17. Hill R, Ayling J. Towards blended psychotherapy: a COVID-19 legacy. In: Psychotherapy and COVID-19: reflections on the pandemic. London, UK: UK Council for Psychotherapy; Mar 2021:6-9.
18. Jameel L, Valmaggia L, Barnes G, Cella M. mHealth technology to assess, monitor and treat daily functioning difficulties in people with severe mental illness: A systematic review. J Psychiatr Res 2021 Nov 24;145:35-49. [doi: 10.1016/j.jpsychires.2021.11.033] [Medline: 34856524]
19. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone Apps for the Treatment of Mental Disorders: Systematic Review. JMIR Mhealth Uhealth 2020 Apr 02;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
20. Torous J, Bucci S, Bell I, Kessing L, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
21. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
22. Dülsen P, Bendig E, Küchler AM, Christensen H, Baumeister H. Digital interventions in adult mental healthcare settings: recent evidence and future directions. Curr Opin Psychiatry 2020 Jul;33(4):422-431. [doi: 10.1097/YCO.0000000000000614] [Medline: 32427592]
23. Taylor C, Graham A, Flatt R, Waldherr K, Fitzsimmons-Craft E. Current state of scientific evidence on Internet-based interventions for the treatment of depression, anxiety, eating disorders and substance abuse: an overview of systematic reviews and meta-analyses. Eur J Public Health 2021 Jul 07;31(31 Suppl 1):i3-i10 [FREE Full text] [doi: 10.1093/eurpub/ckz208] [Medline: 32918448]
24. Paterson C, Karatzias T, Dickson A, Harper S, Dougall N, Hutton P. Psychological therapy for inpatients receiving acute mental health care: A systematic review and meta-analysis of controlled trials. Br J Clin Psychol 2018 Nov 16;57(4):453-472. [doi: 10.1111/bjc.12182] [Medline: 29660770]
25. Gaynes B, Brown C, Lux LJ, Sheitman B, Ashok M, Boland E, et al. Relationship Between Use of Quality Measures and Improved Outcomes in Serious Mental Illness (Report No.: 15-EHC003-EF). Rockville, MD: Agency for Healthcare Research and Quality; Jan 2015.
26. Chiang K, Tsai J, Liu D, Lin CH, Chiu HL, Chou KR. Efficacy of cognitive-behavioral therapy in patients with bipolar disorder: A meta-analysis of randomized controlled trials. PLoS One 2017;12(5):e0176849 [FREE Full text] [doi: 10.1371/journal.pone.0176849] [Medline: 28472082]
27. Pfammatter M, Junghan UM, Brenner HD. Efficacy of psychological therapy in schizophrenia: conclusions from meta-analyses. Schizophr Bull 2006 Oct 18;32 Suppl 1(Suppl 1):S64-S80 [FREE Full text] [doi: 10.1093/schbul/sbl030] [Medline: 16905634]
28. Thase ME, Kingdon D, Turkington D. The promise of cognitive behavior therapy for treatment of severe mental disorders: a review of recent developments. World Psychiatry 2014 Oct 01;13(3):244-250 [FREE Full text] [doi: 10.1002/wps.20149] [Medline: 25273290]
29. Delespaul PH, consensusgroep EPA D. [Consensus regarding the definition of persons with severe mental illness and the number of such persons in the Netherlands]. Tijdschr Psychiatr 2013;55(6):427-438 [FREE Full text] [Medline: 23864410]
30. Beentjes T, van Gaal BGI, Vermeulen H, Nijhuis-van der Sanden MWG, Goossens P. A Blended Electronic Illness Management and Recovery Program for People With Severe Mental Illness: Qualitative Process Evaluation Alongside a Randomized Controlled Trial. JMIR Ment Health 2021 Jan 20;8(1):e20860 [FREE Full text] [doi: 10.2196/20860] [Medline: 33470945]
31. McCabe R, Priebe S. The therapeutic relationship in the treatment of severe mental illness: a review of methods and findings. Int J Soc Psychiatry 2004 Jun;50(2):115-128. [doi: 10.1177/0020764004040959] [Medline: 15293429]
32. Ruggeri M, Leese M, Thornicroft G, Bisoffi G, Tansella M. Definition and prevalence of severe and persistent mental illness. Br J Psychiatry 2000 Aug 02;177(2):149-155. [doi: 10.1192/bjp.177.2.149] [Medline: 11026955]
33. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
34. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
35. Colquhoun H, Levac D, O'Brien KK, Straus S, Tricco A, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014 Dec;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
36. Tricco A, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]

37. Santos WMD, Secoli SR, Püschel VADA. The Joanna Briggs Institute approach for systematic reviews. Rev Lat Am Enfermagem 2018 Nov 14;26:e3074 [FREE Full text] [doi: 10.1590/1518-8345.2885.3074] [Medline: 30462787]
38. Ellis KR, Raji D, Olaniran M, Alick C, Nichols D, Allicock M. A systematic scoping review of post-treatment lifestyle interventions for adult cancer survivors and family members. J Cancer Surviv 2022 Apr 13;16(2):233-256 [FREE Full text] [doi: 10.1007/s11764-021-01013-x] [Medline: 33713302]
39. Cardi V, Lounes N, Kan C, Treasure J. Meal support using mobile technology in Anorexia Nervosa. Contextual differences between inpatient and outpatient settings. Appetite 2013 Jan;60(1):33-39. [doi: 10.1016/j.appet.2012.10.004] [Medline: 23085476]
40. Duffy D, Enrique A, Connell S, Connolly C, Richards D. Internet-Delivered Cognitive Behavior Therapy as a Prequel to Face-To-Face Therapy for Depression and Anxiety: A Naturalistic Observation. Front Psychiatry 2019;10:902 [FREE Full text] [doi: 10.3389/fpsyt.2019.00902] [Medline: 31998149]
41. Austin SF, Frøsig A, Buus N, Lincoln T, von Malachowski A, Schlier B, et al. Service User Experiences of Integrating a Mobile Solution (IMPACHS) Into Clinical Treatment for Psychosis. Qual Health Res 2021 Apr 23;31(5):942-954. [doi: 10.1177/1049732320986556] [Medline: 33491577]
42. Fitzpatrick M, Nedeljkovic M, Abbott J, Kyrios M, Moulding R. "Blended" therapy: The development and pilot evaluation of an internet-facilitated cognitive behavioral intervention to supplement face-to-face therapy for hoarding disorder. Internet Interv 2018 Jun;12:16-25 [FREE Full text] [doi: 10.1016/j.invent.2018.02.006] [Medline: 30135765]
43. Zwerenz R, Baumgarten C, Becker J, Tibubos A, Siepmann M, Knickenberg R, et al. Improving the Course of Depressive Symptoms After Inpatient Psychotherapy Using Adjunct Web-Based Self-Help: Follow-Up Results of a Randomized Controlled Trial. J Med Internet Res 2019 Oct 24;21(10):e13655 [FREE Full text] [doi: 10.2196/13655] [Medline: 31651403]
44. Zwerenz R, Becker J, Knickenberg RJ, Siepmann M, Hagen K, Beutel ME. Online Self-Help as an Add-On to Inpatient Psychotherapy: Efficacy of a New Blended Treatment Approach. Psychother Psychosom 2017 Nov 3;86(6):341-350. [doi: 10.1159/000481177] [Medline: 29131090]
45. Alvarez-Jimenez M, Koval P, Schmaal L, Bendall S, O'Sullivan S, Cagliarini D, et al. The Horyzons project: a randomized controlled trial of a novel online social therapy to maintain treatment effects from specialist first-episode psychosis services. World Psychiatry 2021 Jun;20(2):233-243 [FREE Full text] [doi: 10.1002/wps.20858] [Medline: 34002511]
46. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
47. Fichter MM, Quadflieg N, Nisslmüller K, Lindner S, Osen B, Huber T, et al. Does internet-based prevention reduce the risk of relapse for anorexia nervosa? Behav Res Ther 2012 Mar;50(3):180-190. [doi: 10.1016/j.brat.2011.12.003] [Medline: 22317754]
48. Jacobi C, Beintner I, Fittig E, Trockel M, Braks K, Schade-Brittinger C, et al. Web-Based Aftercare for Women With Bulimia Nervosa Following Inpatient Treatment: Randomized Controlled Efficacy Trial. J Med Internet Res 2017 Sep 22;19(9):e321 [FREE Full text] [doi: 10.2196/jmir.7668] [Medline: 28939544]
49. Kooistra L, Wiersma J, Ruwaard J, Neijenhuijs K, Lokkerbol J, van Oppen P, et al. Cost and Effectiveness of Blended Versus Standard Cognitive Behavioral Therapy for Outpatients With Depression in Routine Specialized Mental Health Care: Pilot Randomized Controlled Trial. J Med Internet Res 2019 Oct 29;21(10):e14261 [FREE Full text] [doi: 10.2196/14261] [Medline: 31663855]
50. Kooistra LC, Ruwaard J, Wiersma JE, van Oppen P, van der Vaart R, van Gemert-Pijnen JE, et al. Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. Internet Interv 2016 May;4:61-71 [FREE Full text] [doi: 10.1016/j.invent.2016.01.003] [Medline: 30135791]
51. Garety P, Ward T, Emsley R, Greenwood K, Freeman D, Fowler D, et al. Effects of SlowMo, a Blended Digital Therapy Targeting Reasoning, on Paranoia Among People With Psychosis: A Randomized Clinical Trial. JAMA Psychiatry 2021 Jul 01;78(7):714-725 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.0326] [Medline: 33825827]
52. Bell I, Rossell S, Farhall J, Hayward M, Lim M, Fielding-Smith S, et al. Pilot randomised controlled trial of a brief coping-focused intervention for hearing voices blended with smartphone-based ecological momentary assessment and intervention (SAVVy): Feasibility, acceptability and preliminary clinical outcomes. Schizophr Res 2020 Feb;216:479-487. [doi: 10.1016/j.schres.2019.10.026] [Medline: 31812327]
53. Bendig E, Küchler AM, Baumeister H, Becker T. Blended Care in In-Patient Acute Psychiatric Care. The Example of a Group Training for Social Competences in Adults-A Pretest-Posttest Feasibility Study. Int J Environ Res Public Health 2021 Sep 07;18(18):9433 [FREE Full text] [doi: 10.3390/ijerph18189433] [Medline: 34574357]
54. Sedgwick O, Hardy A, Greer B, Newbery K, Cella M. "I wanted to do more of the homework!"-Feasibility and acceptability of blending app-based homework with group therapy for social cognition in psychosis. J Clin Psychol 2021 Dec 08;77(12):2701-2724. [doi: 10.1002/jclp.23193] [Medline: 34101177]
55. Blankers M, van Emmerik A, Richters B, Dekker J. Blended internet care for patients with severe mental illnesses: An open label prospective controlled cohort pilot study. Internet Interv 2016 Sep;5:51-55 [FREE Full text] [doi: 10.1016/j.invent.2016.07.004] [Medline: 30135807]

56. Thomas N, Farhall J, Foley F, Leitan N, Villagonzalo K, Ladd E, et al. Promoting Personal Recovery in People with Persisting Psychotic Disorders: Development and Pilot Study of a Novel Digital Intervention. Front Psychiatry 2016;7:196 [FREE Full text] [doi: 10.3389/fpsyt.2016.00196] [Medline: 28066271]
57. Klein J, Hauer-von Mauschwitz A, Berger T, Fassbinder E, Mayer J, Borgwardt S, et al. Effectiveness and safety of the adjunctive use of an internet-based self-management intervention for borderline personality disorder in addition to care as usual: results from a randomised controlled trial. BMJ Open 2021 Sep 08;11(9):e047771 [FREE Full text] [doi: 10.1136/bmjopen-2020-047771] [Medline: 34497078]
58. Jacob G, Hauer A, Köhne S, Assmann N, Schaich A, Schweiger U, et al. A Schema Therapy-Based eHealth Program for Patients with Borderline Personality Disorder (priovi): Naturalistic Single-Arm Observational Study. JMIR Ment Health 2018 Dec 17;5(4):e10983 [FREE Full text] [doi: 10.2196/10983] [Medline: 30559092]
59. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry 2018 Jan;5(1):31-40 [FREE Full text] [doi: 10.1016/S2215-0366(17)30427-3] [Medline: 29175276]
60. Granholm E, Holden J, Dwyer K, Mikhael T, Link P, Depp C. Mobile-Assisted Cognitive Behavioral Therapy for Negative Symptoms: Open Single-Arm Trial With Schizophrenia Patients. JMIR Ment Health 2020 Dec 01;7(12):e24406 [FREE Full text] [doi: 10.2196/24406] [Medline: 33258792]
61. Beentjes TAA, Goossens PJJ, Vermeulen H, Teerenstra S, Nijhuis-van der Sanden MWG, van Gaal BGI. BMC Health Serv Res 2018 Dec 12;18(1):962 [FREE Full text] [doi: 10.1186/s12913-018-3767-5] [Medline: 30541536]
62. Beintner I, Jacobi C. Internet-based aftercare for women with bulimia nervosa following inpatient treatment: The role of adherence. In: Internet Interv. Internet-based aftercare for women with bulimia nervosa following inpatient treatment: The role of adherence. Internet Interv Elsevier; Mar 2019:67-75.
63. Haynes RB, Taylor DW, Sackett DL, editors. Adherence in Health Care. Baltimore, MD: Johns Hopkins University Press; 1979.
64. Musiat P, Johnson C, Atkinson M, Wilksch S, Wade T. Impact of guidance on intervention adherence in computerised interventions for mental health problems: a meta-analysis. Psychol Med 2021 Nov 22;52(2):229-240. [doi: 10.1017/s0033291721004621]
65. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
66. Aref-Adib G, McCloud T, Ross J, O'Hanlon P, Appleton V, Rowe S, et al. Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. The Lancet Psychiatry 2019 Mar;6(3):257-266. [doi: 10.1016/s2215-0366(18)30302-x]
67. Naslund J, Marsch L, McHugo G, Bartels S. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
68. Chivilgina O, Wangmo T, Elger B, Heinrich T, Jotterand F. mHealth for schizophrenia spectrum disorders management: A systematic review. Int J Soc Psychiatry 2020 Nov;66(7):642-665. [doi: 10.1177/0020764020933287] [Medline: 32571123]
69. Gühne U, Weinmann S, Riedel-Heller S, Becker T. Psychosocial therapies in severe mental illness: update on evidence and recommendations. Curr Opin Psychiatry 2020 Jul;33(4):414-421. [doi: 10.1097/yco.0000000000000618]
70. Bighelli I, Salanti G, Huhn M, Schneider-Thoma J, Krause M, Reitmeir C, et al. Psychological interventions to reduce positive symptoms in schizophrenia: systematic review and network meta-analysis. World Psychiatry 2018 Oct;17(3):316-329 [FREE Full text] [doi: 10.1002/wps.20577] [Medline: 30192101]
71. Grubaugh AL, Brown WJ, Wojtalik JA, Myers US, Eack SM. Meta-Analysis of the Treatment of Posttraumatic Stress Disorder in Adults With Comorbid Severe Mental Illness. J Clin Psychiatry 2021 May 25;82(3):33795. [doi: 10.4088/jcp.20r13584]
72. Xuan R, Li X, Qiao Y, Guo Q, Liu X, Deng W, et al. Mindfulness-based cognitive therapy for bipolar disorder: A systematic review and meta-analysis. Psychiatry Res 2020 Aug;290:113116. [doi: 10.1016/j.psychres.2020.113116] [Medline: 32480120]
73. Jauhar S, Laws K, McKenna P. CBT for schizophrenia: a critical viewpoint. Psychol. Med 2019 Feb 13;49(8):1233-1236. [doi: 10.1017/s0033291718004166]
74. Zimmerman M. Severity and the Treatment of Depression. J Nerv Ment Dis 2019;207(4):219-223. [doi: 10.1097/nmd.0000000000000960]
75. O'Connor LK, Yanos PT. Where are all the psychologists? A review of factors impacting the underrepresentation of psychology in work with serious mental illness. Clin Psychol Rev 2021 Jun;86:102026. [doi: 10.1016/j.cpr.2021.102026] [Medline: 33813162]
76. Ince P, Haddock G, Tai S. A systematic review of the implementation of recommended psychological interventions for schizophrenia: Rates, barriers, and improvement strategies. Psychol Psychother 2016 Sep;89(3):324-350. [doi: 10.1111/papt.12084] [Medline: 26537838]
77. Sickel AE, Seacat JD, Nabors NA. Mental health stigma update: A review of consequences. Advances in Mental Health 2015 Jan 30;12(3):202-215. [doi: 10.1080/18374905.2014.11081898]

78. Batra S, Baker R, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. Med Devices (Auckl) 2017 Oct;10:237-251 [FREE Full text] [doi: 10.2147/MDER.S144158] [Medline: 29042823]
79. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim H, et al. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. Curr Psychiatry Rep 2019 Nov 07;21(11):116 [FREE Full text] [doi: 10.1007/s11920-019-1094-0] [Medline: 31701320]
80. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022 Jan 26;1(2):87-100. [doi: 10.1038/s44159-021-00003-1]
81. Velligan D, Weiden P, Sajatovic M, Scott J, Carpenter D, Ross R, et al. Assessment of adherence problems in patients with serious and persistent mental illness: recommendations from the Expert Consensus Guidelines. J Psychiatr Pract 2010 Jan;16(1):34-45. [doi: 10.1097/01.pra.0000367776.96012.ca] [Medline: 20098229]
82. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 02;2(1):118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
83. Nicaise P, Giacco D, Soltmann B, Pfennig A, Miglietta E, Lasalvia A, et al. Healthcare system performance in continuity of care for patients with severe mental illness: A comparison of five European countries. Health Policy 2020 Jan;124(1):25-36. [doi: 10.1016/j.healthpol.2019.11.004] [Medline: 31831211]
84. Wiersma D. Needs of people with severe mental illness. Acta Psychiatr Scand Suppl 2006(429):115-119. [doi: 10.1111/j.1600-0447.2005.00728.x] [Medline: 16445493]

Abbreviations

APA: American Psychological Association
BDI-II: Beck's Depression Inventory Second Edition
bPT: blended psychotherapy
CBT: cognitive behavioral therapy
GAD: Generalized Anxiety Disorder (Scale)
mHealth: mobile health
NICE: National Institute for Health and Care Excellence
PHQ-9: 9-item Patient Health Questionnaire
PRISMA-ScR: PRISMA Extension for Scoping Reviews
SMI: severe mental illness
TAU: treatment as usual
TBI: technology-based intervention

Edited by J Torous; submitted 29.10.22; peer-reviewed by R Büscher, M Poppelaars; comments to author 10.02.23; revised version received 06.07.23; accepted 29.10.23; published 26.12.23.

Please cite as:
Ehrt-Schäfer Y, Rusmir M, Vetter J, Seifritz E, Müller M, Kleim B
Feasibility, Adherence, and Effectiveness of Blended Psychotherapy for Severe Mental Illnesses: Scoping Review
JMIR Ment Health 2023;10:e43882
URL: https://mental.jmir.org/2023/1/e43882
doi:10.2196/43882
PMID:38147373

Original Paper

Virtual Partnership Addressing Mental Health Crises: Mixed Methods Study of a Coresponder Program in Rural Law Enforcement

M Muska Nataliansyah[1], MPH, MD, PhD; Kimberly A S Merchant[2], MA; J Priyanka Vakkalanka[3], PhD; Luke Mack[4,5], MA; Seth Parsons[4,6], MD; Marcia M Ward[2], PhD

[1]Department of Surgery, Collaborative for Healthcare Delivery Science, Medical College of Wisconsin, Milwaukee, WI, United States
[2]Department of Health Management and Policy, College of Public Health, University of Iowa, Iowa City, IA, United States
[3]Department of Emergency Medicine, University of Iowa Carver College of Medicine, Iowa City, IA, United States
[4]Avel eCare, Sioux Falls, SD, United States
[5]Department of Family Medicine, University of South Dakota Sanford School of Medicine, Sioux Falls, SD, United States
[6]Department of Psychiatry, University of South Dakota Sanford School of Medicine, Sioux Falls, SD, United States

Corresponding Author:
M Muska Nataliansyah, MPH, MD, PhD
Department of Surgery
Collaborative for Healthcare Delivery Science
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI, 53005
United States
Phone: 1 4149551402
Email: mnataliansyah@mcw.edu

Abstract

Background: A mental health crisis can create challenges for individuals, families, and communities. This multifaceted issue often involves different professionals from law enforcement and health care systems, which may lead to siloed and suboptimal care. The virtual crisis care (VCC) program was developed to provide rural law enforcement with access to behavioral health professionals and facilitated collaborative care via telehealth technology.

Objective: This study was designed to evaluate the implementation and use of a VCC program from a telehealth hub for law enforcement in rural areas.

Methods: This study used a mixed methods approach. The quantitative data came from the telehealth hub's electronic record system. The qualitative data came from in-depth interviews with law enforcement in the 18 counties that adopted the VCC program.

Results: Across the 181 VCC encounters, the telehealth hub's recommended disposition and the actual disposition were similar for remaining in place (n=141, 77.9%, and n=137, 75.7%, respectively), voluntary admission (n=9, 5.0%, and n=10, 5.5%, respectively), and involuntary committal (IVC; n=27, 14.9%, and n=19, 10.5%, respectively). Qualitative insights related to the VCC program's implementation, use, benefits, and challenges were identified, providing a comprehensive view of the virtual partnership between rural law enforcement and behavioral health professionals.

Conclusions: Use of a VCC program likely averts unnecessary IVCs. Law enforcement interviews affirmed the positive impact of VCC due to its ease of use and the benefits it provides to the individuals in need, the first responders involved, law enforcement resources, and the community.

(JMIR Ment Health 2023;10:e42610) doi:10.2196/42610

KEYWORDS

mental health; telehealth; rural health; coresponder model; implementation; mixed methods; community; technology; virtual care

Introduction

Background

A mental health crisis can cause an emergency response, disruption, and long-term problems for individuals, families, and communities [1,2]. This multifaceted issue often involves law enforcement, health care, and justice systems, potentially leading to siloed and suboptimal care [3]. Enhancement in collaborative responses to and outcomes of mental health crises requires improving the care delivery model and organization of resources [4,5].

Various crisis response models have been developed to address the need for appropriate emergency responses for people with mental health conditions [6]. Among these models, 2 major examples are the crisis intervention team (CIT) model and the coresponder model [7]. The CIT model requires 40 hours of training to equip law enforcement officers with the needed knowledge and skills to coordinate mental health care. In contrast, the coresponder model emphasizes pairing law enforcement officers with behavioral health professionals [6,7]. Although law enforcement officers are often provided with some crisis intervention training, there are indications that working with behavioral health professionals can add layers of expertise and improve the service to communities [7,8]. Considering the potential value of partnership in addressing mental health crises, this study was designed to evaluate the implementation and use of a virtual crisis care (VCC) program from a telehealth hub for law enforcement in rural areas.

Coresponder Model

Studies have shown that having behavioral health professionals collaborate with law enforcement officers on mental health crisis calls reduces the burden on the justice system and improves access to mental health care [9,10]. Most studies have been conducted on in-person mental health services in urban areas [11-14]. However, many rural areas are experiencing a workforce shortage of qualified behavioral health professionals, which makes the in-person coresponder model largely infeasible. An alternative is bringing a behavioral health professional virtually to a crisis encounter via telehealth to support law enforcement officers as they assist people in rural and underserved communities by offering timely behavioral health services.

One primary challenge in mental health emergencies encountered by law enforcement officers is the decision to commit people with mental health conditions involuntarily. This decision can be resource intensive and lead to suboptimal care [15]. A survey of involuntary admitted individuals showed that most of them felt pressured to be hospitalized and were more likely to report a subjective lack of improvement [16]. Studies also found that involuntary committal (IVC) is associated with poorer outcomes [17,18]. Legal procedures related to IVCs can vary by state and may include extensive requirements for documentation [19]. However, in most states, the IVC process requires a mental illness diagnosis and a judge’s signature to issue the IVC order [19].

Since geographical boundaries, local policies, and limitation of resources can shape the delivery of care, there is a need to explore the implementation of a coresponder model in rural areas [3,4,20]. This is especially relevant in South Dakota, where more than half of the population lives in rural areas and an estimated 10% of all emergency calls involve a person who may have a mental health crisis [21].

Virtual Crisis Care Program

The Leona M. and Harry B. Helmsley Charitable Trust provided Avera eCARE (now known as Avel eCare) with a grant to establish and implement telehealth programs that would help meet the behavioral health needs of rural communities in South Dakota. Specifically, the grant supported the development of a mobile crisis care service for law enforcement and probation officers called VCC.

The VCC's pilot program was initiated in the field in January 2020. Over the subsequent first 9 months of the pilot program, the service was implemented in 18 county sheriff departments and 1 judicial circuit court probation district representing probation offices within 8 (44.4%) of those counties. Avera eCARE served as the hub for the VCC program and staffed it at all times with qualified behavioral health professionals who had experience in providing virtual and in-person consultations. Law enforcement officers in the participating rural counties were trained and supplied with electronic tablets to connect with the VCC hub for virtual mental health crisis support. The hub provided training materials and conducted sessions to help law enforcement officers get accustomed to the service. Law enforcement officers used VCC to acquire immediate and timely mental health evaluations by professionals trained in communicating and recognizing when someone would be a danger to themselves or others. The recommendations provided by the VCC connection informed the decisions made by law enforcement officers in caring for individuals in crisis.

The COVID-19 pandemic provided an opportunity to showcase the value of using telehealth to bridge this partnership. The use of familiar technology also contributed to the positive response and willingness to implement the innovative program. Overall, the virtual consultation aimed to de-escalate and stabilize the mental health condition of an individual during a crisis situation. After responding to a crisis call and conducting an initial assessment, law enforcement officers decided whether to offer and initiate the VCC service. Following a telehealth consultation with the person in crisis using the provided tablets, the VCC behavioral health professional shared their recommendation with law enforcement officers, who ultimately decided the disposition of the individual. Means restriction (ie, reducing a suicidal individual’s access to lethal means) and safety planning (ie, reducing the risk of harm to oneself) were strategies used to help individuals remain in place [22,23]. The hub team strived to collaboratively partner with local mental health community organizations to further assist the individuals establish outpatient resources, if needed.

Methods

Study Design

Using a mixed methods explanatory design, an independent team of researchers gathered data using both quantitative and qualitative approaches. The telehealth hub collected all quantitative data in its electronic record system. Variables assessed included the gender of the person served, mental health conditions identified at the time of the encounter, and the location of use. Specifically for location, *home* was identified when the incidence occurred at a residence, the *jail* was the county detention facility, and the *community* included any other public place or property. The recommended disposition by the telehealth hub and the actual disposition, which was verified by each county, were also captured and evaluated. These included remaining in place or getting admitted to a mental health facility (further classified as voluntary or involuntary). There may have also been continued law enforcement involvement until the individual was stabilized.

We tabulated several county-level measures to describe the rurality of these areas, health care coverage, and area deprivation. For rurality, we used the second-level rural-urban commuting area codes to classify as rural and also report the population density of these counties [24]. For health care access measures, we used data from the County Health Rankings data for these counties, including ratios of primary care providers and mental health providers [25]. Additionally, we described these counties by a measure of the area deprivation index percentile, comparing the rankings of these counties with national percentiles [26].

In-depth qualitative interviews were conducted to collect insights from users of the service. We used purposive sampling by inviting law enforcement officers from counties that adopted the VCC program to participate in a semistructured interview. These contacts were provided to the research team by the VCC hub team, indicating that the counties had experience using VCC and were willing to share perspectives. In the end, we were able to collect insights from the listed key informants who were most familiar with the VCC program from 10 different counties. This completion allowed us to be more confident in the generated findings, since it included the vast majority and full range of usage in the total sample. The telephone interviews—each ranging between 30 and 60 minutes—took place between January and May 2021. Two experienced qualitative researchers conducted the interviews. Both have years of experience conducting telehealth research and qualitative methods. Questions addressed topics including why their county was interested in VCC, the motivation to use the service, descriptions of typical and unique situations when used, community reactions and the willingness to use, and overall opinion of the service. We used these questions not only to gain insights into the implementation process but also to acquire some idea of how the community responded to the availability of VCC service.

The qualitative component included 14 interviews with 15 law enforcement officers in sheriff's departments in 10 counties, with 1 (10%) county extending interviews to city police department users. The interviews were audio-recorded, transcribed, and reviewed independently by 2 research team members. One coder had a background in health organization and implementation science, and the other was a qualitative specialist, having studied telehealth applications in multiple settings. They used an inductive qualitative analysis approach to identify relevant quotes and themes related to VCC use [27,28]. The researchers then met to discuss their findings, which led to the identification of themes and a composition of relevant quotes. The coders discussed the emerging findings with another research team member to enhance the validity of the results. Any discrepancies were discussed until consensus was reached. A final step was to classify the themes into domains and review all coding with the full research team. The process generated 3 domains that contained associated themes: (1) the *implementation of VCC* domain includes themes relevant to the initiation of the service, the need for the program, and why the program is important to law enforcement; (2) the *use of VCC* domain includes themes relevant to the experience of using the service or how law enforcement officers include the VCC program in their daily routine; and (3) the *benefits and challenges of VCC* domain includes themes relevant to the opinions of the service with lessons learned. The availability of multiple collaborators with different backgrounds and familiarity with the subject facilitated discussions with valuable perspectives that complemented each other. We used Microsoft Excel to manage and categorize quotes from different interviewees into themes and domains. The research followed the Consolidated Criteria for Reporting Qualitative Research (COREQ) for reporting purposes [29].

Ethical Considerations

The project protocol was reviewed by the Institutional Review Board (IRB) of the University of Iowa and did not meet the regulatory definition of human subjects' research, because the information sought from interviewees would not be private information concerning the interviewee—it would be information about the VCC services. All methods were carried out in accordance with relevant guidelines and regulations. Participation was voluntary, and no compensation was provided. Participants' identities were confidential and de-identified, and data were presented anonymously.

Results

VCC Use in a Rural Setting

In total, 18 counties and 1 judicial circuit court probation district implemented VCC during the pilot program, accounting for 181 encounters during the study period (Table 1). Usage varied from 0 to 34 encounters per site and was unrelated to the length of time in use. About two-thirds of the consultations were with men (n=113, 62.4%), one-third with women (n=67, 37.0%), and 1 (0.6%) was missing this data. The most common location for using the service was in the community (n=81, 44.8%), followed by home (n=73, 40.3%), with a smaller percentage occurring in jail (n=27, 14.9%).

Table 1. Characteristics of individuals by VCC[a] encounter setting.

Characteristics	Total (N=181), n (%)	Community (n=81, 44.8%), n (%)	Home (n=73, 40.3%), n (%)	Jail (n=27, 14.9%), n (%)
Gender				
Male	113 (62.4)	43 (53.1)	46 (63.0)	24 (88.9)
Female	67 (37.0)	37 (45.7)	27 (37.0)	3 (11.1)
Missing	1 (0.6)	1 (1.2)	0	0
Nature of request				
Suicidal ideation or self-harm	97 (53.6)	44 (54.3)	37 (50.7)	16 (59.3)
Depression or anxiety	48 (26.5)	24 (29.6)	19 (26.0)	5 (18.5)
Aggressive behavior	29 (16.0)	11 (13.6)	13 (17.8)	5 (18.5)
Delusion or hallucinations	7 (3.9)	2 (2.5)	4 (5.5)	1 (3.7)

[a]VCC: virtual crisis care.

Encounters resulted in VCC consultations for various mental health situations, of which suicidal ideation or self-harm represented over half of all encounters (n=97, 53.6%). Other mental health situations identified for VCC consultations included depression or anxiety (n=48, 26.5%), aggressive or disruptive behavior or bullying (n=29, 16.0%), and delusion or hallucinations (n=7, 3.9%).

The 18 counties in this assessment have 78 census tracts included in them, of which 50 (64%) are considered rural, as defined by their rural-urban commuting area codes. The median population density (people per square mile) is 166 (IQR 8-1139). The average ratios for primary care providers and mental health providers per population are 1:1935 and 1:1004, respectively. Using the area deprivation index percentile comparing these counties to national estimates, the median percentile was 72 (IQR 58-82), indicating a relatively high level of socioeconomic disadvantage.

The behavioral health professional at the telehealth hub made recommendations directly to law enforcement officers at the conclusion of the VCC encounter (Table 2). The most common recommendation was that the person remain in place (n=141, 77.9%). The behavioral health professional occasionally recommended that the person in crisis be voluntarily admitted (n=9, 5.0%) or be involuntarily committed (n=27, 14.9%). Overall, the actual disposition agreed with the recommended disposition by VCC for 92% of encounters with a known disposition (ie, neither disposition was missing). The telehealth hub's recommended disposition and actual disposition were similar for remaining in place (n=141, 77.9%, and n=137, 75.7%, respectively) and voluntary admission (n=9, 5.0%, and n=10, 5.5%, respectively). In most encounters, the law enforcement officers followed the VCC providers' recommendation but gave a rationale for the few instances when other actions were taken. Differences in recommended and actual dispositions for IVCs were primarily due to encounters in the community, such that only 8 (9.9%) of the community encounters resulted in an IVC compared to the recommended 14 (17.3%). Considering the limited sample, we only presented a descriptive statistic and did not conduct further statistical analyses to avoid misinterpretation of the results.

Table 2. Distribution of recommended and actual dispositions by location of incident.

Dispositions	Community (n=81, 44.8%), n (%)		Home (n=73, 40.3%), n (%)		Jail (n=27, 14.9%), n (%)		Total rec.[a] (N=181), n (%)	Total act.[b] (N=181), n (%)
	Rec.	Act.	Rec.	Act.	Rec.	Act.		
Remain in place	63 (77.8)	64 (79.0)	51 (69.9)	46 (63.0)	27 (100.0)	27 (100.0)	141 (77.9)	137 (75.7)
Voluntary admission[c]	2 (2.5)	3 (3.7)	7 (9.6)	7 (9.6)	0	0	9 (5.0)	10 (5.5)
IVC[d]	14 (17.3)	8 (9.9)	13 (17.8)	11 (15.1)	0	0	27 (14.9)	19 (10.5)
Law enforcement (other)[e]	0	4 (4.9)	0	1 (1.4)	0	0	0	5 (2.8)
Missing	2 (2.5)	2 (2.5)	2 (2.7)	8 (11.0)	0	0	4 (2.2)	10 (5.5)

[a]Rec: recommended disposition by the virtual crisis care (VCC) program.

[b]Act: actual disposition by law enforcement personnel.

[c]Admissions included higher levels of care in a health care setting, such as inpatient stay, emergency department visit, or transfer to a psychiatric facility.

[d]IVC: involuntary committal.

[e]Continuous monitoring by law enforcement until the individual was stabilized.

Insights Into VCC Implementation in Rural Settings

In-depth interviews with 15 law enforcement officers from 10 counties provided helpful information about learning points from the VCC implementation in rural areas. These 10 counties experienced 155 (85.6%) of the 181 VCC encounters completed over the 18-month pilot.

Identified themes were grouped into 3 main categories: the initiation/implementation of VCC, the use of VCC, and the benefits and challenges of VCC (Table 3). Selected quotes representing the themes are presented in Table 4.

Table 3. Qualitative insights into the VCC[a] program.

Themes and subthemes	Details
Theme 1: implementation of VCC	• Demographic and geographic setting • Resource-intensiveness of IVCs[b] • Relevant technology and service training
Theme 2: use of VCC	• Presenting the service • The consultation process • Recommendations and decisions
Theme 3: benefits and challenges of VCC	
Subtheme 3A: benefits	• Mitigation of IVC procedures • Better documentation for mental health crisis encounters • Providing quality and efficient access to mental health experts
Subtheme 3B: challenges	• Technical concerns • Limited network availability

[a]VCC: virtual crisis care.

[b]IVC: involuntary committal.

Table 4. Selected quotes on the VCC[a] program in the rural United States.

Themes and subthemes	Quotes
Theme 1: implementation of VCC	
Demographic and geographic setting	"We serve about 2350 square miles. We're a big county, but not a big population." [Site C]
Resource-intensiveness of IVCs[b]	"Usually, we burn about 24 man-hours every time we have a mental health commitment that's involuntary. And under state law, we're obligated. If we feel that a person is a harm to them self or to others, we're obligated to act. And so, it utilizes a lot of manpower and man-hours to accomplish that goal." [Site I]
Relevant technology and service training	"The majority of the training was how to use the tablets, how to log-in to those, how to make sure they're working, and then real simple instructions on going forward. We went through the criteria on what was needed for us to make the phone call, the questions we needed to ask of the person in crisis, the information we needed to gather before we made the phone call and got online with the doc." [Site M]
Theme 2: use of VCC	
Presenting the service	"So, if it's in the home, we'll talk to them about that we've got these options that we could take them in or we have a nice new program where they can talk with a behavioral health professional on an iPad within the comfort of their own home. And then that way somebody who is more trained than I am with mental health and counseling will have an opportunity to talk with them. If they agree to that, then we'll call the iPad, get that set up, call ahead for the online behavioral health professional, and inform them of the situation and the person's name." [Site A]
The consultation process	"If it's a person who I feel shouldn't be left alone, I stay in the room. But, normally, after gaining rapport with them and speaking with them, I tell them, "Hey, I'm going to give you some one-on-one time with the mental health professional. If you need anything, I'll be in the next room." I feel like by me stepping out of the room and giving them some one-on-one time, it really opens them up to saying more to the mental health professional who's on the other end of the tablet." [Site H]
Recommendations and decisions	"And then when they're done, we would talk to the behavioral health professional and see what their recommendations are knowing that we're on scene and we still have the final say. But we'll take that into consideration." [Site N]
Theme 3A: benefits of VCC	
Potential mitigation of IVC procedures	"I think any of them that we've done the [VCC] on and follow the [mental health professional] recommendations to either turn to a safety plan or return to their home environment, I think all of those were prevented IVCs. I'm not saying they would've all been ordered for continued detention after the IVC, but I think every one of those 34 that the recommendation was to not continue with the IVC, which I think we've had two out of the 34 that they said, yeah, you should do an IVC anyway. So, 32 of them I think were prevented involuntary commitments." [Site J]
Better documentation for mental health crisis encounters	"It's documented very well. The whole encounter is very detailed in how it's logged. I think it's beneficial for us to get [the VCC report] back so we can write our report and be thorough." [Site N]
Providing quality and efficient access to mental health experts	"Using the [VCC] system has been fantastic both from the officer's point of view and from the client's point of view, the person who's in crisis. Because they tell us that. They don't get hauled away, taken away to a facility in another community where they're held until they're evaluated and deemed whether or not they can go home. In these cases [where VCC is used], they've been able to talk with somebody. A lot of times that seems to help, just being able to vent to the right person and then come up with some sort of game plan, a plan of action for them to follow up on whether it's seeing a behavioral health professional or whatever it may be." [Site K]
Theme 3B: challenges of VCC	
Technical concerns	"I know the one time with my juvenile, I called the number. I got a hold of the behavioral health professional. We set up an appointment via the iPad, but for some reason it wasn't connecting. I thought I had good cell reception. I don't know if it was some technical issue, but we just overcame that and she just talked to him on my cell phone." [Site F]
Limited network availability	"It happened one time where the internet was a little tough to deal with. And the only thing that happened was the subject and [mental health professional], they could hear each other, but as far as the screen and stuff, the screen would kind of pause. That has happened the one time, but everybody could hear everything. It was just you couldn't have that screen-to-screen interaction." [Site B]

[a]VCC: virtual crisis care.

[b]IVC: involuntary committal.

The Implementation of VCC

The themes presented in this section highlight factors that influenced VCC implementation. The first factor that motivated VCC implementation was the demographic and geographic setting. The South Dakota county sheriff departments serve extensive rural areas with limited resources. Beyond the regular population of local counties, tourist activity can skyrocket the number of people and increase the workload for small county sheriff departments.

The second factor that motivated VCC implementation was the resource-intensiveness of IVCs. One important responsibility of county sheriff departments in South Dakota is to transport to mental health facilities individuals who are being involuntarily committed due to mental health crises. These transfers take time and money due to long travel distances—up to 500 miles in some cases; if the transfer is unnecessary, it creates undue strain on the person in crisis, the health system, and local law enforcement.

The third factor influencing VCC implementation was relevant training from the telehealth hub. As soon as county sheriff departments were on board to participate in this pilot VCC program, the VCC hub implementation staff provided equipment and training to the county law enforcement officers. The interviewees indicated the equipment was easy to use and simple to implement and that the training was straightforward. Test calls were encouraged, and ongoing training, troubleshooting, and communication were always available.

The Use of VCC

The themes in this section illustrate the process of using the VCC service. The first step in the consultation process was presenting the service to individuals with mental health conditions. Law enforcement officers responded to calls—most of them coming from family members or neighbors who had concerns for friends or loved ones—that presented requests ranging from welfare checks to crisis intervention. As shown in Table 1, more than half of the requests involved suicidal ideation or self-harm. Once law enforcement responded to a call of this nature, they then determined whether to use VCC based on their experience, comfort in offering, and the person in crisis.

The second step was to allow the individual in crisis time and space to consult with the VCC behavioral health professional, while maintaining safety for all. The telehealth encounter using the electronic tablet took between 15 and 60 minutes, depending on the circumstances. At the final step of the encounter, the behavioral health professional reconnected with law enforcement and provided a recommendation. Law enforcement officers had the option of carrying through with the VCC recommendation or choosing a different course of action based on additional knowledge or extenuating circumstances.

In this study, law enforcement overwhelmingly followed the VCC behavioral health professional's recommendation, but in the few instances when other actions were taken, they gave a rationale for the departure. These exceptions included hearing the person lie, intoxication, having the person attempt self-harm after the call, being called back to the home within hours, having the safety plan refused or not implemented after the call, or, in several situations, having to follow through with an arrest warrant.

The Benefits and Challenges of VCC

The themes in this section summarized lessons learned from the VCC pilot program and included insights into the benefits and challenges of the VCC service.

Benefits

According to law enforcement, the various benefits of the VCC program have shown a positive direct impact on individuals experiencing mental health crises and the law enforcement officers involved, plus the broader value of having the service in these rural communities.

The first benefit of VCC is potential mitigation of IVC procedures, which could lighten the use of already limited resources. South Dakota has a legal procedure that law enforcement must follow for IVC. This includes proper documentation, evaluations by a qualified mental health professional, and forms to be signed, followed by County Bureau of Mental Illness review and decision, which can be a lengthy process [30]. One interviewee shared that before a resource like VCC was available, just the mention or threat of suicide would prompt an IVC in past scenarios. Multiple interviewees emphasized how working with the VCC hub behavioral health professional facilitates an opportunity to acquire more information, promote crisis reduction, work out a safety plan, and frequently prevent an IVC from occurring.

The second benefit was better documentation for mental health crisis encounters. Postencounter reports from VCC behavioral health professionals shared with law enforcement provided a detailed, timed report about the encounter and interaction with the individual in crisis and the rationale for the disposition recommendation. This documentation has proven helpful to the sheriff departments, which attach the information to law enforcement incident reports, by developing a history and patterns for individuals that could be beneficial in the future, as verification for decisions to pursue IVCs when necessary and as documentation of actions supporting stay-in-place decisions.

A third benefit of VCC was providing the rural community with quality and efficient access to mental health experts. Quality elements of VCC, such as timeliness, convenience, and expertise in assessment, were noted as benefits. It was noted that the virtual conversation between the individual in crisis and the behavioral health professional often diffuses the situation. Officers commented that they greatly value the involvement of expertly trained behavioral health professionals. In addition, interviewees reported that VCC helps them do their jobs better for the communities they serve.

Challenges

As in any pilot program, the law enforcement participants understood VCC to be a work in progress. From their experience, interviewees provided suggestions for improvements to further meet law enforcement needs.

One of the challenges was presenting the VCC technology out in the field. When the VCC program was initialized, some

officers experienced technical difficulties or connectivity issues. Another difficulty was encouraging an irrational person in crisis to interact with the telehealth tablet and engage the VCC behavioral health professional. However, as 1 officer stated, the timing worked in favor of VCC because COVID-19 has led to many people becoming more familiar with using virtual platforms for communication.

Another challenge was limited network availability in highly rural areas. To address this issue, law enforcement officers used an internet hotspot or cell phone booster technology in patrol cars and were aware of locations in the county where broadband service was available. In fact, when preparing to launch VCC in some counties, the officers did sweeps of their counties' geography to note good broadband locations. Consequently, some encounters moved from home to car or a nearby open space or building.

Discussion

Principal Findings

Law enforcement officers are commonly the first responders to a mental health crisis in rural communities. By partnering virtually with behavioral health professionals in these situations, law enforcement can address the gap in having suitable local experts to determine appropriate care. South Dakota sheriff departments got an opportunity to use technology with the VCC pilot program to virtually bring a behavioral health professional to individuals in crisis. This mixed methods study examined VCC encounter characteristics and perceptions by rural law enforcement users.

VCC consultations were frequently used for males and individuals with suicidal ideation or depression. These findings are similar to the results of existing studies on different crisis response models [7,12], although requests for aggressive or disruptive behaviors in crisis responses occurred infrequently during the VCC pilot program. This result shows the importance of having a behavioral health professional address the issue of potential self-harm and recommend proper care that is potentially lifesaving [6,9]. By having the behavioral health professional's evaluation and recommendation for next-step care, law enforcement saved time and resources previously expended to conduct IVCs [10]. Moreover, this type of program provides an opportunity to connect individuals in mental health crises with appropriate care and resources and to avert their involvement in the criminal justice system [31,32].

The potential mitigation of IVCs was seen as 1 of VCC's primary benefits conveyed by law enforcement during qualitative interviews. Other benefits include access to behavioral health professionals virtually because those experts are rarely available locally in rural areas. This virtual access led to crisis situations being diffused, safety plans being initiated, and better documentation of care plans and rationales. Otherwise, the limited resources and expertise in rural settings often serve as significant hurdles to providing appropriate care for individuals with mental health issues [20,33,34]. The use of technology and the availability of virtual experts can potentially overcome multiple hurdles, as shown in this study. However, technology-related challenges (ie, technical concerns and network availability) should be anticipated and addressed before program implementation.

Insights from the interviews showed that introducing technology and relevant collaborative training in the implementation phase is crucial to building law enforcement confidence and familiarity with the program. This finding is consistent with the implementation of a coresponder program in an urban area, where initial agency collaboration and team building were identified as key facilitators of the program [35]. Moreover, it is essential for law enforcement to understand their jurisdiction characteristics and resource limitations to foster motivation and buy-in for program implementation. Indeed, 1 barrier previously identified for crisis response program effectiveness is a lack of buy-in and support from related agencies [10,20]. Future implementation efforts should consider team-building activities and identify key values of the service in the local context to facilitate successful preparation.

Our study also gathered insights into service use patterns following the VCC preparation. The process of presenting the VCC service, providing space for the virtual consultation, and making a decision based on expert recommendations via a structured protocol can be considered for future use. Indeed, systematic studies have highlighted the importance of building a standardized practice for replicability and evaluation purposes without neglecting the local context [5,9]. The availability of virtual experts does not diminish law enforcement's role in this situation. This study shows that law enforcement officers are still the primary contact and decision makers. They provided the service with a proactive approach, while considering individuals' safety. The VCC program and technology provided law enforcement with tools to make informed decisions and facilitated better care for needed individuals. This social partnership model is perceived positively not only by law enforcement but also by individuals in crisis [36].

A prominent model for implementation in public service sectors has outlined 4 implementation phases (ie, exploration, preparation, implementation, and sustainment), including the influencing factors [37]. This study provides informative insights into law enforcement's experience during the early phases of VCC implementation. Future studies can build upon the findings of this study and develop a focus on ways to make this crisis response model sustainable and analyze its cost-effectiveness.

Limitations

There are a couple of limitations to this study. First, the interviewees were limited to those who participated in the VCC program and research and reasons for not participating were not captured. However, the quantitative data covered a diverse group of jurisdictions and the qualitative data provided rich insights into the VCC implementation experience. Moreover, the current results are similar to the existing literature in terms of the characteristics of the persons served, while extending the application of the coresponder program model into rural settings using virtual technology. Second, for outcome disposition, there were several encounters for which no final disposition could be ascertained from counties. As a result, the true proportion of concordance between recommended and actual dispositions

may differ. Third, we do not have information about the counterfactual condition, meaning that we cannot yet compare the effectiveness of the VCC program with counties that do not have access to the program or the condition before the implementation of the program. This limitation can be the next direction of future relevant projects.

Conclusion

Per the VCC pilot program operating in South Dakota, we found evidence from both quantitative and qualitative data that using telehealth to engage behavioral health professionals in mental health crisis encounters may positively impact efforts in addressing mental health crises. It has potentially saved individuals the stress and negative health outcomes associated with IVCs, it helps law enforcement by better preparing them to serve their local communities, and it helps counties preserve resources and reallocate time and energy otherwise dedicated toward IVCs.

Noting the acceptance and benefit to rural counties, Avera eCARE reached out to South Dakota legislators to promote the service statewide to all 66 counties. In the 2021 session, the South Dakota legislature supported the continuation and expansion of VCC services in the state [38]. As of this writing, VCC has been implemented in and serves 40 counties in South Dakota [39].

Acknowledgments

The authors thank the interviewees who participated in this research for sharing their knowledge and experiences. The authors thank Amanda Bell of Avera Health/Avel eCare for guidance on the study design and Mark Pals, formerly of Avera Health, for contributions during data collection. The paper does not include copyrighted materials. The paper's content is solely the authors' responsibility and does not necessarily represent the official position, policy, or views of the Leona M and Harry B Helmsley Charitable Trust.

Conflicts of Interest

MMN, KASM, JPV, and MMW have no competing interests. LJM and SP are employed by Avera Health/Avel eCare, which provides the VCC service.

References

1. Saleh AZ, Appelbaum PS, Liu X, Scott Stroup T, Wall M. Deaths of people with mental illness during interactions with law enforcement. Int J Law Psychiatry 2018;58:110-116. [doi: 10.1016/j.ijlp.2018.03.003] [Medline: 29853001]
2. Pollack HA, Humphreys K. Reducing violent incidents between police officers and people with psychiatric or substance use disorders. Ann Am Acad Polit Soc Sci 2020 Feb 13;687(1):166-184. [doi: 10.1177/0002716219897057]
3. Pepler EF, Barber CG. Mental health and policing: picking up the pieces in a broken system. Healthc Manag Forum 2021 Mar;34(2):93-99. [doi: 10.1177/0840470420979635] [Medline: 33499683]
4. National guidelines for behavioral health crisis care: best practice toolkit. Substance Abuse and Mental Health Services Administration (SAMHSA). 2020. URL: https://www.samhsa.gov/sites/default/files/national-guidelines-for-behavioral-health-crisis-care-02242020.pdf [accessed 2023-02-27]
5. Balfour ME, Hahn Stephenson A, Delany-Brumsey A, Winsky J, Goldman ML. Cops, clinicians, or both? Collaborative approaches to responding to behavioral health emergencies. Psychiatr Serv 2022 Jun;73(6):658-669. [doi: 10.1176/appi.ps.202000721] [Medline: 34666512]
6. Watson A, Compton M, Pope L. Crisis Response Services for people with Mental Illnesses or Intellectual and Developmental Disabilities: A Review of the Literature on Police-Based and Other First Response Models. New York, NY: Vera Institute of Justice; 2019.
7. Seo C, Kim B, Kruis NE. Variation across police response models for handling encounters with people with mental illnesses: a systematic review and meta-analysis. J Crim Justice 2021 Jan;72:101752. [doi: 10.1016/j.jcrimjus.2020.101752]
8. Virtual crisis care in South Dakota. Crime and Justice Institute. 2021. URL: https://www.cjinstitute.org/publication/south-dakotas-virtual-crisis-care-pilot-program-a-model-for-rural-states/ [accessed 2023-02-27]
9. Puntis S, Perfect D, Kirubarajan A, Bolton S, Davies F, Hayes A, et al. A systematic review of co-responder models of police mental health 'street' triage. BMC Psychiatry 2018 Aug 15;18(1):256 [FREE Full text] [doi: 10.1186/s12888-018-1836-2] [Medline: 30111302]
10. Shapiro GK, Cusi A, Kirst M, O'Campo P, Nakhost A, Stergiopoulos V. Co-responding police-mental health programs: a review. Adm Policy Ment Health 2015 Sep;42(5):606-620. [doi: 10.1007/s10488-014-0594-9] [Medline: 25239523]
11. Blais E, Landry M, Elazhary N, Carrier S, Savard A. Assessing the capability of a co-responding police-mental health program to connect emotionally disturbed people with community resources and decrease police use-of-force. J Exp Criminol 2020 Jun 12;18(1):41-65. [doi: 10.1007/s11292-020-09434-x]
12. Helfgott JB, Hickman MJ, Labossiere AP. A descriptive evaluation of the Seattle Police Department's crisis response team officer/mental health professional partnership pilot program. Int J Law Psychiatry 2016;44:109-122. [doi: 10.1016/j.ijlp.2015.08.038] [Medline: 26314890]

13. Kirst M, Francombe Pridham K, Narrandes R, Matheson F, Young L, Niedra K, et al. Examining implementation of mobile, police-mental health crisis intervention teams in a large urban center. J Ment Health 2015 Dec;24(6):369-374. [doi: 10.3109/09638237.2015.1036970] [Medline: 26383041]
14. Kisely S, Campbell LA, Peddle S, Hare S, Pyche M, Spicer D, et al. A controlled before-and-after evaluation of a mobile crisis partnership between mental health and police services in Nova Scotia. Can J Psychiatry 2010 Oct 01;55(10):662-668. [doi: 10.1177/070674371005501005] [Medline: 20964945]
15. Road runners: the role and impact of law enforcement in transporting individuals with severe mental illness. A national survey. Treatment and Advocacy Centers. 2019. URL: https://www.treatmentadvocacycenter.org/storage/documents/Road-Runners.pdf [accessed 2023-02-27]
16. Bonsack C, Borgeat F. Perceived coercion and need for hospitalization related to psychiatric admission. Int J Law Psychiatry 2005;28(4):342-347. [doi: 10.1016/j.ijlp.2005.03.008] [Medline: 15936078]
17. Houston KG, Mariotto M. Outcomes for psychiatric patients following first admission: relationships with voluntary and involuntary treatment and ethnicity. Psychol Rep 2001 Jun;88(3 Pt 2):1012-1014. [doi: 10.2466/pr0.2001.88.3c.1012] [Medline: 11597046]
18. Fennig S, Rabinowitz J, Fennig S. Involuntary first admission of patients with schizophrenia as a predictor of future admissions. Psychiatr Serv 1999 Aug;50(8):1049-1052. [doi: 10.1176/ps.50.8.1049] [Medline: 10445653]
19. Civil commitment and the mental health care continuum: historical trends and principles for law and practice. Substance Abuse and Mental Health Services Administration (SAMHSA). 2019. URL: https://www.samhsa.gov/resource/ebp/civil-commitment-mental-health-care-continuum-historical-trends-principles-law [accessed 2023-02-27]
20. Bratina MP, Carsello JA, Carrero KM, Antonio ME. An examination of crisis intervention teams in rural jurisdictions. Community Ment Health J 2021 Oct;57(7):1388-1398. [doi: 10.1007/s10597-021-00797-7] [Medline: 33638058]
21. Virtual crisis care program. South Dakota Unified Judicial System. 2022. URL: https://ujs.sd.gov/Resources/VirtualCrisisCare.aspx [accessed 2023-02-27]
22. Daigle MS. Suicide prevention through means restriction: assessing the risk of substitution. A critical review and synthesis. Accid Anal Prev 2005 Jul;37(4):625-632. [doi: 10.1016/j.aap.2005.03.004] [Medline: 15949453]
23. Yip PSF, Caine E, Yousuf S, Chang S, Wu KC, Chen Y. Means restriction for suicide prevention. Lancet 2012 Jun 23;379(9834):2393-2399 [FREE Full text] [doi: 10.1016/S0140-6736(12)60521-2] [Medline: 22726520]
24. Rural-urban commuting area codes. USDA Economic Research Service. 2020. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/ [accessed 2023-02-27]
25. Rankings data and documentation. County Health Rankings and Roadmaps. 2022. URL: https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation [accessed 2023-02-27]
26. Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible - the neighborhood atlas. N Engl J Med 2018 Jun 28;378(26):2456-2458 [FREE Full text] [doi: 10.1056/NEJMp1802313] [Medline: 29949490]
27. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2016 Jun 30;27(2):237-246. [doi: 10.1177/1098214005283748]
28. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
29. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
30. Qualified mental health professional and Board of Mental Illness. South Dakota Department of Social Services. 2020. URL: https://dss.sd.gov/docs/behavioralhealth/board_of_mental_illness/Online_Training_PowerPoint.pdf [accessed 2023-02-27]
31. Kennedy-Hendricks A, Huskamp HA, Rutkow L, Barry CL. Improving access to care and reducing involvement in the criminal justice system for people with mental illness. Health Aff (Millwood) 2016 Jun 01;35(6):1076-1083. [doi: 10.1377/hlthaff.2016.0006] [Medline: 27269025]
32. Hartford K, Carey R, Mendonca J. Pre-arrest diversion of people with mental illness: literature review and international survey. Behav Sci Law 2006;24(6):845-856. [doi: 10.1002/bsl.738] [Medline: 17171772]
33. Tailoring crisis response and pre-arrest diversion models for rural communities. Substance Abuse and Mental Health Services Administration (SAMHSA). 2019. URL: https://store.samhsa.gov/product/Tailoring-Crisis-Response-and-Pre-Arrest-Diversion-Models-for-Rural-Communities/PEP19-CRISIS-RURAL [accessed 2023-02-27]
34. Vakkalanka JP, Nataliansyah MM, Merchant KAS, Mack LJ, Parsons S, Mohr NM, et al. Telepsychiatry services across an emergency department network: a mixed methods study of the implementation process. Am J Emerg Med 2022 Sep;59:79-84. [doi: 10.1016/j.ajem.2022.06.053] [Medline: 35810736]
35. Bailey K, Paquet SR, Ray BR, Grommon E, Lowder EM, Sightes E. Barriers and facilitators to implementing an urban co-responding police-mental health team. Health Justice 2018 Nov 22;6(1):21 [FREE Full text] [doi: 10.1186/s40352-018-0079-0] [Medline: 30467739]

36. Boscarato K, Lee S, Kroschel J, Hollander Y, Brennan A, Warren N. Consumer experience of formal crisis-response services and preferred methods of crisis intervention. Int J Ment Health Nurs 2014 Aug;23(4):287-295. [doi: 10.1111/inm.12059] [Medline: 24575860]
37. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Adm Policy Ment Health 2011 Jan;38(1):4-23 [FREE Full text] [doi: 10.1007/s10488-010-0327-7] [Medline: 21197565]
38. Ki N. South Dakota looks to adopt 988 national suicide hotline. Argus Leader. 2021. URL: https://www.argusleader.com/story/news/2021/09/14/south-dakota-legislators-discuss-establishing-988-suicide-hotline-virtual-crisis-care-program/8335471002/ [accessed 2023-02-27]
39. Ki N. Virtual crisis intervention program expands to 40 counties. Here's what that means. Argus Leader. 2022. URL: https://www.argusleader.com/story/news/2022/01/04/virtual-crisis-care-partnership-south-dakota-police-force-expands/9079004002/ [accessed 2023-02-27]

Abbreviations

CIT: crisis intervention team
IVC: involuntary committal
VCC: virtual crisis care

Edited by J Torous; submitted 11.09.22; peer-reviewed by N Natafgi, C Melillo, D Madhusudhan; comments to author 26.10.22; revised version received 17.12.22; accepted 11.01.23; published 20.03.23.

Please cite as:
Nataliansyah MM, Merchant KAS, Vakkalanka JP, Mack L, Parsons S, Ward MM
Virtual Partnership Addressing Mental Health Crises: Mixed Methods Study of a Coresponder Program in Rural Law Enforcement
JMIR Ment Health 2023;10:e42610
URL: https://mental.jmir.org/2023/1/e42610
doi:10.2196/42610
PMID:36939827

Original Paper

A Digital Peer Support Platform to Translate Online Peer Support for Emerging Adult Mental Well-being: Randomized Controlled Trial

GeckHong Yeo[1,2], PhD; Gladys Loo[3], BSc; Matt Oon[4], BSc; Rachel Pang[4], BA; Dean Ho[1,2,5,6], PhD

[1]N.1 Institute for Health, National University of Singapore, Singapore, Singapore

[2]The Institute for Digital Medicine (WisDM), Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[3]National University of Singapore, Singapore, Singapore

[4]Acceset, Singapore, Singapore

[5]Department of Biomedical Engineering, College of Design and Engineering, National University of Singapore, Singapore, Singapore

[6]Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
GeckHong Yeo, PhD
N.1 Institute for Health
National University of Singapore
28 Medical Drive
Singapore, 117456
Singapore
Phone: 65 6601 7766
Email: geckhongyeo@gmail.com

Abstract

Background: Emerging adulthood (ages 19 to 25 years) is a developmental phase that is marked by increased mental health conditions, especially depression and anxiety. A growing body of work indicates that digital peer emotional support has positive implications for the psychological functioning of emerging adults. There is burgeoning interest among health care professionals, educational stakeholders, and policy makers in understanding the implementation and clinical effectiveness, as well as the associated mechanism of change, of digital peer support as an intervention.

Objective: This randomized controlled trial (RCT) examined the effectiveness of a digital peer support intervention over a digital platform—Acceset—for emerging adult psychological well-being with 3 primary aims. First, we evaluated the implementation effectiveness of digital peer support training for individuals providing support (befrienders) and of the digital platform for peer support. Second, we assessed the clinical outcomes of digital peer support in terms of the intervening effect on emerging adult psychological well-being. Third, we investigated the mechanism of change linking the digital peer support intervention to emerging adult psychological well-being.

Methods: This RCT involving 100 emerging adults from the National University of Singapore follows the published protocol for this trial.

Results: This RCT found effectiveness in digital peer support training—specifically, befrienders' peer support responses demonstrating significantly higher post- than pretraining scores in selfhood (posttraining score: mean 62.83, SD 10.18, and SE 1.72; pretraining score: mean 54.86, SD 7.32, and SE 1.24; t_{34}=3.88; P<.001). The digital peer support intervention demonstrated clinical effectiveness in enhancing selfhood, compassion, and mindfulness and lowering depressive and anxiety symptoms among seekers in the intervention group after the intervention (mean 7.15, SD 5.14; SE 0.88) than among seekers in the waitlist control group before the intervention (mean 11.75, SD 6.72; SE 0.89; t_{89}=3.44; P<.001). The effect of the intervention on seekers' psychological well-being was sustained beyond the period of the intervention. The mechanism of change revealed that seekers' engagement with the intervention had both immediate and prospective implications for their psychological well-being.

Conclusions: This RCT of a digital peer support intervention for emerging adult psychological well-being harnesses the interventional potential of 4 components of psychological well-being and elucidated a mechanism of change. By incorporating and validating the digital features and process of a peer support platform, our RCT provides the parameters and conditions for

deploying an effective and novel digital peer support intervention for emerging adult psychological well-being in real-world settings.

Trial Registration: ClinicalTrials.gov NCT05083676; https://clinicaltrials.gov/ct2/show/NCT05083676

(JMIR Ment Health 2023;10:e43956) doi:10.2196/43956

KEYWORDS

mental health; digital health; peer support intervention; peer emotional disclosure; randomized controlled trial

Introduction

Background and Rationale

Emerging adulthood is a development phase that spans ages 19 to 25 years in most high-income economies [1]. This developmental period is characterized by heightened fluctuations in positivity (ie, the degree to which individuals experience positive emotions) and negativity (ie, the degree to which individuals experience negative emotions) [1]. These normative emotional fluctuations are associated with greater propensity for anxiety and depression, which establishes mental health as a pressing concern among emerging adults [1]. The onset of the global pandemic has resulted in increased distress and systematic changes in positivity and negativity among college students [2,3]. They experienced a decrease in positivity and an increase in negativity that were associated with greater depressive symptoms and anxiety [2,3].

COVID-19's social implications, which involve uncertainty, insecurity, and a reduced sense of agency and self-directedness, relate most directly to emerging adulthood [4]. The prolonged isolation and reduced social connections brought about by the pandemic have driven college students' social connections on the web [5,6]. In particular, their use of digital platforms for peer emotional sharing helped them manage intensifying concerns and intense negative feelings associated with the pandemic [7]. There is accumulating evidence underscoring the importance of peers in providing remote emotional support for emerging adults that bears on their psychological functioning [8-10]. Consequently, there is burgeoning interest in understanding digital peer emotional disclosure and support as an intervention, evaluating relevant clinical outcomes, and assessing the associated mechanisms of change [11]. To this end, this randomized controlled trial (RCT) assessed the efficacy of digital peer emotional disclosure and support systems as well as the mechanism of change.

Digital Peer Support Intervention for Psychological Well-being

Overview: The Role of Emotional Disclosure

According to the social sharing of emotions framework, emotional disclosure refers to the interpersonal communication of important and personal experiences with mild to strong positivity and negativity with one's close social network (eg, "I was exhilarated to receive a good grade on an exam" or "I was distressed being involved in a car accident") [12,13]. It is a psychological mechanism that facilitates emotional regulation and recovery [12,13]. Theoretically, emotional sharing has important psychoemotional and social implications [12,13]. Emotional sharing maintains one's exposure to the emotional content and establishes emotional closeness and bonds among sharing partners [12,13]. Specifically, sharing negative emotions helps individuals deal with mental rumination and intrusive thoughts and search for meanings to integrate these experiences into one's existing narrative. Negative emotions also stimulate social interactions to garner emotional support [12,13]. In contrast, positive emotional experiences facilitate goal achievement by providing positive feedback on successful experiences that can be integrated into one's knowledge repertoire. Positive emotional episodes represent opportunities for capitalization of positive emotions that enhance one's positive affect and social bonds [12,13].

The social sharing of emotions framework was originally developed to understand emotional sharing in the offline context, particularly face-to-face interactions [12], and Rime [13] has also extended the framework to the web-based context. Applying the social sharing of emotions framework, scholars have found that college students share emotions on web-based platforms and such sharing may have similar psychological implications as that documented in face-to-face interactions [14,15]. Owing to the affordances of digital platforms, such as reduced nonverbal cues and high accessibility (ie, content is easily available and shared across physical locations [16]), individuals are less apprehensive and able to share emotions with a large audience unconfined by locality. Furthermore, users can elicit immediate feedback and amplify experiences on digital platforms [16]. These features can create a climate of emotional bonding and help individuals manage emotional load, enhance social connections and support, and contribute to social integration [12,17]. An extensive body of work with college students has demonstrated that peers are the primary targets for emotional disclosure [18]. Thus, digital peer emotion sharing has the potential to increase emotion sharers' sense of social support and the reassurance they are seeking [12,17], especially among college students, at scale.

Effective Peer Support

For digital peer support to be effective in mitigating emerging adults' psychological distress, including anxiety and depression, existing literature underscores the role of 4 components [19-21]. They are (1) enhancing one's sense of mattering (the extent to which we are important to the surrounding world and people), (2) strengthening selfhood (one's sense of identity), (3) exploring compassion (the degree of sensitivity to one's and others' pain and distress, with a desire to alleviate that distress), and (4) cultivating mindfulness (paying attention to the present moment with intention and acceptance; refer to the *Methods* section for details). Research has revealed that cultivating mattering and selfhood through meaningful and intimate

relationships involving peer support has positive implications for young people's mental well-being [22,23]. For instance, Marshall and Tilton-Weaver [23] reported that mattering to one's close social networks, specifically peers, that is cultivated through interacting with and caring for them reduces symptoms of depression and anxiety.

As a result of the social implications of the global pandemic, there has been a corresponding surge in research to understand the role of compassion and mindfulness in relation to individuals' psychological well-being [20,24]. Recent compassion- and mindfulness-based web-based interventions to support college students' mental health in coping with COVID-19 conditions have demonstrated positive outcomes, including clinical efficacy [20]. Students reported significantly lower levels of stress and anxiety coupled with enhanced levels of self-compassion. These findings provide evidence of the interventional potential of these 4 components in digital peer support. Thus, building college students' capacity to deliver these components to provide effective support may enhance their peers' psychological well-being.

Our review of existing literature revealed a limited understanding of how peer emotional disclosure on digital platforms functions as an intervention in support of emerging adults' psychological well-being [25-27]. The administration of peer support in the offline context, specifically over face-to-face interactions, intervening in various psychological and health outcomes among college students is well documented [28]. Traditional (offline) peer support interventions have demonstrated consistent evidence for implementation and feasibility outcomes, user acceptability, and clinical effectiveness in mediating the psychological well-being of emerging adults, particularly college students [28]. However, with regard to assessing peer support interventions for the mental health of young people, it is important to consider the mechanism of change to elucidate whether and how the intervention affects the development of psychological symptoms—how they unfold over time among emerging adults.

With the onset of the pandemic, the increase in web-based social connections among college students has underscored digital peer emotional disclosure as an intervention that may address specific areas of well-being involving anxiety, depression, and suicidal ideation [28]. As a result, there is burgeoning interest and impetus among researchers, education stakeholders, and health care professionals in establishing the role of peer support on digital platforms and evaluating its clinical effectiveness in improving the psychological well-being of emerging adults, particularly college students [26,27]. Establishing a change model provides the basis for building the therapeutic potential of digital peer support that harnesses the 4 components of emerging adult mental well-being to reduce psychological symptoms. Such findings may provide actionable knowledge for timely and relevant digital peer support for emerging adults' psychological well-being.

This Study

Overview

Following the published protocol [29], this RCT examined the effectiveness of a digital peer support intervention for the psychological well-being of emerging adults by establishing evidence for clinical outcomes (refer to the study by Yeo et al [29] for details of the trial). In brief, this RCT validated the applicability of harnessing the mattering, selfhood, compassion, and mindfulness components in a digital peer support intervention to promote emerging adults' mental well-being. By evaluating the mechanism of change involving digital peer support and emerging adults' psychological functioning, this RCT provided insights into whether or how digital peer support intervened in the development of psychological symptoms. These findings may provide actionable knowledge for timely digital peer support for intervening effectively in emerging adults' psychological functioning.

Aims, Research Questions, and Hypotheses

This study had 2 primary aims. First, we evaluated the digital peer support intervention in terms of training effectiveness for emerging adults providing support (befrienders), who received training in digital peer support skills from a clinical psychologist and certified counselors (moderators). Second, we investigated the clinical effectiveness for the emerging adults who anonymously share emotional experiences to seek support (seekers)—whether it enhanced the 4 components of psychological well-being, specifically, mattering, selfhood, compassion, and mindfulness, and improved the psychological well-being of emerging adults—and the associated mechanism of change linking the digital peer support intervention to emerging adult psychological well-being. As a secondary aim, we also examined the implementation outcomes in terms of feasibility and acceptability of the intervention in offering an ongoing mechanism of support and in identifying individuals with a high risk of having a mental health condition (refer to Multimedia Appendix 1).

To address the first aim, we examined the research question and hypothesis outlined in Textbox 1.

For the second aim, we investigated the research questions and hypotheses outlined in Textbox 2.

Textbox 1. Research question and hypothesis for the first aim of this study.

- Research question 1
  - Is the digital peer support training effective in building befrienders' capacity to provide effective peer support that harnesses the interventional potential of the 4 components of psychological well-being (mattering, selfhood, compassion, and mindfulness) peer support?
- Hypothesis 1
  - The digital peer support training is effective, as shown by the befrienders' adoption of and fidelity to the training curriculum that applies the 4 components of psychological well-being in providing effective peer support.

Textbox 2. Research questions and hypotheses for the second aim of this study.

- Research question 2a
  - Does the digital peer support intervention enhance the 4 components of psychological well-being among seekers?
- Hypothesis 2a
  - The digital peer support intervention increases the 4 components of psychological well-being among seekers over the course of the study across 4 time points.
- Research question 2b
  - Is the digital peer support intervention effective in improving emerging adult psychological well-being?
- Hypothesis 2b
  - The digital peer support intervention leads to significantly greater psychological well-being of seekers in the intervention group, as indexed by their lower symptoms of anxiety and depression compared with the waitlist control group. This effect is sustained beyond the period of the intervention at follow-up assessments on all seekers (carryover effects).
- Research question 2c
  - What is the mechanism explaining the change in seekers' psychological well-being in relation to digital peer support?
- Hypothesis 2c
  - The initial level and rate of change (ie, growth factors) of befrienders' online support predict positively the growth factors of seekers' psychological well-being.

Methods

Trial Design

This trial is registered with the US National Library of Medicine ClinicalTrials.gov (NCT05083676). This prospective interventional study evaluated both the implementation of the digital peer support intervention (eg, fidelity, adoption, and utility) and the clinical outcomes (eg, anxiety and depressive symptoms) via an RCT. The seekers (n=50), befrienders (n=30), and moderators (n=2) in arm 1 engaged in the digital peer support intervention (Acceset platform) for 3 weeks, and the control group (n=50) in arm 2 was wait-listed for the intervention. The 4–time-point questionnaire battery on psychological well-being for arm 1 was compared with that for arm 2.

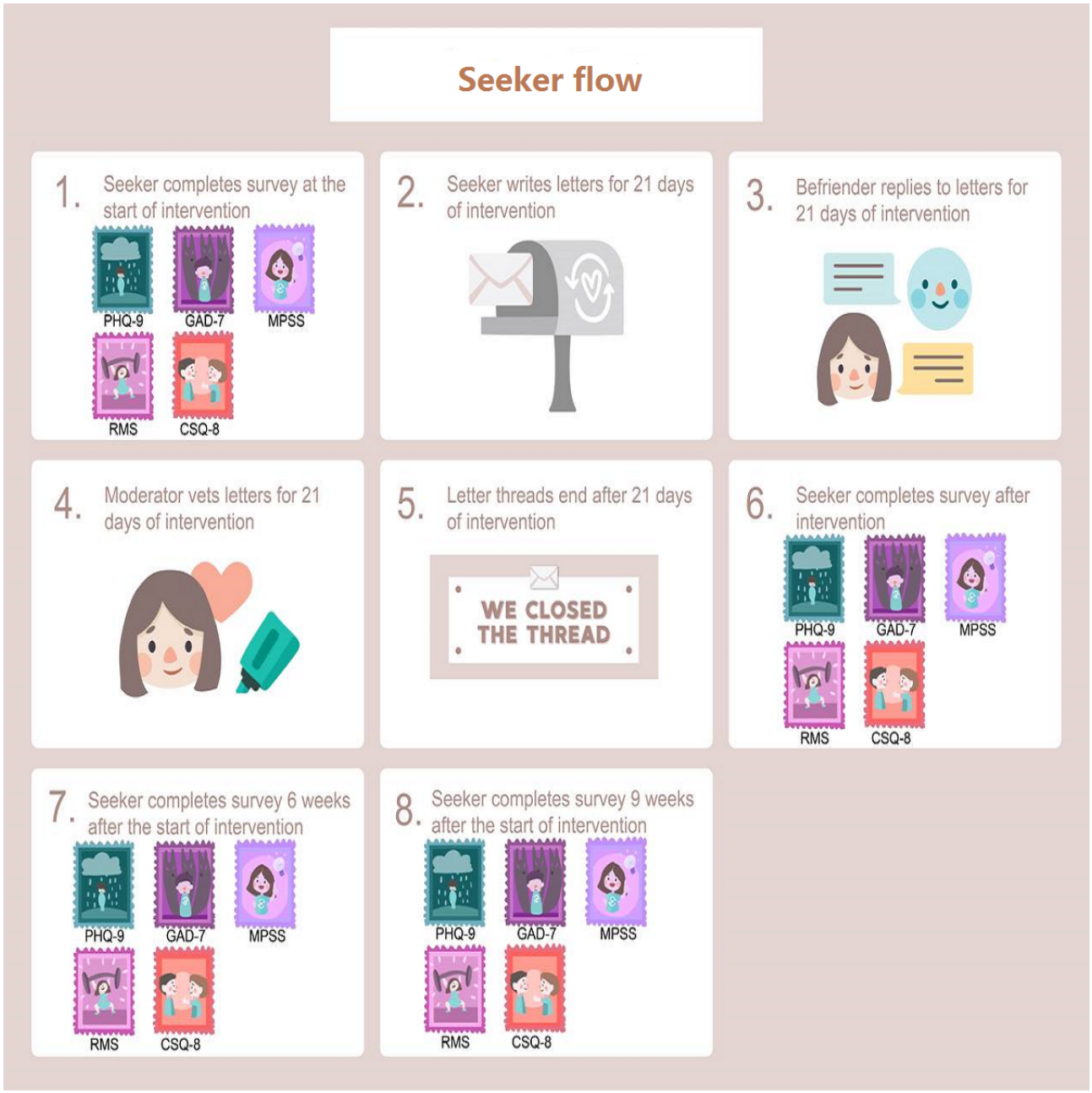
Ethics Approval

The study protocol received National University of Singapore (NUS) Institutional Ethics Review Board (IRB) approval in October 2021 (NUS-IRB reference S-20-144).

Intervention

In this study, we engaged Acceset—a stand-alone digital peer support platform that uses a digital text-based intervention involving peer disclosure for emerging adult mental well-being. Acceset incorporates characteristics of successful interventions that affect the psychological well-being of emerging adults [30,31]. Through Acceset, users (seekers) can anonymously share their emotional experiences and receive support from their peers (befrienders). These peers receive training in digital peer support skills from a clinical psychologist and certified counselors (moderators). Acceset digital peer support training for this trial intended to harness mattering, selfhood, compassion, and mindfulness—4 components that reduce emerging adults' anxiety and depression [19-21]. The Acceset platform incorporates a collection of digital features such as emotion stamps, motivation Graphics Interchange Format emotionality, and functional adjustment stickers as markers of psychological well-being, including emotionality (ie, positivity and negativity), motivations, and functional adjustment (ie, internalizing and externalizing behaviors), respectively. During the course of the intervention, user engagement with these features on the platform served as a source of self-reported information on their psychological well-being status. The Acceset text-based peer disclosure process begins when seekers engage with the platform to seek support with managing their emotional experiences. Details on the Acceset digital peer support training, the platform, and the seeker-befriender interaction (peer support workflow; Figure 1) are provided in the protocol that outlined the RCT [29].

Figure 1. Diagrammatic flow or procedure of the intervention. CSQ-8: 8-item Client Satisfaction Questionnaire; GAD-7: 7-item Generalized Anxiety Disorder Scale; MPSS: Multidimensional Scale of Perceived Social Support; PHQ-9: 9-item Patient Health Questionnaire; RMS: Rapid Mood Screener.



Sample

Befrienders

Participants comprised emerging adults from NUS [29] (mean age 19.60, SD 0.63 years; 17/30, 57% female). The ethnic composition of the sample (23/30, 77% Chinese; 4/30, 13% Malay; 2/30, 7% Indian; and 1/30, 3% others) approximated the Singapore population (74%, 13%, and 12%, respectively [32]). On the basis of participant-reported monthly family income, most participants were from low-middle–income families (<SGD 3000 [US $2224.28]; 13/30, 43%), and the rest were from middle-high–income (>SGD 4000 [US $2965.71]; 10/30, 33%) and middle-income (SGD 3000 to SGD 4000 [US $2224.28-$2965.71]; 7/30, 23%) families. Almost two-thirds of the participants (20/30, 67%) were first-generation college students, and approximately half of them (16/30, 53%) resided at home with parents, with the rest living in campus housing (13/30, 43%) or having an alternative housing arrangement such as an off-campus rental apartment (1/30, 3%).

Seekers in the Treatment Arm

Participants (n=291; mean age 19.44, SD 1.28 years; 16/30, 53% female; 24/30, 80% Chinese; 3/30, 10% Malay; 2/30, 7% Indian; and 1/30, 3% others) were emerging adults from NUS [29]. The ethnic composition of the sample was similar to the Singapore population [32]. On the basis of participant-reported monthly family income, most of them were from middle-high–income families (>SGD 4000 [US $2965.71]; 16/30, 53%), and the rest were from middle-income (SGD 3000 to SGD 4000 [US $2224.28-$2965.71]; 8/30, 27%) and low-middle–income (<SGD 3000 [US $2224.28]; 6/30, 20%) families. Slightly more than half (17/30, 57%) of the participants

were first-generation college students, and a similar number resided at home with parents (17/30, 57%), with the rest living in campus housing (11/30, 37%) or having an alternative housing arrangement such as an off-campus rental apartment (2/30, 7%).

Seekers in the Control Arm

Participants (mean age 19.43, SD 0.66 years; 14/30, 47% female) were emerging adults from NUS [29]. A total of 80% (24/30) of them were Chinese, 7% (2/30) were Malay, 7% (2/30) were Indian, and 7% (2/30) were of other races. These figures were comparable with the ethnic composition of the Singapore population [32]. In total, 63% (19/30) of the participants were first-generation college students. Most of them (26/30, 87%) were living in campus housing, with the rest residing at home with parents (2/30, 7%) or having an alternative housing arrangement such as an off-campus rental apartment (2/30, 7%). On the basis of participant-reported monthly family income, most of them were from middle-income families (>SGD 4000 [US $2965.71]; 14/30, 47%), and the rest were from low-middle– (SGD 3000 to SGD 4000 [US $2224.28-$2965.71]; 7/30, 23%) and middle-high–income (>SGD 4000 [US $2965.71]; 9/30, 30%) families.

The following details are listed in the published protocol that outlined the trial [29]: (1) study setting, (2) eligibility criteria, and (3) peer support workflow.

Measures

The 4 Components of Psychological Well-being (Mattering, Selfhood, Compassion, and Mindfulness)

For each letter exchange between befrienders and seekers, 2 undergraduate assistants extracted and coded the letter content on 4 components. A total of 192 letter exchanges over the 3-week intervention were extracted and coded. On average, 4.67 letters (SD 1.43) were exchanged per day. These letter exchanges were deidentified by the first author and assigned to a team of 2 undergraduate assistants who were trained on the coding procedure outlined in the study by Krippendorff [33] over 1 month until they consistently exceeded adequate reliability standards (κ>0.6 on all variables of interest; refer to Table S1 in Multimedia Appendix 1 for the coding process and samples of participants' extracted letter exchanges). All codes presented in the following paragraphs were coded independently. The κ coefficients ranged from 0.70 to 0.85, which indicated good interrater reliability.

To determine mattering, coders used the Rosenberg Mattering Scale with 5 items ([34]; eg, "how important are you to others?" and "How interested are others in what you have to say?") to assess the degree to which befrienders' responses indicated how important seekers are to others and the degree to which seekers' responses indicated how important they are to others. To assess selfhood, coders rated the extent to which befrienders' responses built seekers' self-knowledge, interpersonal self, and self-agency using the Rosenberg Global Self-Esteem Scale ([35]; eg, "the peer is a person of worth, at least on an equal plane with others" and "the peer has a number of good qualities"), the Self-Consciousness Scale ([36]; eg, "the peer is constantly examining his/her motives" and "the peer is concerned about what other people think of him/her"), the General Self-Efficacy Scale ([37]; eg, "When the peer makes plans, he/she is certain that he/she can make them work" and "the peer is a self-reliant person"), and the Social Self-Efficacy Scale ([38]; eg, "The peer does not handle himself/herself well in social gatherings" and "It is difficult for the peer to make new friends"). For seekers, coders rated the extent to which their emotional disclosure reflected self-knowledge, interpersonal self, and self-agency using the same subscales and items except with a different item stem (eg, "he/she is a person of worth, at least on an equal plane with others" and "he/she has a number of good qualities"). A composite selfhood score was obtained by averaging the ratings across all 4 selfhood scales for both befrienders and seekers.

The compassion of befrienders' peer support responses and seekers' peer emotional disclosure was assessed by coders using the Compassion Cultivation Training program [37] as reflecting (1) an awareness of distress, (2) affective concern, and (3) a responsiveness or readiness to help relieve that distress (motivational). Mindfulness was measured as a flexible cognitive state in which individuals are actively present and notice novel aspects in both the environment and one's perspectives [39]. Befrienders' support responses and seekers' peer emotional disclosure were evaluated based on the degree to which they harnessed mindful self-acceptance: (1) identifying novel aspects of the situation or perspective; (2) demonstrating "work in progress" by using possibility words such as "could be" and offering other interpretations of the situation; (3) highlighting puzzles and paradoxes, for example, how individuals may feel victimized yet they are responsible for being in that situation; (4) noticing humorous aspects of the situation; (5) perceiving the situation from multiple perspectives; (6) considering alternative (useful) aspects of a problematic context or the silver lining; (7) emphasizing a mental file of positive memories; and (8) encouraging mindfulness journaling.

To assess mattering, selfhood, compassion, and mindfulness, a 4-point scale ranging from 1 (not at all) to 4 (a lot) was used by raters. On the basis of raters' scores, the Cronbach α for the befrienders' mattering was .86 for the first letter exchange, .84 for the second letter exchange, .87 for the third letter exchange, and .91 for the fourth letter exchange. For the seekers' mattering, the Cronbach α was .83, .82, .77, and .74 for the first, second, third, and fourth letter exchanges, respectively. The Cronbach α for the befrienders' selfhood was .77 for the first letter exchange, .76 for the second letter exchange, .71 for the third letter exchange, and .70 for the fourth letter exchange. For the seekers' selfhood, the Cronbach α was .85, .88, .89, and .90 for the first, second, third, and fourth letter exchanges, respectively.

The Cronbach α for the befrienders' compassion was .88 for the first letter exchange, .71 for the second letter exchange, .74 for the third letter exchange, and .75 for the fourth letter exchange. For the seekers' compassion, the Cronbach α was .88, .91, .90, and .91 for the first, second, third, and fourth letter exchanges, respectively. The Cronbach α for the befrienders' mindfulness was .76 for the first letter exchange, .80 for the second letter exchange, .73 for the third letter exchange, and .70 for the fourth letter exchange. For the seekers' mindfulness, the Cronbach α was .84, .88, .89, and .88 for the first, second,

third, and fourth letter exchanges, respectively. The Cronbach α demonstrated adequate internal consistency in this study, similar to that established in other Asian contexts such as China—mattering (Cronbach α=.69 [40]), selfhood (self-esteem: Cronbach α=.85; self-efficacy: Cronbach α=.82; self-consciousness: Cronbach α=.6 [41,42]), compassion (Cronbach α=.88 [43]), and mindfulness (Cronbach α=.85 [43]).

Befrienders' Peer Support Responses

We tested 1-factor versus 4-factor models for befrienders' support (with the 4 components of psychological well-being as indicators measuring a latent construct). Results indicated acceptable to excellent fit of the 1-factor model, which was a better fit than the 4-factor model, across the 4 time points (refer to Table S2 in Multimedia Appendix 1). Thus, we created a composite peer support score with mattering, selfhood, compassion, and mindfulness scales as a set of metrics for assessing befrienders' peer support responses for each of the 4 letter exchanges, with higher scores representing greater peer support.

Psychological Well-being (Times 1-4)

We created a latent construct with 2 indicators—anxiety and depression. Befrienders and seekers provided self-report responses to the 7-item General Anxiety Disorder Scale (GAD-7), which provides rapid screening for the presence of clinically significant anxiety symptoms, particularly in outpatient settings ([44]; eg, "feeling nervous, anxious, or on edge" and "not being able to stop or control worrying"), and the 9-item Patient Health Questionnaire (PHQ-9), which is a 9-item scale that quantifies depressive symptoms and monitors their severity ([45]; eg, "little interest or pleasure in doing things" and "feeling down, depressed or hopeless"). All the items on the GAD-7 and PHQ-9 were rated on a 4-point scale ranging from 0 (not sure at all) to 4 (nearly every day) across 4 time points: time 1 (before the intervention), time 2 (postintervention assessment; 3 weeks), time 3 (first follow-up; 6 weeks), and time 4 (second and final follow-up; 9 weeks).

We tested a 1-factor versus a 2-factor model for the seekers' psychological well-being (with the GAD-7 and PHQ-9 items as indicators measuring a latent construct). Results indicated acceptable to excellent fit of the 1-factor model, which was a better fit than the 2-factor model, across the 4 time points (refer to Table S2 in Multimedia Appendix 1). Thus, we created a composite psychological well-being score using the GAD-7 and PHQ-9 scores as a metric for assessing psychological well-being, with lower scores representing greater well-being. The Cronbach α for seekers' psychological well-being was .88 at time 1 (before the intervention), .94 at time 2 (postintervention assessment; 3 weeks), .93 at time 3 (first follow-up; 6 weeks), and .95 at time 4 (second and final follow-up; 9 weeks). The GAD-7 and PHQ-9 have demonstrated adequate internal consistency in Asian contexts such as China (Cronbach α=.85 and .89, respectively [46]).

Perceived Social Support (Times 1-4)

Seekers responded to the 12-item Multidimensional Scale of Perceived Social Support, which assessed perceived social support from 3 sources—family, friends, and significant others ([47]; eg, "There is a special person who is around when I am in need" and "I get the emotional help and support I need from my family")—using a 7-point scale (1=very strongly disagree; 7=very strongly agree). The Cronbach α for seekers' perceived social support was .88 at time 1 (before the intervention), .90 at time 2 (postintervention assessment; 3 weeks), .94 at time 3 (first follow-up; 6 weeks), and .93 at time 4 (second and final follow-up; 9 weeks). The Multidimensional Scale of Perceived Social Support has demonstrated adequate internal consistency in Asian contexts such as China (Cronbach α=.86 [48]).

Data Management and Planned Analyses

Overview

Consistent with the procedure for data management and planned analyses outlined in the study by Yeo et al [29], Acceset provided a cloud storage for depositing data on participants' responses to all the measures as well as written letters between seekers and befrienders on the Acceset platform. Following the NUS-IRB protocol, each participant had a nonidentifying unique ID, and all data were deidentified before analyses, conducted by the lead author and 2 student interns. On the basis of a power analysis conducted before the commencement of the RCT, 50 participants each for the intervention and waitlist control groups were needed to find significant effects with the total number of outcomes examined in the RCT. Using a Monte Carlo stimulation [49,50], our power analysis on the mechanism of change indicated that 100 participants provided 80% power to detect significant linkages between support from befrienders and psychological well-being of seekers based on criteria listed in the studies by Muthén and Muthén [49,50]. Our power revealed that the range of biases of the parameter and SE (0.04% to 1.40%), SE (0.05% to 1.72%), and coverage (0.94 to 1.00) met the criteria for power to approximate or exceed 80% listed in the studies by Muthén and Muthén [49,50]. For instance, the parameter and SE biases do not exceed 10% for all parameters in the model. We performed all analyses using RStudio (Posit) [51].

Aim 1

To address research question 1 and hypothesis 1, a total of 2 student assistants conducted qualitative analyses of all letter exchanges between befrienders and seekers by extracting, coding, and evaluating the degree to which befrienders' support responses reflected the 4 components of psychological well-being—mattering, selfhood, compassion, and mindfulness. We then examined if befrienders' pre- and posttraining peer support responses were significantly different using the independent-sample *t* test (2-tailed) with Bonferroni post hoc tests to control for multiple comparisons.

Aim 2

For research question 2a and hypothesis 2a, we ascertained whether the intervention increased seekers' experience of the 4 components of psychological well-being by performing 4 sets of latent growth curve modeling to examine the changes in mattering, selfhood, compassion, and mindfulness. To model baseline mean levels and rates of change (measured as intercepts and slope factors, respectively), we used repeated-measure data comprising the 4 time points of a given composite score as

indicators of latent variables. In each of the 4 baseline latent growth models (ie, mattering, selfhood, compassion, and mindfulness), we estimated and compared the linear, quadratic, and nonlinear curve-fitting models. In all 4 models, the loadings of the intercept factor were constrained to 1 on each time point. In the linear model, the factor loadings for slope were set at 0, 1, 2, and 3; in the quadratic model, the factor loadings were set at 0, 1, 4, and 9; and in the nonlinear curve-fitting model, the factor loadings were set at 0 and 1 for time 1 and time 2, respectively, and freely estimated for times 3 and 4.

In evaluating each model, we examined the following fit indexes: the chi-square statistic, the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), and the comparative fit index (CFI). Acceptable model fit was indicated by an RMSEA of <0.08 (90% CI 0.05-0.10), an SRMR with values of ≤1.0, and a CFI of >0.90. RMSEA and SRMR values of <0.05 (90% CI 0.05-0.10) and CFI values of >0.95 were considered a good fit [52]. The linear model was nested within the quadratic and optimal-fit models, and the quadratic and optimal-fit models were not nested. Thus, the Bayesian information criterion (BIC) was used for model comparisons [53].

As for research question 2b and hypothesis 2b, we investigated whether the intervention affected seekers' psychological well-being using independent-sample *t* tests (2-tailed) with Bonferroni post hoc tests. First, seekers' experience of anxiety and depressive symptoms in the intervention group at postintervention assessment (week 3) was compared with the symptoms of the waitlist control group before the intervention (week 0). Second, we assessed the sustained effects of the intervention on the psychological well-being of all seekers (intervention and waitlist) by comparing baseline symptoms (week 0) with postintervention symptoms (week 3), as well as symptoms at each follow-up (weeks 6 and 9). Similarly, we assessed seekers' support-seeking behaviors by comparing their perceived social support at baseline (week 0) with that at the postintervention (week 3) and follow-up (weeks 6 and 9) assessments.

We addressed research question 2c and hypothesis 2c on the mechanism of change by evaluating a conditional growth model on befrienders' peer support predicting seekers' psychological well-being. Befrienders' sex was included as a covariate as female individuals are more likely to provide support than male individuals [54]. The seekers' sex and socioeconomic status (ie, monthly household income) were included as covariates as female students and those of low socioeconomic status are more likely to experience anxiety and depression [54]. Following the procedure outlined in the study by Kaplan [53], a random sample of 20 participants was used to plot and explore separately the variability in the level and trend of befrienders' support and seekers' psychological well-being in college. We then examined the trajectories of befrienders' support and seekers' well-being by fitting individual baseline growth models following the same procedure outlined previously on assessing the trajectories of the 4 components of psychological well-being among seekers. These steps were necessary to prevent problems with specification (ie, negative residual variances) when modeling growths in digital peer support and psychological well-being simultaneously.

Results

Preliminary Analyses

The nonsignificant result from the Little Missing Completely at Random test showed that the missing data involving all the variables of interest in this study (refer to the *Measures* section), which ranged from 19% to 58%, were nonsystematic (χ^2_{3783}=766.0; *P*>.99). Thus, we handled missing data using full information maximum likelihood imputation [53]. Maximum likelihood imputation is a method that ascertains the parameter values of a model using mean and variance; it maximizes the chance that the values generated are closest to those observed.

We used multiple group confirmatory factor analyses to establish sequentially the different levels of measurement invariance of the 1-factor befriender support and 1-factor seeker psychological well-being models across 4 time points. Results revealed configural invariance (in factorial structure), metric invariance (in factor loadings), and scalar invariance (in item intercepts) with acceptable model fit indexes (RMSEA<0.08; SRMR<0.10; CFI>0.90). Comparing constrained versus unconstrained measurement models, the differences in goodness-of-fit indexes were below the recommended cutoff value of CFI=0.01 [55]. These results confirmed the longitudinal invariance of the 1-factor model for both befrienders' support and seekers' psychological well-being for meaningful comparisons of scores across time.

Tables 1 and 2 present the descriptive statistics and correlations among variables for each letter exchange for seekers and befrienders, respectively. Of the 4 components of psychological well-being, mattering and selfhood displayed consistently moderate and positive associations, and similar associations were documented between compassion and mindfulness for both befrienders and seekers. For seekers, perceived social support had low to moderate negative associations with psychological symptoms at each time point (eg, support and well-being at time 1 [before the intervention]) and across time (support at time 1 and well-being at time 3 [first follow-up at 6 weeks]; refer to Table S3 in Multimedia Appendix 1).

Table 1. Correlations and descriptive statistics for seekers on study variables for letter exchanges over the 3-week intervention.[a]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Mean (SD)
1. Ma[b] 1[c]	—[d]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3.12 (1.09)
2. Ma 2[e]	0.26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2.86 (0.87)
3. Ma 3[f]	0.43	0.53	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2.58 (1.02)
4. Ma 4[g]	0.60	0.58	0.78	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2.56 (0.99)
5. S[h] 1	0.18	0.12	0.11	0.10	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2.47 (1.29)
6. S 2	0.21	0.06	0.18	0.22	0.17	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3.07 (1.76)
7. S 3	0.27	0.13	0.10	0.22	0.14	0.38	—	—	—	—	—	—	—	—	—	—	—	—	—	3.39 (1.05)
8. S 4	0.27	0.27	0.28	0.18	0.21	0.58	0.41	—	—	—	—	—	—	—	—	—	—	—	—	3.63 (1.77)
9. C[i] 1	0.15	0.26	0.14	0.15	0.28	0.13	0.10	0.24	—	—	—	—	—	—	—	—	—	—	—	2.63 (1.02)
10. C 2	0.10	0.21	0.28	0.32	0.18	0.15	0.11	0.12	0.44	—	—	—	—	—	—	—	—	—	—	3.53 (1.24)
11. C 3	0.15	0.19	0.26	0.33	0.16	0.13	0.22	0.20	0.35	0.70	—	—	—	—	—	—	—	—	—	3.53 (0.75)
12. C 4	0.26	0.29	0.16	0.46	0.29	0.10	0.14	0.24	0.44	0.78	0.70	—	—	—	—	—	—	—	—	3.84 (0.81)
13. M[j] 1	0.19	0.11	0.16	0.18	0.19	0.16	0.16	0.14	0.16	0.18	0.17	0.22	—	—	—	—	—	—	—	3.30 (1.00)
14. M 2	0.18	0.19	0.11	0.17	0.12	0.13	0.18	0.14	0.18	0.30	0.17	0.13	0.25	—	—	—	—	—	—	3.16 (1.90)
15. M 3	0.11	0.15	0.16	0.17	0.14	0.15	0.19	0.21	0.13	0.15	0.39	0.17	0.45	0.60	—	—	—	—	—	2.91 (1.35)
16. M 4	0.20	0.24	0.17	0.13	0.27	0.21	0.12	0.17	0.26	0.11	0.35	0.30	0.56	0.56	0.64	—	—	—	—	2.31 (1.18)
17. PWB[k] 1	0.15	0.12	0.34	0.15	0.21	0.11	0.17	0.38	0.12	0.10	0.19	0.16	0.27	0.14	0.20	0.28	—	—	—	10.72 (1.06)
18. PWB 2	0.05	0.38	0.24	0.15	0.24	0.45	0.23	0.20	0.17	0.25	0.26	0.19	0.20	0.21	0.19	0.11	0.52	—	—	9.68 (1.08)
19. PWB 3	0.17	0.33	0.42	0.28	0.32	0.13	0.15	0.14	0.18	0.18	0.11	0.18	0.31	0.17	0.13	0.15	0.54	0.76	—	8.41 (1.12)
20. PWB 4	0.08	0.29	0.36	0.26	0.10	0.33	0.23	0.13	0.13	0.28	0.10	0.13	0.34	0.12	0.25	0.11	0.52	0.69	0.78	8.35 (1.16)

[a]Coefficients above the diagonal refer to those for befrienders, and coefficients below the diagonal refer to those for seekers. All correlation coefficients >0.15 are significant.

[b]Ma: mattering.

[c]1: first letter exchange.

[d]Not applicable.

[e]2: second letter exchange.

[f]3: third letter exchange.

[g]4: fourth letter exchange.

[h]S: selfhood.

[i]C: compassion.

[j]M: mindfulness.

[k]PWB: psychological well-being.

Table 2. Correlations and descriptive statistics for befrienders on study variables for letter exchanges over the 3-week intervention.[a]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Mean (SD)
1. Ma[b] 1[c]	—[d]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3.14 (0.17)
2. Ma 2[e]	0.74	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2.91 (0.83)
3. Ma 3[f]	0.46	0.80	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2.59 (0.60)
4. Ma 4[g]	0.89	0.87	0.83	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2.20 (0.52)
5. S[h] 1	0.56	0.66	0.35	0.57	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2.74 (1.87)
6. S 2	0.30	0.34	0.34	0.34	0.72	—	—	—	—	—	—	—	—	—	—	—	—	—	—	3.43 (1.87)
7. S 3	0.24	0.20	0.50	0.50	0.51	0.26	—	—	—	—	—	—	—	—	—	—	—	—	—	3.68 (1.17)
8. S 4	0.13	0.38	0.14	0.17	0.31	0.35	0.61	—	—	—	—	—	—	—	—	—	—	—	—	3.87 (0.32)
9. C[i] 1	0.34	0.30	0.27	0.36	0.48	0.16	0.34	0.53	—	—	—	—	—	—	—	—	—	—	—	1.29 (1.97)
10. C 2	0.19	0.66	0.30	0.10	0.37	0.45	0.39	0.29	0.82	—	—	—	—	—	—	—	—	—	—	2.27 (1.20)
11. C 3	0.27	0.10	0.20	0.51	0.21	0.24	0.35	0.35	0.66	0.45	—	—	—	—	—	—	—	—	—	2.49 (0.53)
12. C 4	0.52	0.42	0.16	0.40	0.19	0.20	0.13	0.13	0.40	0.54	0.73	—	—	—	—	—	—	—	—	2.36 (1.13)
13. M[j] 1	0.35	0.13	0.20	0.30	0.23	0.57	0.11	0.11	0.19	0.27	0.23	0.16	—	—	—	—	—	—	—	1.51 (0.63)
14. M 2	0.24	0.42	0.30	0.49	0.60	0.50	0.15	0.15	0.16	0.47	0.29	0.14	0.85	—	—	—	—	—	—	2.87 (1.59)
15. M 3	0.17	0.50	0.35	0.30	0.49	0.35	0.14	0.14	0.15	0.24	0.21	0.17	0.55	0.63	—	—	—	—	—	3.17 (1.18)
16. M 4	0.18	0.21	0.16	0.37	0.35	0.35	0.17	0.17	0.18	0.28	0.11	0.18	0.58	0.60	0.53	—	—	—	—	3.24 (1.61)
17. PWB[k] 1	0.12	0.23	0.57	0.15	0.31	0.30	0.48	0.48	0.40	0.68	0.25	0.56	0.27	0.28	0.22	0.41	—	—	—	2.61 (1.33)
18. PWB 2	0.21	0.18	0.54	0.27	0.33	0.34	0.41	0.41	0.55	0.30	0.30	0.39	0.38	0.17	0.38	0.44	0.54	—	—	2.92 (0.74)
19. PWB 3	0.11	0.43	0.57	0.35	0.36	0.15	0.46	0.46	0.76	0.55	0.25	0.56	0.19	0.28	0.23	0.39	0.76	0.79	—	3.25 (1.52)
20. PWB 4	0.10	0.20	0.24	0.10	0.35	0.14	0.48	0.48	0.49	0.34	0.38	0.57	0.13	0.20	0.53	0.49	0.53	0.38	0.66	3.60 (1.62)

[a]All correlation coefficients >0.15 are significant.

[b]Ma: mattering.

[c]1: first letter exchange.

[d]Not applicable.

[e]2: second letter exchange.

[f]3: third letter exchange.

[g]4: fourth letter exchange.

[h]S: selfhood.

[i]C: compassion.

[j]M: mindfulness.

[k]PWB: psychological well-being.

Main Analyses

Aim 1

To address research question 1 and hypothesis 1, we examined if Acceset digital peer support training was effective in training befrienders to provide effective peer support by assessing the adoption of and fidelity to the training curriculum through the letter exchanges between seekers and befrienders. Specifically, we evaluated the text of the letter exchanges in terms of the extent to which befrienders displayed fidelity in applying the 4 components of psychological well-being—specifically, mattering, selfhood, compassion, and mindfulness—in their peer responses to seekers' disclosures. To control for multiple comparisons, we applied the Bonferroni post hoc tests. Befrienders' peer responses demonstrated greater selfhood (support) after training (mean 62.83, SE 10.18) than before training (mean 54.86, SE 7.32; t_{34}=3.88; P<.001). However, their support responses did not indicate greater mattering after training (mean 9.03, SE 3.87) than before training (mean 9.34, SE 3.87; t_{34}=0.40; P=.69). The same outcomes were observed for compassion after training (mean 8.57, SE 3.20) and before training (mean 9.09, SE 3.97; t_{34}=0.40; P=.50) and mindfulness after training (mean 10.00, SE 2.63) and before training (mean 11.17, SE 3.59; t_{34}=1.64; P=.11).

Aim 2

Research Question 2a and Hypothesis 2a

To address research question 2a and hypothesis 2a, we examined if the Acceset peer support training of befrienders enhanced the 4 components of psychological well-being among seekers. We compared the change in mattering, selfhood, compassion, and mindfulness scores of seekers over the course of the study. We conducted 4 sets of latent growth curve modeling to examine the trajectories of mattering, selfhood, compassion, and mindfulness by fitting individual baseline growth models. Table 3 presents the parameter estimates and fit statistics for the latent growth curve models for the 4 components of psychological well-being. For each component, we found a best-fit model. The chi-square difference tests, BICs, and the various alternative fit indexes indicated that alternative models were poorer fits. Figure 2 shows the models examined and the best-fit models for each of the 4 components of psychological well-being.

Table 3. Latent growth models for the 4 components of psychological well-being: fit indexes and estimates.

Component and model	Goodness of fit					Estimates	
	RMSEA[a] (90% CI)	CFI[b]	SRMR[c]	BIC[d]	Chi-square (*df*)	Intercept β[e] (SE)	Slope (SE)
Mattering							
L[f,g]	*0.00 (0.00-0.05)*	*1.00*	*0.04*	*1284.69*	*1.45 (5)*	*3.14*[h] *(0.27)*	*−0.48 (0.13)*
Q[i]	0.02 (0.01-0.05)	0.98	0.08	1288.67	0.91 (4)	2.42[h] (0.30)	−0.46 (0.42)
NL[j]	0.04 (0.00-0.09)	0.96	0.09	1292.76	0.39 (3)	3.03[h] (0.28)	0.13 (0.22)
Selfhood							
L	*0.03 (0.00-0.05)*	*0.99*	*0.08*	*836.88*	*5.48 (5)*	*2.74*[h] *(0.12)*	*0.42*[h] *(0.08)*
Q	0.04 (0.00-0.16)	0.98	0.09	869.02	4.60 (4)	2.77[h] (0.12)	0.59[h] (0.20)
NL	0.09 (0.00-0.17)	0.90	0.08	871.90	2.87 (3)	2.79[h] (0.12)	0.60[h] (0.16)
Compassion							
L	0.22 (0.15-0.30)	0.67	0.44	1748.36	28.74[h] (5)	1.57[h] (0.64)	0.49[h] (0.66)
Q	0.11 (0.00-0.21)	0.92	0.15	1733.34	9.11[k] (4)	1.44[h] (0.61)	0.45[h] (1.13)
NL	*0.00 (0.00-0.04)*	*1.00*	*0.06*	*1730.62*	*1.79 (3)*	*1.29*[h] *(0.01)*	*0.61*[h] *(0.91)*
Mindfulness							
L	0.29 (0.22-0.37)	0.00	0.30	1669.95	47.16[h] (5)	1.49[h] (0.96)	0.33[h] (0.62)
Q	0.17 (0.08-0.26)	0.00	0.20	1642.34	14.95[h] (4)	1.30[h] (0.80)	0.23[h] (0.10)
NL	*0.06 (0.00-0.09)*	*0.92*	*0.06*	*1605.62*	*7.78 (3)*	*1.51*[h] *(0.60)*	*0.87*[h] *(0.93)*

[a]RMSEA: root mean square error of approximation.

[b]CFI: comparative fit index.

[c]SRMR: standardized root mean square residual.

[d]BIC: Bayesian information criterion.

[e]Standardized estimate.

[f]L: linear.

[g]Italics indicate the best-fit models.

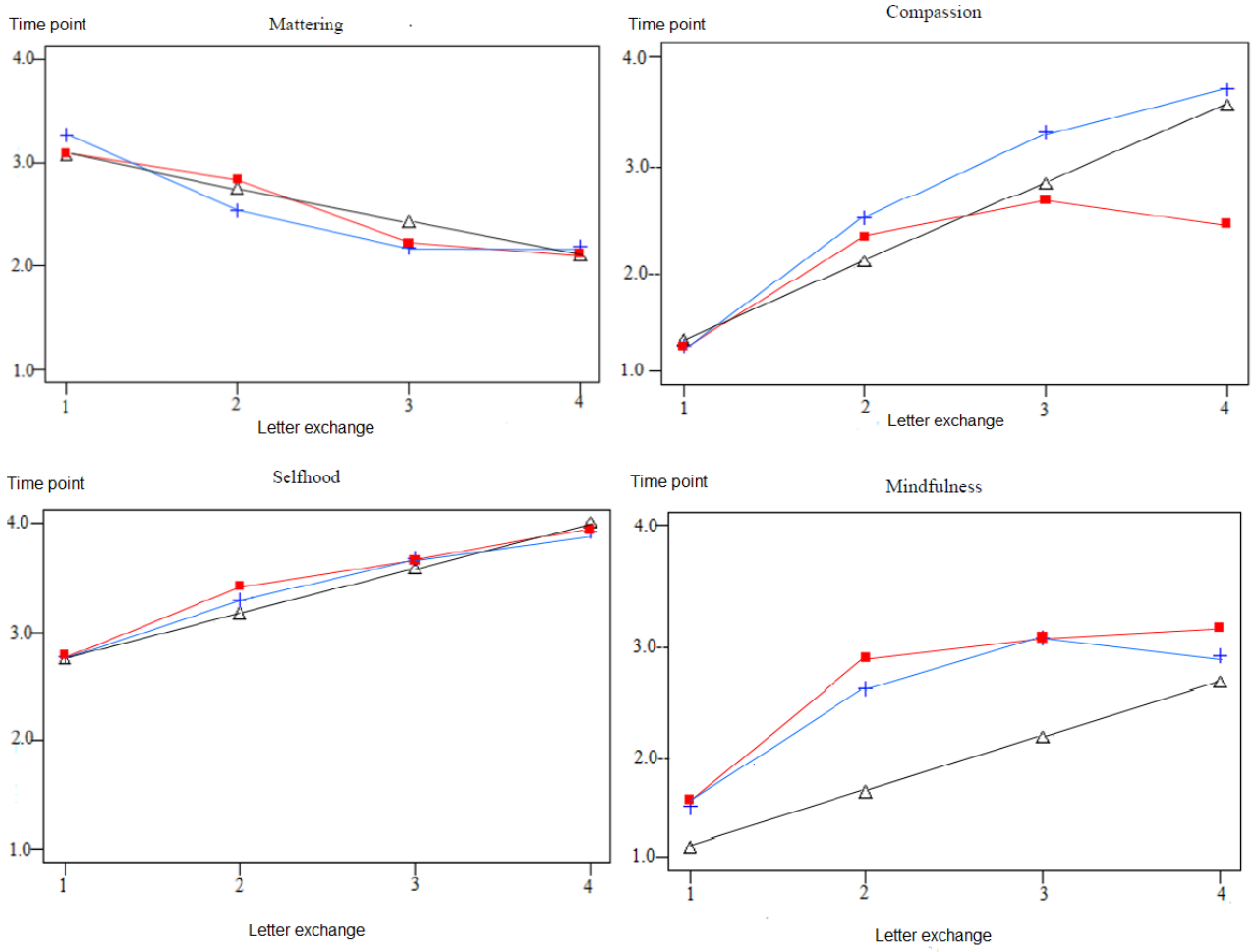
[h]*P*<.01.

[i]Q: quadratic.

[j]NL: nonlinear curve.

[k]*P*<.05.

Figure 2. Latent growth curve models for the 4 components of psychological well-being among seekers. For mattering and selfhood, the linear baseline model was the best-fitting model (in black); for compassion and mindfulness, the nonlinear curve-fitting baseline model was the best-fitting model (in red).



Mattering

The linear model demonstrated excellent fit, but rate of change was not significant. The initial level and rate of change in mattering did not indicate significant association (r_{100}=0.17; *P*=.12). Seekers' sense of mattering demonstrated limited change at 0 (baseline), 3 (postintervention assessment), 6 (follow-up assessment), and 9 (follow-up assessment) weeks.

Selfhood

The linear model demonstrated excellent fit. The moderate initial level of selfhood increased linearly from baseline to the final follow-up assessment, and the initial level was unrelated to its rate of change (r_{100}=0.03; *P*=.95). Seekers had moderate level of selfhood development before the intervention, and the increase in selfhood at postintervention assessment was sustained at the follow-up assessments at 6 and 9 weeks.

Compassion

The optimal curve-fitting model revealed excellent fit. The low initial level of compassion had a sharp increase from baseline to postintervention assessment, which increased modestly at subsequent follow-up assessments. The initial level of mindfulness was not significantly related to its rate of change (r_{100}=0.27; *P*=.21). Before the Acceset peer support intervention, seekers indicated a low level of compassion that increased substantially following the intervention and modestly at the 6- and 9-week follow-ups.

Mindfulness

The optimal curve-fitting model demonstrated excellent fit and revealed a low initial level with sharp increase following the intervention and continued to increase subsequently but dipped slightly at the last follow-up assessment. The initial level of mindfulness was not significantly related to its rate of change (r_{100}=0.33; *P*=.56). Seekers experienced a pronounced increase in mindfulness from baseline to postintervention assessment, and this continued to increase at the 6-week follow-up. However, mindfulness decreased slightly at 9 weeks—the final follow-up.

Research Question 2b and Hypothesis 2b

As for research question 2b and hypothesis 2b, congruent with our hypothesis, engagement with Acceset digital peer support led to improved mental well-being of seekers in the intervention group, particularly lower psychological symptoms (postintervention assessment; mean 7.15, SE 0.88; Table 4), as compared with the waitlist control group (before the intervention; mean 11.75, SE 0.89; t_{89}=3.44; *P*<.001). We assessed the sustained effect of the digital peer support intervention beyond the period of the intervention by evaluating

the change in mental well-being of the participants in both groups (intervention and control) after 3, 6, and 9 weeks from the baseline using self-report questionnaires. In conducting independent-sample *t* tests, we controlled for multiple comparisons using Bonferroni post hoc tests. Seekers experienced less psychological symptoms at postintervention assessment (ie, 3 weeks; mean 9.49, SE 1.16; t_{46}=3.14; *P*=.003) and at the first follow-up (ie, 6 weeks; mean 8.45, SE 1.30; t_{38}=3.20; *P*=.003), but not at the second follow-up (ie, 9 weeks; mean 10.26, SE 1.51; t_{38}=1.74; *P*=.09), than they did at baseline (mean 12.06, SE 1.20). Seekers' support-seeking behaviors beyond the Acceset platform were assessed through their perceived social support at weeks 6 and 9. Compared with baseline (mean 63.38, SE 1.50), seekers' perceived social support did not differ significantly at 3 weeks (mean 63.92, SE 1.35; t_{62}=0.43; *P*=.67), 6 weeks (mean 62.25, SE 2.09; t_{54}=0.44; *P*=.66), and 9 weeks (mean 62.60, SE 2.38; t_{42}=0.46; *P*=.65).

Table 4. Descriptive statistics for support and psychological well-being among seekers in the treatment and control arms at baseline, postintervention assessment, and follow-ups.

	Baseline[a], mean (SD)	Postintervention assessment[b], mean (SD)	First follow-up[c], mean (SD)	Second follow-up[d], mean (SD)
Treatment				
PSS[e]	63.47 (9.71)	66.06 (7.92)	61.35 (9.15)	60.60 (11.20)
PWB[f]	10.16 (3.22)	7.15 (0.88)	9.71 (2.99)	10.18 (3.45)
Control				
PSS	62.39 (8.99)	61.78 (8.33)	63.15 (9.04)	64.60 (10.31)
PWB	11.75 (2.97)	8.13 (1.15)	7.19 (1.89)	10.34 (3.99)

[a]First time point (before the intervention; 0 weeks).

[b]Second time point (3 weeks).

[c]Third time point (6 weeks).

[d]Fourth time point (9 weeks).

[e]PSS: perceived social support.

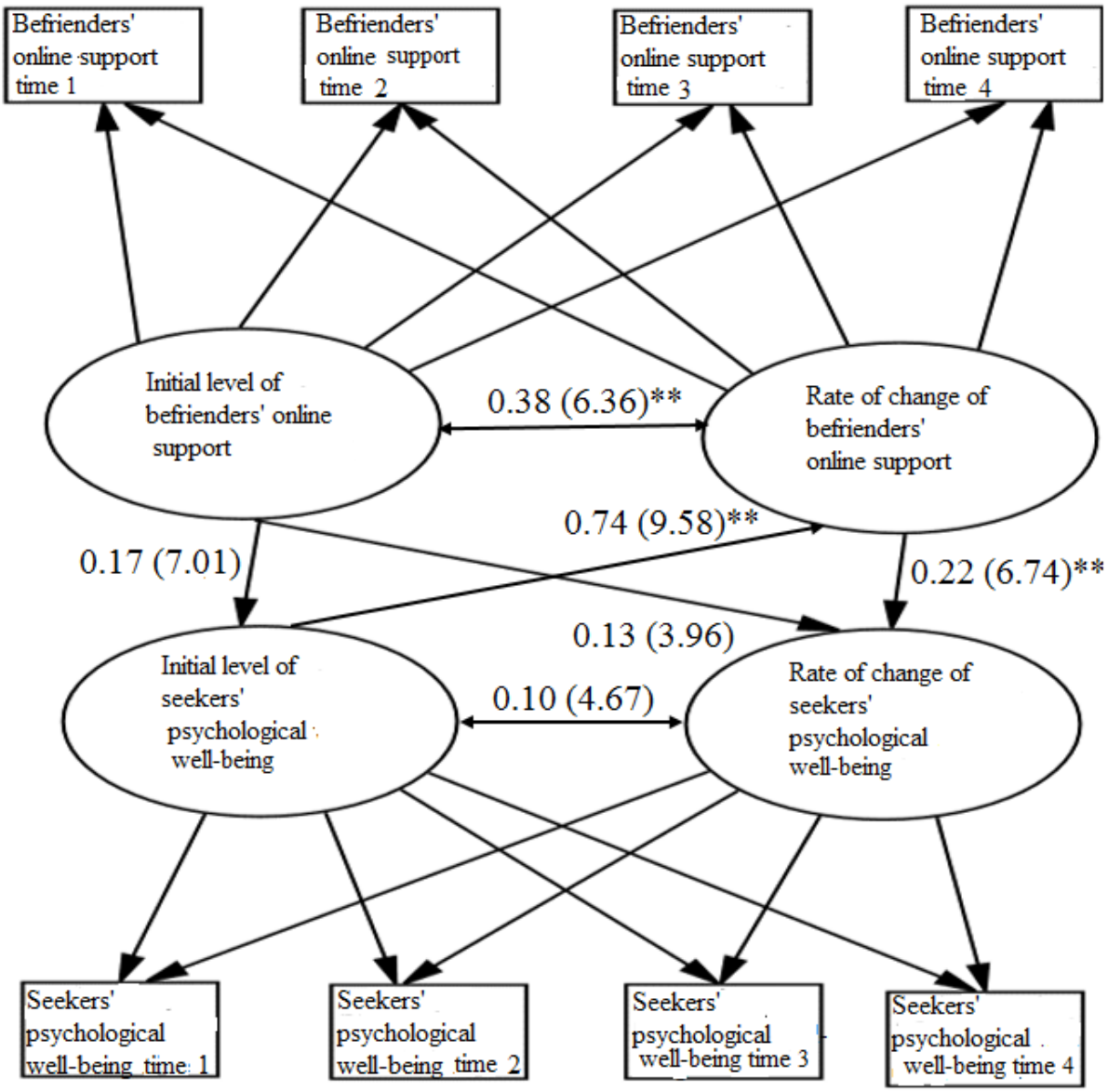
[f]PWB: psychological well-being.

Additional independent *t* tests were conducted to determine if seekers who provided data across all 4 time points were different from those who discontinued the study and provided less data points, which could bias our results. We expected instances when seekers and befrienders dropped out of the study, for example, when seekers indicated high risk of depression and suicidality during the course of the study. It was also possible that befrienders discontinued the letter exchange with seekers because the letter content caused distress to the befrienders. In such cases, the platform informed the seekers and sought their consent to end the letter thread or continue the letter exchange with a new befriender.

For research question 2c and hypothesis 2c, we elucidated the mechanism of change linking the digital peer support intervention to emerging adult mental well-being (Figure 3) by assessing whether and how the initial level and rate of change from baseline to weeks 3, 6, and 9 in befrienders' support related to seekers' initial level and rate of change in mental well-being.

The findings revealed that the conditional growth model on the growth factors of befrienders' support that predicted seekers' psychological well-being demonstrated acceptable fit (N=100, χ^2_{19}=71.4; *P*<.001; RMSEA=0.06, 90% CI 0.02-0.08; SRMR=0.08; CFI=0.98; BIC=4599.26). None of the covariates for befrienders' support and seekers' psychological well-being were significant. The initial level of seekers' experience of psychological symptoms positively predicted the rate of change (ie, increase) in befrienders' support (r_{100}=0.74; *P*<.001). Seekers with lower mental well-being at the start of the study, who reported more depressive and anxiety symptoms, experienced a greater rate of increase in support from befrienders over the course of 21 days or 3 weeks of the intervention using the Acceset platform. The rate of change or increase in befrienders' support positively predicted the rate of change or decrease in seekers' psychological symptoms (r_{100}=0.22; *P*<.001). In other words, befrienders' support that increased over the course of the intervention led to the decrease in depressive and anxiety symptoms among seekers at postintervention assessment (3 weeks) and at the follow-up assessments (6 and 9 weeks). Thus, seekers' engagement with the digital peer support intervention over the Acceset platform demonstrated both immediate (at 3 weeks after the intervention) and prospective implications for enhanced mental well-being.

Figure 3. Conditional growth model on befrienders' support and seekers' psychological well-being (befrienders: n=30 emerging adults; seekers: n=100 emerging adults). All path coefficients are standardized, with SEs in parentheses; none of the covariates were significant, and they were not included in this model. ***P*<.01.



Discussion

Principal Findings

This RCT assessed the effectiveness of a digital peer support intervention on emerging adults' psychological well-being—specifically, anxiety and depressive symptoms. In total, 2 aims were explored. First, we assessed the training effectiveness of the digital peer support training and intervention. Befrienders' adoption of and fidelity to the training program was indicated by the extent to which their peer support responses demonstrated the 4 components of well-being—mattering, selfhood, compassion, and mindfulness. Hypothesis 1 was partially supported; befrienders' peer support responses demonstrated significantly higher post- than pretraining scores for selfhood but not for the other 3 components.

Second, we examined the clinical efficacy of the digital peer support intervention on seekers' psychological well-being. We found evidence for hypothesis 2a in that the intervention enhanced 3 components of psychological well-being among seekers over the course of the study; these were selfhood, compassion, and mindfulness. Our findings provided evidence for hypothesis 2b. Specifically, seekers in the intervention group had improved psychological well-being with lower symptoms of anxiety and depression after the intervention than seekers in the waitlist control group before the intervention. Furthermore, the effect of the intervention on seekers' psychological well-being was sustained beyond the period of the intervention. In particular, seekers reported less psychological symptoms at postintervention assessment (3 weeks from baseline) and at the first follow-up assessment (6 weeks from baseline) but not at the second follow-up assessment (9 weeks from baseline). We

also assessed seekers' support-seeking behaviors beyond the Acceset platform and found that their perceived social support at 3, 6, and 9 weeks did not differ significantly from baseline—before the intervention.

We elucidated the mechanism of change that linked befrienders' digital peer support to the psychological well-being of seekers. Our results lend credence to hypothesis 2c—seekers' engagement with the digital peer support intervention demonstrated both immediate and prospective implications for their psychological well-being. The increasing trajectory of befrienders' support over the course of the intervention predicted seekers' experience of decrease in anxiety and depression symptoms at the postintervention (3 weeks) and follow-up (6 and 9 weeks) assessments. This study on a digital peer support intervention for emerging adult psychological well-being is the first to harness the interventional potential of 4 components of emerging adult psychological well-being and elucidate a mechanism of change involving digital peer support in enhancing psychological well-being. Our intervention also incorporated and validated digital markers of psychological well-being, especially emotion stamps, motivational Graphic Interface Formats, and functional adjustment stamps as reflecting emotionality, motivations, and psychological symptoms, respectively. Findings from our RCT provide the parameters and conditions for a novel mobile health (mHealth) peer support intervention that is effective in intervening in emerging adult psychological well-being in real-world settings.

Effectiveness of Digital Peer Support Intervention: Training Befrienders (Aim 1)

Of the 4 components of psychological well-being, specifically, mattering, selfhood, compassion, and mindfulness, our RCT revealed that digital peer support training is effective in building emerging adults' capacity to provide peer support responses that enhance their peers' selfhood. This finding is consistent with an extensive body of work documenting the role of peers that is crucial in young people's self and identity development as they establish and maintain complex social relationships, especially peer relationships [56-58]. In contrast, the lack of evidence regarding training emerging adults, specifically peer befrienders who provided support, in harnessing mattering, compassion, and mindfulness in their peer responses suggests that these components may require more deliberate efforts to be incorporated into young individuals' support responses. Selfhood development is routine and happens in everyday life [56-58], especially during emerging adulthood, when negotiating identity exploration is a key developmental task [1], whereas compassion and mindfulness are more intentional and require active engagement of one's cognition and insights [37-39]. Future RCTs are necessary to validate the adoption of and fidelity to training efficacy to harness the 4 components of psychological well-being in emerging adults' online peer support responses.

Digital Peer Support Intervention: Clinical Outcomes for Seekers and Mechanism of Change (Aim 2)

Our RCT findings on the increasing trajectories of selfhood, compassion, and mindfulness among seekers—emerging adults who sought support on the digital platform—provide empirical evidence for these 3 components as active ingredients of psychological well-being [19-21]. The characteristics of the change trajectories of selfhood, compassion, and mindfulness among seekers were aligned with the findings on befrienders' adoption of and fidelity to the peer support training curriculum. In particular, we found that, among seekers, the moderate initial level of selfhood increased linearly throughout the course of the study, which contrasted with the low initial levels of compassion and mindfulness that increased sharply following the intervention but modestly at the follow-ups. Thus, selfhood may present greater interventional potential as an active ingredient of emerging adults' psychological well-being—both for training peers in providing support and for those who seek support—in light of how a key developmental feature of emerging adulthood is self-exploration [1]. These findings are consistent with our ongoing meta-analytic review on how young people's selfhood and psychopathology indicate the negative moderate relations of multiple self-variables with specific mental health conditions—anxiety and depression (PROSPERO registration: CRD42021248495).

Examining clinical efficacy, we found that the digital peer support intervention led to improved psychological well-being among emerging adults, with lower self-report symptoms of anxiety and depression at the postintervention assessment and the first follow-up assessment at 6 weeks but not at 9 weeks. Clinical outcomes associated with traditional peer support that takes place face to face are well established [28]. Importantly, our RCT provides robust empirical evidence for the clinical effectiveness of peer support delivered on a digital platform. Our ongoing systematic review and meta-analysis on peer support and mental health also found that interventional peer support has a high positive effect on mental health functioning with limited carryover effect (PROSPERO registration: CRD42022353624). Our RCT indicated that the window for harnessing the effective, therapeutic potential of online peer support for emerging adult psychological well-being via the selected peer support platform may not extend beyond 6 weeks—a finding that warrants future empirical investigation. Although we assessed and expected individuals to indicate higher perceived social support during the intervention and beyond (as compared with the level before the intervention), our findings were contrary to expectations. Perceived social support is typically conceptualized and operationalized as a regular form of peer support that one obtains from close social networks [47], which may be distinct from active engagement in and solicitation of peer support from peers with common lived experiences from the community [59].

By elucidating the mechanism of change, we found that emerging adults with lower psychological well-being at the start of the study, who reported more depressive and anxiety symptoms, experienced greater online peer support from befrienders over the course of the intervention using the Acceset platform. This result provides important insights into the therapeutic benefits of active solicitation of support from peers in the community for enhancing psychological well-being among emerging adults, especially college students [60], and lends credence to the social sharing of emotions framework [12,13]. In particular, emotional disclosure on digital platforms functions

as a psychological process of support that facilitates emotion regulation and recovery and mitigates the experience of anxiety and depressive symptoms among emerging adults. More importantly, our findings revealed that online peer support increased over the course of the intervention, which predicted the pronounced decrease in emerging adults' (seekers) psychological symptoms at the postintervention assessment (3 weeks) and the sustained level of symptoms at the follow-up assessments (6 and 9 weeks). Thus, engagement with the digital peer support intervention demonstrated both immediate and prospective implications for enhanced psychological well-being among emerging adults.

Limitations and Conclusions

A possible limitation of this RCT of a digital peer support intervention for emerging adult psychological well-being is the use of self-report measures for assessing clinical outcomes, which could be subject to under- or overestimation of anxiety and depressive symptoms [45]. However, the use of self-report measures was aligned with the confidentiality and feasibility considerations given the anonymity of seeker-befriender-moderator interactions on the platform and is consistent with the standard procedure of peer-led mHealth interventions for young people [61]. Another limitation is the preliminary results on validating the digital markers of psychological well-being on the Acceset platform. Future research may assess additional evidence to further evaluate the actionability of these digital markers. In turn, these markers could potentially be used to triangulate evidence from self-report measures and physiological biomarkers, such as cortisol levels and blood pressure, to enhance the validity and reliability of clinical outcomes involving emerging adults' psychological well-being [62]. Furthermore, the evidence built on the single-site superiority trial outlined in this study limits the external validity of the digital peer support intervention in real-world settings [30]. Future research should consider properly powered and rigorous studies using multiple trials of strategies and multiple sites for potential expansion into larger cohorts and potential downstream studies that could provide insight into longer-term retention to build the evidence on the effects of digital interventions in addressing emerging adult mental health.

Notwithstanding these limitations, this study's development and validation of a novel digital innovation realize important contributions to the field of emerging adult mental health [25,30]. The key strengths of the proposed intervention are the scalability and sustainability of the digital peer support intervention. Existing evidence suggests that for scalability of web-based nonprofessional peer support training to attain the desired reach, it should incur minimal cost, not be constrained by geographical locations, and be available to individuals from diverse backgrounds and abilities [26,61]. Congruent with these findings, the studied digital peer support intervention is designed to maximize reach and is widely accessible at scale by adopting dual community engagement approaches (active and consultative) to drive optimal user engagement [30]. Our intervention uses the active method of community engagement in providing web-based safety support for young people. This form of peer support is affordable and readily available as it draws on the common lived experiences of the community, with peers functioning as befrienders and moderators who provide their peer seekers with emotional support. By using the community consultative method, college students aged 19 to 25 years from institutes of higher learning (IHLs) in Singapore were consulted in co-designing the Acceset platform. Relatedly, a unique aspect of our intervention is that it taps into community-driven cocreation, validation, and potential deployment of mHealth interventions for mental health by leveraging technology and engaging clinicians, psychologists, and counselors in the community to further the sustainability of digital peer support. This intervention may represent a scalable, sustainable, and low-cost prevention strategy that has therapeutic potential in supporting the psychological well-being of young people.

There are potential recommendations for future design and implementation of digital peer support for youth mental health based on results from our RCT and those from a systematic review and meta-analysis of RCTs on the sustainable effects of mental health interventions for students from IHLs [63-65]. First, digital peer support interventions need to maximize the effectiveness and sustainability of psychological interventions for mental health, especially for emerging adults, by emphasizing the importance of a multi-systemic approach [63-65] that entails contributions from the individual, community, and societal levels. Second, to establish and maintain the dynamics of the digital peer support platform, future interventions can consider a system comprising seekers, befrienders, and moderators. This dynamic system is characterized by individual contributions and the opportunity for seekers to join the peer support network as befrienders to expand the online support groups. Third, to normalize digital peer support as a new service delivery model, collaborations with policy makers and IHLs in designing safety standards and protocols for emerging adult engagement with digital peer support is necessary as a next step forward.

Acknowledgments

DH gratefully acknowledges funding from the National Research Foundation Singapore under its Artificial Intelligence Singapore Programme (award AISG-GC-2019-002); the Singapore Ministry of Health National Medical Research Council under its Open Fund – Large Collaborative Grant (MOH-OFLCG18May-0028); and the Institute for Digital Medicine Translational Research Programme (grant R-719-000-037-733) at the Yong Loo Lin School of Medicine, National University of Singapore, Ministry of Education Tier 1 Academic Research Fund Grant (R-397-000-333-114). The funding sources had no role in the investigation outline or in the decision to submit the manuscript for publication. The authors would also like to acknowledge Ngiam Ze Hui, Li Jia Shun, Cameron Tan, and Mitali Kothari for assistance in research and literature review; Nikhil Gupta, Manish Gupta, Uday

Pratap Singh, and Woo Wai Leng for product maintenance and support; Vanessa Goh, Clarice Goh, Liao Yihan, Liao Yijou, and Kwok Yanqi for design support; and Dr Matthew Lim Sheng Mian, Dr Chang Weining, Dr Jimmy Lee, Dr Chua Hong Choon, Marissa Teo, and Hwee Boon for training material advisory.

Authors' Contributions

GHY conceptualized and designed the study, obtained ethics approval and clinical trial registration, and drafted the manuscript. GL, MO, and RP were in charge of data collection and management and responsible for participant recruitment and follow-up. DH was involved in advising the ethical and clinical trial registration and drafting the manuscript. All authors approved the manuscript.

Conflicts of Interest

MO is cofounder and chief executive officer of Acceset Pte. Ltd. RP is an employee of Acceset Pte. Ltd. DH is a scientific cofounder and shareholder of KYAN Therapeutics, which is developing digital medicine-based platforms to optimize cancer therapy. DH is also an inventor of pending patents pertaining to personalized medicine.

Multimedia Appendix 1
Coding scheme and supplementary analyses.
[DOCX File , 43 KB - mental_v10i1e43956_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3623 KB - mental_v10i1e43956_app2.pdf]

References

1. Arnett JJ, Žukauskienė R, Sugimura K. The new life stage of emerging adulthood at ages 18-29 years: implications for mental health. Lancet Psychiatry 2014 Dec;1(7):569-576. [doi: 10.1016/S2215-0366(14)00080-7] [Medline: 26361316]
2. Magson NR, Freeman JY, Rapee RM, Richardson CE, Oar EL, Fardouly J. Risk and protective factors for prospective changes in adolescent mental health during the COVID-19 pandemic. J Youth Adolesc 2021 Jan;50(1):44-57 [FREE Full text] [doi: 10.1007/s10964-020-01332-9] [Medline: 33108542]
3. Shanahan L, Steinhoff A, Bechtiger L, Murray AL, Nivette A, Hepp U, et al. Emotional distress in young adults during the COVID-19 pandemic: evidence of risk and resilience from a longitudinal cohort study. Psychol Med 2022 Apr;52(5):824-833 [FREE Full text] [doi: 10.1017/S003329172000241X] [Medline: 32571438]
4. Ohannessian CM. Introduction to the special issue: the impact of the COVID-19 pandemic on the lives of emerging adults. Emerg Adulthood 2021 Oct;9(5):431-432. [doi: 10.1177/21676968211030875]
5. Lippke S, Fischer MA, Ratz T. Physical activity, loneliness, and meaning of friendship in young individuals - a mixed-methods investigation prior to and during the COVID-19 pandemic with three cross-sectional studies. Front Psychol 2021 Feb 2;12:617267 [FREE Full text] [doi: 10.3389/fpsyg.2021.617267] [Medline: 33603702]
6. Loades ME, Chatburn E, Higson-Sweeney N, Reynolds S, Shafran R, Brigden A, et al. Rapid systematic review: the impact of social isolation and loneliness on the mental health of children and adolescents in the context of COVID-19. J Am Acad Child Adolesc Psychiatry 2020 Nov;59(11):1218-39.e3 [FREE Full text] [doi: 10.1016/j.jaac.2020.05.009] [Medline: 32504808]
7. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
8. Arenas DL, Viduani AC, Bassols AM, Hauck S. Peer support intervention as a tool to address college students' mental health amidst the COVID-19 pandemic. Int J Soc Psychiatry 2021 May;67(3):301-302. [doi: 10.1177/0020764020954468] [Medline: 32883124]
9. Drysdale MT, McBeath ML, Callaghan SA. The feasibility and impact of online peer support on the well-being of higher education students. J Ment Health Train Educ Pract 2022;17(3):206-217. [doi: 10.1108/jmhtep-02-2021-0012]
10. Grégoire S, Beaulieu F, Lachance L, Bouffard T, Vezeau C, Perreault M. An online peer support program to improve mental health among university students: a randomized controlled trial. J Am Coll Health (forthcoming) 2022 Aug 09:1-13. [doi: 10.1080/07448481.2022.2099224] [Medline: 35943903]
11. Suresh R, Alam A, Karkossa Z. Using peer support to strengthen mental health during the COVID-19 pandemic: a review. Front Psychiatry 2021 Jul 12;12:714181 [FREE Full text] [doi: 10.3389/fpsyt.2021.714181] [Medline: 34322045]
12. Rimé B. Emotion elicits the social sharing of emotion: theory and empirical review. Emot Rev 2009 Jan;1(1):60-85. [doi: 10.1177/1754073908097189]

13. Rimé B. Self disclosure. In: Friedman HS, editor. Encyclopedia of Mental Health. 2nd edition. Waltham, MA, USA: Academic Press; 2016:66-74.
14. Choi M, Toma CL. Social sharing through interpersonal media: patterns and effects on emotional well-being. Comput Human Behav 2014 Jul;36:530-541. [doi: 10.1016/j.chb.2014.04.026]
15. Vermeulen A, Vandebosch H, Heirman W. #Smiling, #venting, or both? Adolescents' social sharing of emotions on social media. Comput Human Behav 2018 Jul;84:211-219. [doi: 10.1016/j.chb.2018.02.022]
16. Nesi J, Choukas-Bradley S, Prinstein MJ. Transformation of adolescent peer relations in the social media context: part 1-a theoretical framework and application to dyadic peer relationships. Clin Child Fam Psychol Rev 2018 Sep;21(3):267-294 [FREE Full text] [doi: 10.1007/s10567-018-0261-x] [Medline: 29627907]
17. Rimé B. The social sharing of emotion in interpersonal and in collective situations. In: Holyst JA, editor. Cyberemotions: Collective Emotions in Cyberspace. Cham, Switzerland: Springer; 2017:53-69.
18. Pennebaker JW, Zech E, Rimé B. Disclosing and sharing emotion: psychological, social, and health consequences. In: Stroebe MS, Hansson RO, Stroebe W, Schut H, editors. Handbook of Bereavement Research: Consequences, Coping, and Care. Washington, DC, USA: American Psychological Association; 2001:517-543.
19. Flett GL, Zangeneh M. Mattering as a vital support for people during the COVID-19 pandemic: the benefits of feeling and knowing that someone cares during times of crisis. J Concurr Disord 2020 Apr 24;2(1):106-123. [doi: 10.54127/almc5515]
20. González-García M, Álvarez JC, Pérez EZ, Fernandez-Carriba S, López JG. Feasibility of a brief online mindfulness and compassion-based intervention to promote mental health among university students during the COVID-19 pandemic. Mindfulness (N Y) 2021;12(7):1685-1695 [FREE Full text] [doi: 10.1007/s12671-021-01632-6] [Medline: 34025814]
21. Rossi A, Panzeri A, Pietrabissa G, Manzoni GM, Castelnuovo G, Mannarini S. The anxiety-buffer hypothesis in the time of COVID-19: when self-esteem protects from the impact of loneliness and fear on anxiety and depression. Front Psychol 2020 Nov 10;11:2177 [FREE Full text] [doi: 10.3389/fpsyg.2020.02177] [Medline: 33240140]
22. McComb SE, Goldberg JO, Flett GL, Rose AL. The double jeopardy of feeling lonely and unimportant: state and trait loneliness and feelings and fears of not mattering. Front Psychol 2020 Dec 17;11:563420 [FREE Full text] [doi: 10.3389/fpsyg.2020.563420] [Medline: 33391078]
23. Marshall SK, Tilton-Weaver L. Adolescents' perceived mattering to parents and friends: testing cross-lagged associations with psychosocial well-being. Int J Behav Dev 2019 Nov;43(6):541-552. [doi: 10.1177/0165025419844019]
24. MacDonald HZ, Neville T. Promoting college students' mindfulness, mental health, and self-compassion in the time of COVID-19: feasibility and efficacy of an online, interactive mindfulness-based stress reduction randomized trial. J College Stud Psychother (forthcoming) 2022 Apr 10:1-19. [doi: 10.1080/87568225.2022.2028329]
25. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015 May 19;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
26. Bernecker SL, Williams JJ, Caporale-Berkowitz NA, Wasil AR, Constantino MJ. Nonprofessional peer support to improve mental health: randomized trial of a scalable web-based peer counseling course. J Med Internet Res 2020 Sep 21;22(9):e17164 [FREE Full text] [doi: 10.2196/17164] [Medline: 32955451]
27. Myers AL, Collins-Pisano C, Ferron JC, Fortuna KL. Feasibility and preliminary effectiveness of a peer-developed and virtually delivered community mental health training program (emotional CPR): pre-post study. J Particip Med 2021 Mar 04;13(1):e25867 [FREE Full text] [doi: 10.2196/25867] [Medline: 33661129]
28. Kaess M, Moessner M, Koenig J, Lustig S, Bonnet S, Becker K, ProHEAD Consortium. Editorial perspective: a plea for the sustained implementation of digital interventions for young people with mental health problems in the light of the COVID-19 pandemic. J Child Psychol Psychiatry 2021 Jul;62(7):916-918 [FREE Full text] [doi: 10.1111/jcpp.13317] [Medline: 32924134]
29. Yeo G, Chang W, Lee LN, Oon M, Ho D. A digital peer support platform to translate web-based peer support for emerging adult mental well-being: protocol for a randomized controlled trial. JMIR Res Protoc 2022 Sep 20;11(9):e34602 [FREE Full text] [doi: 10.2196/34602] [Medline: 36125855]
30. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
31. Andalibi N, Flood MK. Considerations in designing digital peer support for mental health: interview study among users of a digital support system (Buddy Project). JMIR Ment Health 2021 Jan 04;8(1):e21819 [FREE Full text] [doi: 10.2196/21819] [Medline: 33393909]
32. Census of Population 2010 Statistical Release 1: Demographic Characteristics, Education, Language and Religion. Department of Statistics, Singapore. 2011 Jan. URL: https://www.singstat.gov.sg/-/media/files/publications/cop2010/census_2010_release1/cop2010sr1.ashx [accessed 2022-10-12]
33. Krippendorff K. Content Analysis: An Introduction to Its Methodology. 4th edition. Thousand Oaks, CA, USA: Sage Publications; 2018.
34. Rosenberg M, McCullough BC. Mattering: inferred significance and mental health among adolescents. Community Ment Health J 1981;2:163-182.

35. Rosenberg M. Society and the Adolescent Self-Image. Middleton, CT, USA: Wesleyan University Press; 1989.
36. Fenigstein A, Scheier MF, Buss AH. Public and private self-consciousness: assessment and theory. J Consult Clin Psychol 1975 Aug;43(4):522-527. [doi: 10.1037/h0076760]
37. Jazaieri H, Jinpa GT, McGonigal K, Rosenberg EL, Finkelstein J, Simon-Thomas E, et al. Enhancing compassion: a randomized controlled trial of a compassion cultivation training program. J Happiness Stud 2013;14(4):1113-1126. [doi: 10.1007/s10902-012-9373-z]
38. Sherer M, Maddux JE, Mercandante B, Prentice-Dunn S, Jacobs B, Rogers RW. The self-efficacy scale: construction and validation. Psychol Rep 1982 Oct;51(2):663-671. [doi: 10.2466/pr0.1982.51.2.663]
39. Carson SH, Langer EJ. Mindfulness and self-acceptance. J Ration Emot Cogn Behav Ther 2006 Jun 20;24(1):29-43. [doi: 10.1007/s10942-006-0022-5]
40. Flett GL, Su C, Ma L, Guo L. Mattering as a unique resilience factor in Chinese children: a comparative analysis of predictors of depression. Int J Child Adolesc Resil 2016 Mar 20;4(1):91-102 [FREE Full text]
41. Chan DW. Self-consciousness in Chinese college students in Hong Kong. Pers Individ Dif 1996 Oct;21(4):557-562. [doi: 10.1016/0191-8869(96)00088-8]
42. Chang CW, Yuan R, Chen JK. Social support and depression among Chinese adolescents: the mediating roles of self-esteem and self-efficacy. Child Youth Serv Rev 2018 May;88:128-134. [doi: 10.1016/j.childyouth.2018.03.001]
43. Yu M, Zhou H, Xu H, Zhou H. Chinese adolescents' mindfulness and internalizing symptoms: the mediating role of rumination and acceptance. J Affect Disord 2021 Feb 01;280(Pt A):97-104. [doi: 10.1016/j.jad.2020.11.021] [Medline: 33207286]
44. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
45. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
46. Musumari PM, Tangmunkongvorakul A, Srithanaviboonchai K, Techasrivichien T, Suguimoto SP, Ono-Kihara M, et al. Grit is associated with lower level of depression and anxiety among university students in Chiang Mai, Thailand: a cross-sectional study. PLoS One 2018 Dec 14;13(12):e0209121 [FREE Full text] [doi: 10.1371/journal.pone.0209121] [Medline: 30550597]
47. Zimet GD, Dahlem NW, Zimet SG, Farley GK. The multidimensional scale of perceived social support. J Pers Assess 1988 Mar;52(1):30-41. [doi: 10.1207/s15327752jpa5201_2]
48. Zhang M, Zhang J, Zhang F, Zhang L, Feng D. Prevalence of psychological distress and the effects of resilience and perceived social support among Chinese college students: does gender make a difference? Psychiatry Res 2018 Sep;267:409-413. [doi: 10.1016/j.psychres.2018.06.038] [Medline: 29960938]
49. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. Struct Equ Model 2002 Oct;9(4):599-620 [FREE Full text] [doi: 10.1207/S15328007SEM0904_8]
50. Muthén LK, Muthén BO. Mplus: Statistical Analysis with Latent Variables: User's Guide. 8th edition. Los Angeles, CA, USA: Muthén & Muthén; 2017.
51. Allaire JJ. R studio: integrated development environment for R. RStudio. Boston, MA, USA: R Studio; 2015. URL: https://www.r-project.org/conferences/useR-2011/abstracts/180111-allairejj.pdf [accessed 2022-12-20]
52. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
53. Kaplan D. Structural Equation Modeling: Foundations and Extensions. 2nd edition. Thousand Oaks, CA, USA: Sage Publications; 2008.
54. Gecková A, van Dijk JP, Stewart R, Groothoff JW, Post D. Influence of social support on health among gender and socio-economic groups of adolescents. Eur J Public Health 2003 Mar 01;13(1):44-50 [FREE Full text] [doi: 10.1093/eurpub/13.1.44] [Medline: 12678313]
55. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model 2002 Apr;9(2):233-255. [doi: 10.1207/S15328007SEM0902_5]
56. Brown BB, Larson J. Peer relationships in adolescence. In: Lerner RM, Steinberg L, editors. Handbook of Adolescent Psychology. 3rd edition. Hoboken, NJ, USA: John Wiley & Sons; 2009:74-103.
57. Erikson EH. Childhood and Society. New York, NY, USA: WW Norton; 1950.
58. Erikson EH. Identity: Youth and Crisis. New York, NY, USA: WW Norton; 1968.
59. Pfeiffer PN, Heisler M, Piette JD, Rogers MA, Valenstein M. Efficacy of peer support interventions for depression: a meta-analysis. Gen Hosp Psychiatry 2011;33(1):29-36 [FREE Full text] [doi: 10.1016/j.genhosppsych.2010.10.002] [Medline: 21353125]
60. Horgan A, McCarthy G, Sweeney J. An evaluation of an online peer support forum for university students with depressive symptoms. Arch Psychiatr Nurs 2013 Apr;27(2):84-89 [FREE Full text] [doi: 10.1016/j.apnu.2012.12.005] [Medline: 23540518]

61. Birrell L, Furneaux-Bate A, Chapman C, Newton NC. A mobile peer intervention for preventing mental health and substance use problems in adolescents: protocol for a randomized controlled trial (the Mind Your Mate study). JMIR Res Protoc 2021 Jul 30;10(7):e26796 [FREE Full text] [doi: 10.2196/26796] [Medline: 34328426]
62. Mendes JP, Moura IR, Van de Ven P, Viana D, Silva FJ, Coutinho LR, et al. Sensing apps and public data sets for digital phenotyping of mental health: systematic review. J Med Internet Res 2022 Feb 17;24(2):e28735 [FREE Full text] [doi: 10.2196/28735] [Medline: 35175202]
63. Greene MC, Huang TT, Giusto A, Lovero KL, Stockton MA, Shelton RC, et al. Leveraging systems science to promote the implementation and sustainability of mental health and psychosocial interventions in low- and middle-income countries. Harv Rev Psychiatry 2021;29(4):262-277 [FREE Full text] [doi: 10.1097/HRP.0000000000000306] [Medline: 34241978]
64. Roystonn K, Vaingankar JA, Chua BY, Sambasivam R, Shafie S, Jeyagurunathan A, et al. The public health impact and policy implications of online support group use for mental health in Singapore: cross-sectional survey. JMIR Ment Health 2020 Aug 04;7(8):e18114 [FREE Full text] [doi: 10.2196/18114] [Medline: 32749231]
65. Winzer R, Lindberg L, Guldbrandsson K, Sidorchuk A. Effects of mental health interventions for students in higher education are sustainable over time: a systematic review and meta-analysis of randomized controlled trials. PeerJ 2018 Apr 2;6:e4598 [FREE Full text] [doi: 10.7717/peerj.4598] [Medline: 29629247]

Abbreviations

BIC: Bayesian information criterion
CFI: comparative fit index
GAD-7: 7-item General Anxiety Disorder Scale
IHL: institute of higher learning
IRB: Institutional Ethics Review Board
mHealth: mobile health
NUS: National University of Singapore
PHQ-9: 9-item Patient Health Questionnaire
RCT: randomized controlled trial
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual

Edited by J Torous, G Eysenbach; submitted 31.10.22; peer-reviewed by A Serlachius, M Schure; comments to author 24.12.22; revised version received 03.02.23; accepted 07.02.23; published 18.04.23.

Please cite as:
Yeo G, Loo G, Oon M, Pang R, Ho D
A Digital Peer Support Platform to Translate Online Peer Support for Emerging Adult Mental Well-being: Randomized Controlled Trial
JMIR Ment Health 2023;10:e43956
URL: https://mental.jmir.org/2023/1/e43956
doi:10.2196/43956
PMID:36756843

Original Paper

Online Mental Health Forums and Rural Resilience: Mixed Methods Study and Logic Model

Artur Steiner[1], PhD; Jane Farmer[2], PhD; Peter Kamstra[3], PhD; Karen Carlisle[4], PhD; Anthony McCosker[2], PhD; Sue Kilpatrick[5], PhD

[1]Glasgow Caledonian University, Yunus Centre, Glasgow, United Kingdom
[2]Swinburne University of Technology, Centre for Social Impact Swinburne, Melbourne, Australia
[3]University of Melbourne, Melbourne, Australia
[4]James Cook University, Townsville, Australia
[5]University of Tasmania, Hobart, Australia

Corresponding Author:
Artur Steiner, PhD
Glasgow Caledonian University
Yunus Centre
M201 George Moore Building
Cowcaddens Road
Glasgow, G4 0BA
United Kingdom
Phone: 44 141 331 8445
Email: artur.steiner@gcu.ac.uk

Abstract

Background: Rural mental health is a growing area of concern internationally, and online mental health forums offer a potential response to addressing service gaps in rural communities.

Objective: The objective of this study was to explore and identify pathways by which online peer support mental health forums help to build resilience for rural residents experiencing mental ill-health by contributing to overcoming their specific contextual challenges.

Methods: We developed a Theoretical Resilience Framework and applied it to 3000 qualitative posts from 3 Australian online mental health forums and to data from 30 interviews with rural forum users.

Results: Drawing on the findings and an abductive approach, a logic model was developed to illustrate links between the resilience resources built and enabling features of forums that make them spaces that facilitate resilience.

Conclusions: The study demonstrated that online forums make valuable contributions to social well-being and access to a range of timely support services for rural people experiencing mental ill-health, and, while doing so, involve users in the processes of resilience building. The study provides a new way for practitioners to frame the work of and value produced by forums. It gives a logic model that can be used in evaluation and audit as it facilitates a causal framing of how forums, as an intervention, link with resilience outcomes. Ultimately, the study contributes to developing new knowledge about how rural resilience building can be conceptualized and measured while showing how forums are part of contemporary health service provision in rural places.

(JMIR Ment Health 2023;10:e47459) doi:10.2196/47459

KEYWORDS

online forums; personal resilience; mental health; rurality; logic model

Introduction

Online Mental Health Peer Support Forums

Since the 2010s, online mental health forums have gained popularity as a way to give and receive information and emotional support between peers experiencing mental health conditions [1]. Factors influencing this trend include increased accessibility of the internet [2], evidence of the benefits of mental health peer support [3,4], and increasing provision of mental health and well-being services by nonprofit organizations focusing on holistic support [5]. Most recently, the COVID-19

pandemic accelerated public acceptability of web-based health services [6].

Studies provide evidence of multiple benefits from using online mental health peer support forums [7]. Forums offer timely access, are efficient in responding to multiple types of needs [8], and overcome geographical boundaries by linking people with similar interests [1]. Compared with the provision of traditional on-premises services, a unique advantage of online peer support forums is their ability to provide access to information from others with similar experiences who can provide advice in nonclinical language [9]. This creation of a space of relating and empathy formed by exchanges between forum users helps to develop trust [10,11], enabling people to discuss their needs and to develop strategies for coping within a safe, anonymous environment [12].

In this study, we were particularly interested in the aspect of developing coping abilities and *the use of positive coping mechanisms such as seeking social support, positive thinking and problem solving*—all associated with the concept of resilience [13]. Interestingly, although previous studies of online forum use show a range of benefits, including some that have looked specifically at rural users [1], we have found no study that specifically considered *how* online forums could help to build resilience for rural people experiencing mental ill-health. As such, we draw together existing discussions about online mental health forums and rural resilience, and in doing so, present new knowledge for the field. Next, we provide an explanation of resilience in the context of this study.

Resilience in This Context

As a latent construct that is not objectively observable [14], resilience is conceptualized in different ways depending on the context. In social scenarios, it is suggested as a desirable collective resource for dealing with adversities, including responding to environmental disasters, lack of economic opportunities, inaccessibility of services, and public health challenges [15,16]. Community resilience has been discussed as building through proactive processes that continually add capacity for populations to respond to risks and effect change and involving persistence, adaptation, and transformation acting as interrelated phenomena allowing resilience to occur [17,18]. Resilience relates to the capacity to adapt and thrive by regarding "disturbances as an opportunity for change and development" [19]. It can be understood as both an ongoing process [20] and an outcome [21]. Magis [22] locates resilient community members in a relationship with resilient communities, suggesting a circular flow where resilient collectives help individuals by making resources available and where individuals then harness these resources to add to community capacity. She says resilience is "the existence, development, and engagement of community resources by community members to thrive in an environment characterized by change, uncertainty, unpredictability, and surprise. Members of resilient communities intentionally develop personal and collective capacity that they engage to respond to and influence the change, to sustain and renew the community, and to develop new trajectories for the communities' future."

A full description of the varying understandings of resilience is beyond the scope of this paper (for a detailed discussion of resilience in the rural context, refer to the paper by Skerratt [16]). Here, we understand resilience as having the individual resources to *adapt and evolve* to new circumstances after or while experiencing adversity [23] and *building strength* to deal with adversity and overcome vulnerabilities [20]. As we feature rural residents, we term this as "individual rural resilience." However, reflecting back to the quote by Magis [22], findings are also relevant to other levels of resilience, involving understandings of collective or community resilience. This is because the collective of people using a forum can be understood as an *online community* where—by interacting on the web—forum users exchange resources for building resilience with each other. In addition, the people using the forums live in *geographically rural communities*, and thus, there is potential for positive gains in individuals' resilience to contribute to collective resilience resources available to their geographical communities. This would happen by enabling people to participate in community life and by adding to the diversity of people who can contribute to community capacity [24].

Although governments and academics have proposed developing resilience as a way to respond to coping in challenging rural contexts, for individuals and communities, there has been less granular study of how specific interventions can influence rural resilience [17]. This results in a gap in tools to assist policy makers, capacity-building practitioners, and citizens to assess what is happening in resilience processes.

Forums and Resilience Building

This study focused on rural people experiencing mental ill-health. It focuses on what happens *in* online mental health peer support forums, and it takes a resilience lens to understand the impacts that forums can have on users' lives. Living in rural places is widely acknowledged to present specific challenges for people experiencing mental ill-health, including inaccessibility of health services, lack of confidentiality, and experiences of stigma and isolation [25]. Given the exacerbated experiences of adversity for rural people experiencing mental ill-health compared with those of urban dwellers [26], there is distinct value in understanding the mechanisms that build rural resilience and respond to their needs.

Building resilience for people with mental ill-health resonates with understandings of recovery for this group as about gaining resources to deal with a fluctuating condition. Duff [27] depicts a good outcome as achieving "health in illness" by navigating "an individual practice of health in perpetual tension with the enduring symptoms of mental illness." Understanding how rural individuals experiencing mental ill-health can build resilience via online forum use is not only important in itself but also—by helping to unpack pathways to resilience—can help to inform ways to develop and assess other interventions targeting rural resilience [23,28].

Considering the rapid development of web-based services, this study aimed *to explore, systematize, and show a pathway by which specialized online peer support forums can act as an intervention to assist resilience building for rural people experiencing mental ill-health*. To do so, we developed a

Theoretical Resilience Framework by using evidence combining rural community development and psychology perspectives. This draws on a chronology of multidisciplinary research by a team including Buikstra et al [20] and Berkes and Ross [23]. The framework (1) guided the analysis of post data generated by rural users of three online mental health forums hosted by Australian nonprofit organizations and (2) informed a topic schedule for interviews with rural forum users. In addition, by identifying resilience resources built, our data analysis reveals features of forums that make them spaces that enable resilience. By discovering *how* resilience builds and what enables it, we aimed to present a logic model that shows a pathway between using forums and rural resilience.

Theoretical Resilience Framework

To address the study aims and assist with data collection and analysis, we used a narrative review approach [29] to inform the development of a Theoretical Resilience Framework (Table 1). A narrative review is a semisystematic literature review that is useful when analyzing complex, latent constructs, such as resilience [30], that are conceptualized diversely, making it challenging to conduct a fully systematic review [31]. Much of the literature about resilience takes a *rural community*–level perspective—that is, it considers collective population resources and structural resources (eg, economic opportunities, amenities, and services [32]). Another stream concerns *individual psychological* resources and is not contextually specific [33]. Context-free perspectives fail to acknowledge the layers of adversity experienced by rural people [26]. Our study targets the nexus of individuals experiencing *mental ill-health* and *living in rural contexts*. Few researchers target this specific area. Studies can be traced back to an initial multidisciplinary study of "shapers" of rural individual resilience [34]. Buikstra et al [20] refined and tested that work, offering a starting set of rural "resilience promoting factors." Building from that study [20], we applied a snowball method [35] to identify other relevant literature. This involved (1) searching for later studies about resilience by authors of the original paper (ie, published after 2010) and (2) searching on Google Scholar for all recent papers that cited or were "related to" the study by Buikstra et al [20] (N=307 [cited or related]). From these studies, we included all those stating resilience factors that could relate to individuals (vs community-level factors; Table 1).

Table 1. Identifying individual resilience factors in the literature.

Studies with individual resilience factors	Resilience factors given	Individual resilience factors
Hegney et al [34]	• The environment • Connection with the place • Family • Culture • Being part of a rural community and community spirit	• Connection with the place • Family • Culture
Buikstra et al [20]	• Social networks and support • Positive outlook • Learning • Early experience • Environment and lifestyle • Sense of purpose • Embracing differences • Beliefs • Leadership	• Social networks and support • Positive outlook • Learning • Early experience • Environment and lifestyle • Sense of purpose • Embracing differences • Beliefs • Leadership
Berkes and Ross [23]	• Social capital and networks • Sense of place • Values • Social identity	• Social capital and networks • Sense of place • Values • Social identity
Maclean et al [36]	• Knowledge, skills, and learning • Community networks • People-place connections • Community infrastructure • Engaged governance	• Knowledge, skills, and learning
Leite et al [37]	• Formal education • Local knowledge • Autonomy • Family support • Social capital and networks of support • Access to psychological counseling • An "invigorating environment"	• Formal education • Local knowledge • Autonomy • Family support • Social capital and networks of support

Using this set of individual resilience factors, we applied an abductive approach to generate a "final" Theoretical Resilience Framework against which to analyze all data in our study. Abduction involves working between an initial general theory

(in this case, the multiple resilience factors suggested in the literature) and specific data (in this case, forum posts) to arrive at a workable theory that accounts for empirical data [38]. To do this, 3 researchers analyzed a sample of 100 forum posts and found consistent evidence of 5 factors (shown in Table 2). On speculating why only these variables were repeatedly evidenced, we found that some other variables appear to be more descriptive of personal attributes that would be less changeable over time, for example, beliefs and early experiences. Other variables are challenging to identify in qualitative data sets—for example, sense of purpose, positive outlook, embracing difference, and leadership. Notably, the factors in the final framework all represent *resources that are subject to change*, which makes them useful for studying the building of resilience in response to an intervention.

Table 2. Identifying factors of the Theoretical Resilience Framework.

Initial rural individual resilience factors (from literature)	Retained or rejected based on sample analysis
Social capital and related social connections, social support, family support, and social networks	Retained
Sense of belonging (including connection to the place)	Retained
Learning knowledge, skills, and education	Retained
Self-efficacy agency, self-organizing, and autonomy	Retained
Adaptive capacity	Retained
Culture, values, social identity, and beliefs	Rejected
Positive outlook	Rejected
Early experiences	Rejected
Sense of purpose	Rejected
Embracing difference	Rejected
Leadership	Rejected
Environment and lifestyle	Rejected
Local knowledge	Rejected

Therefore, the final Theoretical Resilience Framework postulates that individual rural resilience, in the context of this study (involving people experiencing mental ill-health in a rural context), is characterized by (1) *social capital* or access to networks of people and the support, trust, and inclusion they can foster [20]; (2) *a sense* of *belonging*, which concerns acceptance as part of a group or community and the processes that lead to identity formation [16]; (3) access to *learning and knowledge*, including information about multiple topics, leading to increased capacity to navigate change [21]; (4) *self-efficacy* or being able to self-organize or work toward feeling in a controlled state [23]; and (5) *adaptive capacity* in the sense of having resources that enable adaptation and behavior change [39]. The framework was used to guide data collection and analysis as discussed in the following sections.

Methods

Overview

This study derives from a large project called o*ptimizing the roles of online communities in rural resilience* and funded by the Australian Research Council. To explore resilience processes, we obtained qualitative forum post data of rural users and interviewed rural forum users in relation to online peer support forums of 3 well-known Australian mental health nonprofit organizations—SANE Australia, Beyond Blue, and ReachOut. The organizations, and thus their forums, target specific demographic groups (Table 3). We analyzed the post data thematically using a Theoretical Resilience Framework (Table 2). To verify the resilience themes found in the post data and to understand more about the features that enable resilience, in-depth interviews were conducted with 30 rural forum users. Our literature review and the 2 data sets were used together to triangulate findings and to identify pathways between the use of online mental health forums and building resilience.

Table 3. Characteristics of host organizations, forums, and data collected.

	SANE Australia	Beyond Blue	ReachOut
Group targeted	People experiencing complex forms of mental ill-health including schizophrenia, bipolar disorder, and posttraumatic stress disorder	People experiencing anxiety, depression, grief, and posttraumatic stress disorder	Young people aged between 14 and 25 years (inclusive) and experiencing mental health issues
Goals of the forum	Advocacy, research, and support	Reduce stigma and discrimination and deliver better health outcomes for people in relation to anxiety, depression, and suicide	Prevent or delay the onset of mental health issues and reducing their incidence, severity, duration, and frequency
Posts from rural areas (N=193,356), n (%)	12,032 (6.22)	5027 (2.59)	11,905 (6.16)
Authors of posts from rural areas in the sample analyzed, n	251	684	121
Interviewees from rural areas (n=30), n (%)	20 (66.67)	6 (20)	4 (13.33)

In this study, we define rural areas as all areas outside of "Major cities" using the Australian Statistical Geography Standard (ASGS) Remoteness Structure [40], which categorizes locations based on populations' relative accessibility to services. This is a measure of rurality that "makes sense" in an Australian context, as populations and services are concentrated in major cities. According to a recent review, 49% of studies used a measure of relative service accessibility to denote relative rurality [41].

Ethics Approval

Ethics committee approval was granted by Swinburne University (R/2019/033).

Participating Forums

Available nationally to Australians, the forums we studied are accessible to anyone with internet access. People must register to be a member, and this allows them to post and read others' posts. We use the term "users" here to distinguish those who post compared with other members who do not post. Clearly, this is shorthand, and members might "use" the forums by reading and using information in posts while not posting themselves. Those who do not post were not included. Forum users can choose to post on existing threads or to create new topic threads. Forums are moderated by staff paid by the host nonprofit organizations. Moderation broadly involves ensuring that users do not disclose personal or identifiable information, removing spam content, removing discouraging comments, preventing prescriptive advice and abusive language, and diffusing conflict.

Forum Post Data

To ethically collect forum post data, we ensured that our study aligned with the research principles of each of the host web-based mental health organizations [42]. We were able to use forum data because, when people register, they agree that their data may be used or reused for research purposes as long as it remains anonymous. Initially, a sample of deidentified forum posts (N=193,356 posts) was obtained, including from SANE (68,634/193,356, 35.49%), ReachOut (80,174/193,356, 41.46%), and Beyond Blue (44,548/193,356, 23.04%). Posts were collected from the period between August 2018 and December 2020. Data were "cleaned" before analysis to remove content that could be potentially identifiable, for example, web-based pseudonyms and references to named geographical locations, organizations, or people. Then, as we focused on rural residents, we selected all posts in which rural postcodes were provided when registering for membership. To obtain a sample that we could analyze in the time available, we further selected a sample of 1000 rural posts per forum (n=3000 posts). These 3000 posts included all posts from *very remote* (n=269, 8.97%) and *remote* (n=410, 13.67%) ASGS categories and a sample of 2321 (77.37%) *outer regional*–ASGS category posts randomized using the Excel (Microsoft Corporation) random function (ie, =RAND).

Interview Data

A sample of rural user interviewees was recruited using a recruitment advertisement posted on each forum. The recruitment posts explained the study's aims and invited rural users to contact the researchers using a web-based "expression of interest" form. The interviews with 30 rural forum users were conducted between 2021 and 2022. Owing to rural locations, COVID-19 restrictions, and respect for users' anonymity, 80% (24/30) of the interviews was conducted via phone and 20% (6/30) was conducted via Zoom. Duration ranged from 45 to 60 minutes. A "duty of care" protocol was provided that offered support to interviewees to access counseling resources if required.

Interviews were semistructured and included questions about users' experiences with using forums, the extent to which benefits or disadvantages accrue, and reasons for changes to occur. Questions about resilience themes identified in the Theoretical Resilience Framework were included. Interviews encouraged users to comment about other aspects of services and community life that influenced overcoming adversity, allowing for other information to emerge according to the interests, experiences, and views of the interviewees [35]. All interviews were audio recorded with consent and transcribed verbatim. Interviewees' ages ranged between 18 to >65 years. Of the 30 participants, 11 (37%) identified as men and 19 (63%)

as women. All Australian states were represented, but none of the participants were from Northern Territory or Australian Capital Territory.

Data Analysis

Forum post and interview data were analyzed using a thematic qualitative analysis method [43]. First, data were analyzed deductively against the Theoretical Resilience Framework (refer to Table 4 for themes in data coded to resilience factors). Inductive analysis was also applied, to explore and discover (1) themes about challenges experienced by rural people and (2) themes about forum features that enable benefits for users. To increase the reliability of data analysis, 4 researchers (JF, SK, PK, and AM) initially read the data, noting themes independently. Agreement was then reached on an initial codebook outlining what would be included or excluded for each thematic code. Then, PK, AM, and KC systematically coded the forum posts and interview transcripts independently using NVivo (QSR International). These coded data were then discussed with the wide research team, and agreement was reached on any inconsistencies. At this point, the data were collected as rich bodies of posts and quotes for each theme. These were checked, and refinements were made regarding allocation according to theme by AS by rereading all coded material data. Agreement about inconsistencies was reached via discussion among the whole writing team. Some quotes or posts were allocated to >1 theme because content overlapped themes. In the findings, we have provided an overview of key insights along with illustrative forum posts (marked as Beyond Blue, ReachOut, and SANE) and interview quotes (using pseudonyms for anonymity).

Table 4. Data coded to resilience factors.

Resilience factor	Description of data coded to this resilience factor
Social capital	This category included descriptions about social connection, relatedness, or empathy between people, for example, expressions of friendship and friendly encounters such as peers offering encouragement to each other, referring to each other as friends, thanking peers for trusted friendships, and describing the value of these friendships.
Sense of belonging	This category included (1) people joining and posting messages seeking to belong to this community, for example, introducing themselves and telling their story; (2) posts that tell others that they belong; and (3) posts that reference the forum as a beneficial place, perhaps discussing features of the forum as if it is a physical place, and testimonials about forum benefits.
Learning and knowledge	This category included (1) giving knowledge, for example, strategies for coping with symptoms, including from lived experiences; (2) requests for advice or information, for example, how to access services; and (3) sharing web-based resources.
Self-efficacy	This category included descriptions relating to control or loss of control: (1) expressions relating to using the forum as a step to gain or regain control; (2) asking others on the forum to "hold them accountable"; and (3) where people tell their story to unload, sometimes with discussion suggesting that this helps to move on.
Adaptive capacity	This category included descriptions about how interacting on the forum has changed their activity, attitude, feeling, or knowledge.

Developing a Logic Model

Logic models are descriptions of the chain of causes and effects leading from an intervention to an outcome or outcomes of interest, that is, in our case, individual rural resilience. Logic models identify, describe, and arrange critical aspects of an intervention to represent how the intervention produces change, with arrows used to indicate causal relationships between the aspects [44]. They are a useful tool, as they offer an explicit visual statement of the activities that will cause change and the results expected when implementing a specific intervention. As such, they assist in identifying essential project resources and support planning. They can inform the design of future interventions and policies [45,46].

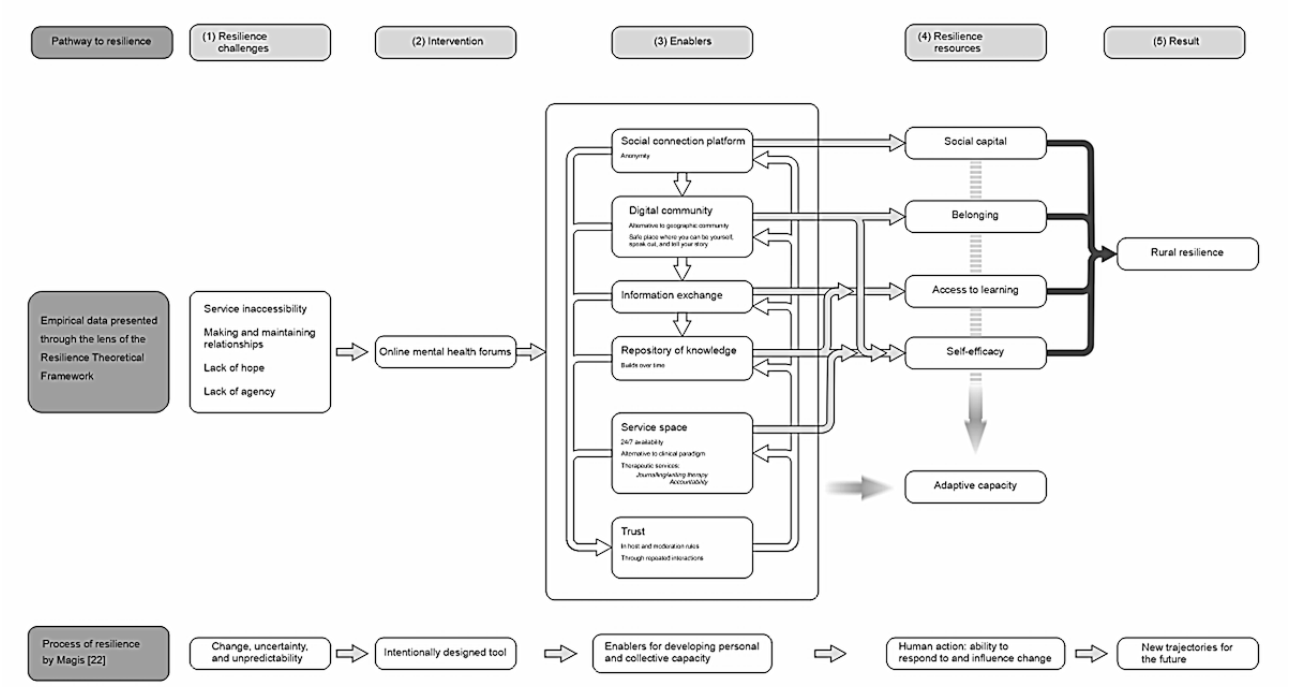
To develop a logic model depicting a pathway by which forums help to build rural resilience, we adopted an abductive approach. As explained by Timmermans and Tavory [38], "theory construction" using abductive analysis requires a "dialectic between data and generalisation as a way to account for empirical findings." That is, it involves a "back and forth" between data and theory to establish whether the data are a case of x. In our study, we focused on the patterns and links between the factors and enablers of resilience in the Theoretical Resilience Framework, the empirical data about these, and any evidence of causality.

Results

Overview

In this section, we first applied inductive analysis to forum post data to reveal the challenges faced by rural people experiencing mental ill-health that they report on forums. Second, we presented findings from deductive analysis applying the Theoretical Resilience Framework to show resilience resources from forum use. We used data from interviews to confirm that forum users acknowledge that they gain these resources. Third, by applying inductive analysis to interview data, we identified features of forums raised in interviews that show how forums enable resilience to build. Wherever we have provided forum posts as examples, the spelling and grammar has been kept exactly as in the posts. These are indicated by the name of the forum and the number of the post, for example, "Beyond Blue; post 1011." In the *Discussion* section, we applied evidence from the *Results* section to develop a logic model that outlines a pathway showing how forums as an intervention can enable resilience resources to be built (Figure 1).

Figure 1. Pathways to rural resilience [22]. Please see the high-resolution version of this figure in Multimedia Appendix 1.



Challenges Experienced

We grouped the challenges identified in the forum data into themes relating broadly to service inaccessibility, making and maintaining relationships, and lack of hope and agency. These are discussed in this section.

Users often paired mentions of their rural living with problems of *inaccessibility* to a spectrum of services including general practitioners, psychologists, counselors, mental health nurses, and psychiatrists. Challenges included nonexistence of some needed services, distance, lack of transport options, costs, and restricted choice—a crucial issue if the person had poor experiences with an available practitioner [47]. Illustrative posts include the following:

> *Today I drove 3 hours to the doctors to get my psych re-evaluation. I am falling apart. It's official, I'm experiencing anxiety at unprecedented levels. To get there wasn't easy.* [Beyond Blue; post 1011]

> *I am on the waiting list to DBT [Dialectical Behavior Therapy] but live in the middle of nowhere so that's not an option right now.* [SANE; post 3]

How distance and inaccessibility exacerbated experiences of mental ill-health was frequently referenced; for example, a user described living on a cattle station 500 km from the closest town and struggling to cope with a new baby (Beyond Blue; post 3). Another resident of a small town referred to the debilitating impact on their agoraphobia, of trying to reach a specialist located >2 hours away (SANE; post 4).

Users discussed that relationships could be difficult to navigate in the context of isolation, with challenges in *making new relationships and navigating existing problematic relationships* affecting people's mental health and ability to thrive and cope:

> *I have a few [friends] who live far away from me, but I am so lonely in the small town I moved to for work. I can't talk to someone unless they talk to me first, and even then, I can't look them in the eye.* [Beyond Blue; post 27]

> *When you live in a small village, everyone knows everything about everyone. You can't date, you can't go to a support group...because everything that you say will be repeated.* [Beyond Blue; interview 3]

Among the analyzed posts, there was a recurrent discussion about how to maintain relationships with friends and family while not "burdening" them with issues associated with users' mental ill-health. These quotes illustrate that—at times—it is not isolation per se, rather concerns about the well-being of others, that prevents people from sharing their experiences and feelings:

> *My wife is my best friend and I do talk to her about it, I have mates but I don't want to burden them with my issues.* [Beyond Blue; post 21]

> *I tried talking to a counsellor at school about it but she didn't help. I can't talk to my mother about it as she already has enough on her plate. I can't tell my friends about it as I don't want to burden them with my baggage. I just need someone to talk to.* [Beyond Blue; post 823]

Forum users talked about *lack of hope* and fears about how they would manage their condition in the future. They questioned what their life might be like going forward and expressed challenges with knowing what to do and how to achieve *the agency to make changes happen*:

> *Feeling like there's just no point to your life. That happens to me a lot because I'm a carer to two special kids. And the prospects of my own life and my babies'*

> *lives, I just wish that they were better but they're not. It is what it is. And sometimes that gets on top of me.* [SANE; interview 1]

> *I am terrified that I will project my insecurities onto my children and cause damage to them in some way, and I am terrified that my issues will ruin my marriage or hurt my husband. I hate being this way, but I truly don't know how to change.* [Beyond Blue; post 93]

Some of these challenges align with the findings of international studies and reviews about the experience of mental ill-health in rural areas. Studies tend to focus on inadequate mental health services [25,48] and experiences of stigma and lack of confidentiality [49]. An issue that is less often raised, but surfaces in forum data, is the personal emotional pain of trying to create satisfying relationships in isolated places and of the inescapability of negative relationships. Studies of lesbians, gay men [50], and transgender individuals in rural places confirm the challenges of "being different" [51]. Findings about feeling burdensome reflect more generic mental health experiences (ie, not confined to rural areas) [52], whereas lack of agency, despair, and hopelessness can be compounded in rural places where appearances must be maintained and there are few support groups [53]. This situation of social stigma and lack of places to relate and be meaningfully occupied could be linked to high rural suicide rates [54].

Resilience Resources

By applying the Theoretical Resilience Framework, findings from the deductive analysis show how resilience resources are generated *for* and *by* users through their interactions on forums.

Building *social capital* can be seen through exchanges of posts in which people connect one-to-one, expressing support, caring, and identifying each other personally. These connections were made in specific threads of the forums, and we found evidence of ongoing conversation—from introductions to dialogues about personal feelings:

> *I'm glad things are starting to feel a little better for you and hope they continue that way. How is the assignment coming along? How amazing is it going to be when we're all finished for the year?!* [ReachOut; post 37]

> *First, I want you to know, that you're not alone in this world. It may not feel like it, but you have many people who love you, very very much- you just can't see it. You are a wonderful person, with an amazing life– you raised children, had a long relationship.* [Beyond Blue; post 100]

Interviews confirmed that new social connections were made on the web. Although some of the connections were relatively abstract, as presented in the following quote, others turned into a form of friendship, with forum users expressing concerns about each other:

> *I've met people on the forums. And they will post something and tag me into the post. You know, I get a notification that someone's mentioned me in a post. So, there's that kind of camaraderie that's built up.* [SANE; interview 1]

The idea of *belonging*, generated via engaging on the forums, was regularly expressed. Posts indicate that the value of being part of the forum as a community sometimes helped to overcome challenges of finding belonging in their geographical locale. Some users stated how, in a rural context, it is hard to find friends who are empathetic and prepared to challenge their own community identity to make people experiencing mental ill-health feel included. Users expressed relief when discovering that multiple people—on the web—understood what they were going through:

> *I don't really have a support network. I live away from my family and have no friends. Because of this I can't get out and meet people. But reading on here and knowing I'm not alone and I can come on here and talk to people who have or are going through the same as me has helped already.* [Beyond Blue; post 400]

Interviewees discussed how a sense of belonging is created on forums. This was often about existing users reaching out to new people to express understanding, support, and solidarity:

> *They weren't telling me to ignore...things. It was more– "we know what you're going through, we know how hard this is. We've been there." "We're here for you, talk as much as you want. Don't worry about thinking what you're saying is trivial. Or we think you're just being stupid or that you're just being a pest for coming on here and saying how terrible you feel." It was the support from people who've been there before.* [Beyond Blue; interviewee 4]

Interviewees referred to belonging being reinforced when they reached a point where they reciprocated support by giving help to others on forums. As such, our data show that belonging involves sufficient embedding and integration, that is, socialization [55], that a user is comfortable and confident to give and to receive assistance:

> *I find it helps me to be there and supporting other people. I tend to stick to areas where I know other people who've been on it, when they're going through a rough patch to support them and vice versa. I try to support newbies, sometimes just by saying I'm listening...It's more two ways now. I find it helps me being there for other people as well.* [Beyond Blue; interview 4]

Access to *learning and knowledge* is manifested in forum post data via direct and indirect requests for information and advice and users sharing information and advice in response or proactively. Information shared may include "factual" information about services, therapies, conditions, and experiences. Examples of such "factual" knowledge posts are as follows:

> *Another option is a Police and Ambulance Intervention Plan. This is a document that can be created to record what strategies would be useful for*

> *the police or ambulance officers attending to you in a mental health crisis.* [SANE; post 117]
>
> *Use aromatherapy and lavender is very calming. I have used it in a diffusor on my skin and on the palms of my hands rubbed together and inhaled it but never heard of the capsules. Would be interested to hear how they go.* [Beyond Blue; post 237]
>
> *I feel overwhelmed by small adult tasks and I am a huge stress head over almost everything. Please tell me it gets better? Or that I am at least not the only one?* [SANE; post 75]

Sometimes, users posted their viewpoints and opinions. Dialogues that involved series of questions and responses and reciprocal relationships in exchanging knowledge or advice were confirmed in interviews; for example, a forum user said the following:

> *Initially, I was using it, to seek help from other people, I guess, thinking that there might be someone who's been through what I've been through. Over time, it became about giving help and information* [Beyond Blue; interviewee 3]

As such, the interview data prove the idea of a generated (online) community of interest of people who support each other by sharing information within an environment of support.

Self-efficacy was most clearly evidenced when users talked about gaining control or acknowledging the loss of control in aspects of their lives. Interacting in forums enabled people to express these feelings or self-assessments. In turn, this helped them to process their situation and how they got there and to set goals for themselves about where they want to be—in their future life or emotional state. Participants mentioned that the digital community helped them to "get things off their chest" and, in ways, to hold themselves accountable:

> *I was starting to lose hope that it can and will get better but I have to try remain positive that I can overcome this and give the medication time to fix the chemical imbalance as well as wait until I can see a mental health professional to learn coping strategies.* [SANE; post 14]

Being unable to speak about emotions with family, some people started using forums to cope with negative feelings as a way of taking control. For example, forums were frequently used for a form of journaling and as a platform for "letting things out" to ease off a negative state of mind:

> *Lost three important people this year. Living in a small Aboriginal community it's hard to find someone who's not a family member to just chat to. It's hard to chat with family because they're grieving too. I don't know how to feel better anymore and it is starting to affect my home life and my work. I just wanted to be heard.* [Beyond Blue; post 115]

An interviewee explained how this element of being able to process things on the forum had become an important assistive tool:

> *It's hard with that pressure to talk to someone because it's difficult to articulate things, even in the moment. I need more time. So, I found that it was a better option to be able to sit there and take my time to write down things...So, as I'm writing I'm exploring, and I can do more, through the process of writing. I tend to come to a sort of light bulb moment in my understanding of myself.* [Beyond Blue; interview 2]

Adaptive capacity was coded when there was an indicated decision or change of activity, attitude, or mental state. In some posts, users discussed navigating life challenges and making decisions to help tackle the causes of these challenges:

> *Hi [username], thanks for ur [sic] reply. Yep, life seems to be picking up for me. Made a decision last week to leave the station [been here 3 yrs on my own]– going back to family farm.* [Beyond Blue; post 79]

Another aspect of adaptive capacity was expressed as awareness of personal issues and how to cope with them by acknowledging ongoing symptoms. This can include informing other users about strategies that worked:

> *I'm okay. I went away for two weeks with my family which kept me distracted and busy...it was kind of good to have a break but it can only last for a short time before I feel like I'm losing myself again, I get antsy and agitated and just need to get back to a space where I'm back on track with my discovery journey.* [Beyond Blue; post 415]
>
> *I was speaking to my doctor today about how much the conversation with you had changed my mood yesterday when things were very dark.* [SANE; post 7]

An interviewee explained how processes that they developed via their activity on the online forum had contributed to changing the way they interact with others in a physical setting:

> *Before I respond to something, I might re-read it two or three times or I might kind of read it, walk away and do something and then come back to it before I respond to it. Just so I can take that in. And that definitely spins off into when I'm actually physically speaking to people. I am a much better listener now than what I probably was previously.* [SANE; interview 20]

Therefore, consistent with those of other studies, findings show benefits for forum users from social support and access to knowledge [1]. We have been able to show that users' engagement in forums contributes to addressing specific rural challenges and that the resources they build for themselves and others can be assessed as contributing to their rural resilience resources. Rural people are growing their social connections—giving access to *social capital*—and generating communities of people with shared experiences that help people feel *belonging*. Forums give users access to repositories of *learning* and enable them to gain control of their lives (*self-efficacy*). Using forums can lead to life changes, with

adaptive capacity being key to resilience. In the following section, we discuss what it is about these forums that enables resilience to build.

Resilience-Enabling Features

Interview data were useful for identifying features of forums that enable rural people to gain resources. These features mean that people use forums to meet their needs including obtaining assistance to solve problems. By being used to address peoples' needs, we argue, forums can help to build resilience resources as shown previously. We now describe the enabling features discussed in user interviews.

Anonymity is described by users as making it easy to ask for advice and exchange information. Web-based anonymity may be particularly helpful in the rural context where people risk stigma if they "open up" to known people. Interviewees also suggested that avoiding in-person interactions can make it easy to start discussions and use them purposefully:

> *I've seen it's mostly men that feel like they have to hide their true feelings in real life. And I guess, that's just some subconscious programming from society. But with the anonymous forums, they can truly open up with what they truly feel as there's no repercussions for them because no one knows who they are.* [SANE; interview 12]

> *Personally, for me, it [anonymity] really does [matter] because I don't want somebody knowing the intricate details of what I might be going through, I like to try and pretend that I'm okay. So, I don't want my name splashed everywhere. I do like being anonymous, but I also feel like that it can create those relationships between people.* [ReachOut; interview 4]

> *I've actually had anxiety myself, which I was diagnosed with three years ago. And in that time, I was looking all over the place for information and support. And I found online was more helpful for me because I didn't know anybody.* [ReachOut; interview 1]

Providing an *alternative community* to their geographical community was noted as important by interviewees. This is depicted as a place where "people are not judging you" or "passing rumours around" (SANE; interviewee 7). An interviewee said the following:

> *I can go to my family when I'm ruminating or down about something or I've got a problem. But I don't want to cross their boundaries either. Yeah, so that's why it's good to have kind of two communities that can support me there.* [ReachOut; interviewee 2]

The people in this alternative community are particularly valuable because they are empathic owing to having similar mental health experiences:

> *At many mental health services, you're just scared of telling the truth because of how they might react to it. The people that are meant to help you...they just want to do stuff to you and control you. Other people [on the forum] have been through it. They get it. They just let you be who you are.* [SANE; interview 7]

Forums are valued as spaces where *you can be yourself, speakout, and tell your story* in a safe environment. Forums also facilitate *information exchange* by encouraging questions and people giving feedback; they facilitate customized responses and language that users understand. Forums were discussed as *repositories* that collect and organize specialized information that can be used at times of different needs:

> *I actually registered quite a long time ago...to view rather than write, to see other people's experiences and to assist my understanding of my own condition. And to see how other people responded and get through their struggles.* [SANE; interview 6]

> *It's a mixture of everything sort of coming together, because you might have one question, and then the next day have another question that's kind of not related. But the answers you get, give you a way forward.* [Beyond Blue; interview 5]

Forums enable people to obtain help *at the times when theyneed it*, without having to wait to access on-premises services. An interviewee said the following:

> *It's somewhere where I can kind of get help when I need it.* [ReachOut; interview 1]

Therapeutic benefits from using forums were also reported. Interviewees described writing on the forum as having a "journalling" function [56], where users can record things to unburden themselves in the moment and to help process difficult issues and emotions—as the following quotes highlight:

> *I found that it was a better option to be able to sit there and take my time to write down things. And that in itself is quite therapeutic, and also been widened to do it in a more targeted way. So that I'm on going to a thread that I relate to.* [Beyond Blue; interview 2]

> *It's more that sort of short term– there's something that's really bothering me at the time, and I can't move past it. [Using the forum] helps get me over that. And then I can bring back in the CBT that I've been doing. And kind of work from there. But it's just getting over that initial hurdle. Sometimes there's something that's bothering me and I can't move past it until I've just spoken it out with someone.* [ReachOut; interview 3]

Interviewees used the forum to post about goals or changes, thus creating a level of *accountability* for personal actions:

> *Being part of the forum made me more accountable to myself...I was going to do this course, blah, blah. And then I had a bit of a think about it while they were talking and realised that I'd been telling myself that for seven years...Anyway, recently, I've managed to enrol in it. And it's actually happening.* [SANE; interview 19]

Interactions on forums are experienced as authentic, with responses from "real people":

> *I have found that really good. It's a bit more personal. I do feel like I've had an interaction afterwards.* [ReachOut; interviewee 2]

Interviewees described how forums work as *an alternative to the clinical paradigm* because "people are so incredibly open, it's like looking into their life...it's written so personally. And everything else that I had been reading was so clinical" (SANE; interview 8).

Regarding the functional benefits from forums, there was repeated discussion about *trust*. This comes from trust in the reputation of the organizations that run forum [57] and trust that is formed through users' ongoing anonymous exchanges. Being on the forums is experienced as a space of consistent, clear rules:

> *I actually trust the forum community more than I trust my local community...there are very few services for mental health here and like the psychiatrist comes once every couple of months so the forum is really important in between times...it's just that I've built that trust over time. Because they've been there for me for so long.* [Beyond Blue; interview 6]

> *This space overall is run by professionals, and they have people that are qualified to deal with everything. So, you know that even if the people on the forum are making you feel a certain way, then the professionals are going to like step in and help you with it.* [ReachOut; interview 2]

As these comments show, interviewees identified forum features that make these web-based spaces particularly useful for meeting their needs as people with experiences of mental ill-health living in rural contexts. We contend that forum features are found to be so helpful that people use the forum for benefits and, *in doing so, build resilience resources*. The forum's features enable users to build "alternative communities" that give safety and connection, and forums facilitate access to useful services that fill gaps and act to complement their mental health service system experiences. In both community and service dimensions, the continuing forum activity is fueled by trust from user interactions within a web-based environment supported and moderated by forum hosts.

Finally, we looked for negative features, but few were raised. Some users said they could not get the specialized information they need (eg, relating to a rare condition). Others commented about challenges arising from moderation practices, such as not being able to specifically reference risks or express anger and, sometimes, when people were banned or suspended, they simply "disappeared" from the forum, causing concern for other users who did not know where they had gone.

Discussion

Principal Findings

This study aimed to explore, identify, and systematize a *pathway* by which moderated online peer support mental health forums can act as an intervention to enable resilience building for rural people experiencing mental ill-health. We found consistent evidence suggesting that users gain access to *resilience resources* from using forums, according to a Theoretical Framework of Resilience that we applied. From interviews, we identified features of forums that enabled people to meet their needs. We understand these features as enablers for building resilience resources through using and interacting on forums.

To illustrate a pathway to resilience, we developed a logic model (Figure 1). A logic model simplifies complex situations, but this "paring down" assists to understand processes associated with specific interventions, while also generating a theory for further testing. Having a way to convey the resilience resources built via an intervention and features that enable these "on a page" is useful for policy makers and practitioners to communicate how initiatives work to effect change (eg, to funders, wide groups of politicians, or communities). Having a model is also useful in highlighting the stages where, in processes of change, assessment could occur. Such assessments could take the form of measurements or checklists (eg, do we see that x, y, and z features are in place because having these should lead to resilience building). Defining stages in this way means applying a logic model that can be useful to guide evaluation or for audit. In our study, via the data sources we used, we also indicate that, increasingly, evaluators can access more diverse sources of data to understand what is happening in complex social processes [42,58]. Here, we used forum post data to show and substantiate the generated resilience resources and a more traditional data source—interviews—to confirm findings and explain causal processes.

Explaining the Logic Model

Overview

The logic model (Figure 1) draws on the findings to show (1) consistent challenges reported by rural forum users; (2) the forum as intervention; and (3) how the forum supports resilience building via its enabling features, with activity reinforced via trust that both builds from ongoing forum interactions and enables future interactions. In the logic model, we grouped the enablers to highlight how they act to formulate types of "enabling spaces" in the forum: a social connection platform, a digital community, an information exchange platform, and a repository of knowledge—with each of these spaces and its features feeding into and supporting the delivery of the next (shown by downward arrows). The forum also provides a space giving access to a range of services. In our conceptualization, the way the forum works to support benefits aligns with the discussion by Duff [27] that mental health coping and recovery are assisted by composite factors and cannot be dependent only on an individual and their response to their condition. Here, the composite factors include the user, other forum users, host organizations, and features enabled by the interplay between people and aspects of the forum technology. In his study, Duff [27] discusses how people experiencing mental ill-health can find alternative spaces where they feel safe. He describes different physical places and the assemblages they comprise. In our study, we are able to show a web-based space that appears to work in supportive and therapeutic ways for rural people who cannot access a physical place of refuge.

Resilience Resources Built

Overview

In this section we show links between enablers and 4 of the 5 resilience resources built. The links with "adaptive capacity" seem less clear. We found evidence of adaptive capacity, but perhaps, it is an emergent feature. Adaptive capacity manifested differently in the forum data (or at least we coded it as a different type of factor) because it is evidenced through data indicating objective change. Adaptive capacity may be a resilience resource or a resilience outcome. Some framings of resilience suggest that adaptive capacity *is* resilience [21,23], whereas others suggest adaptive capacity as a "a latent property, which can be activated when people exercise their agency" [23], suggesting that it is associated with activating self-efficacy. Social learning is also associated with adaptive capacity [59], but learning itself may be antecedent to self-efficacy. Berkes and Ross [23] note that there is a gap in literature explaining how resilience factors relate to each other. Here, we are limited to saying that the resilience factors were observed, but there is inconclusive evidence that some are antecedent to others (note: the logic model does not sequentially follow the sections in the *Results* section because, in analysis, it was first necessary to establish the existence of resilience factors, before identifying enabling features).

To illustrate how the logic model relates to data, we selected themes in findings to explain a pathway to resilience for an imaginary forum user—Jack.

Challenges

Jack has a new diagnosis and is struggling to access information and services in their isolated community. They are losing hope and do not know how to take control.

Intervention

Using a mental health peer support forum is suggested by their mental health nurse.

Enablers

Jack registers and, at first, they observe what other users are saying and perform some searches of the forum threads for topics relevant to their condition. Over time, Jack gains confidence to ask questions. They make friends with users, Happy123 and Hope678, who are supportive, and Jack is able to ask them how they cope and use Jack's own language. After some time, Jack tells them about their own health journey. It is very personal, but the forum is anonymous; therefore, people will never know who Jack really is, and no one in their geographical community will know about these web-based interactions. The forum is a safe place because if anyone posts abusively on the forum, the posts are filtered out. Happy123 and Hope678 are friendly, but they also give Jack coping strategies and support control taking. Jack starts to give themself small challenges such as finding local exercise classes and starting to speak with neighbors. Jack starts to use the forum less for their own needs, but they know it is always there for Jack if their world goes bleak and they need immediate help. Jack also uses the forum to see if they can help and support others. These exchanges make Jack feel as a valued part of a reciprocal community.

Resilience Resources

Jack's access to knowledge and learning, a social network, and belonging to an alternative digital community of shared experience have helped them to gain control. Access to these resilience resources has stimulated Jack's innate capability to adapt.

Outcome

In our terms, we can say that Jack is building their individual rural resilience.

The final line of the logic model compares what has been observed and systematized from the data with the description of resilience by Magis [22]. In a phrase, Magis [22] comments simultaneously about individual resilience, community resilience, and the relationship between individual and community resilience. In relation to our study, we observed what Magis [22] refers to "the existence, development, and engagement of community resources by community members to thrive in an environment characterized by change, uncertainty, unpredictability, and surprise." Magis [22] says that members of resilient communities "intentionally" develop capacity, and it *could* be said that people use the forums with intent—intending initially to help themselves and then to help others. Our findings also resonate with those in the study by Magis [22] in that we see people engaging to develop *personal and collective capacity.* People initially engage to help themselves, but they build a collective, ongoing resource that replenishes, refreshes, and changes—potentially to meet the needs of multiple users. Through our study, we see rural people appreciate forum resources to address challenges of their context. In its organicism and malleability, it is possible to suggest that the forum as a web-based community is resilient; however, for its ongoing existence, it is dependent on external enabling structures—that is, the host organization, its moderation practices, and its ability to generate ongoing funding to enable its continuation [60]. This points perhaps to the difference between a collective of resilient individuals making up a resilient community and the idea that a resilient community requires more than resilient individuals—it also requires high-quality supportive structures and governance [61].

Implications for Academia

Findings of this study suggest that resilience is a composite resource or outcome that can be built through using a web-based intervention if the right conditions (enablers) are in place. Although previous studies show the benefits of online mental health forums [1,3,11], only a small number of them are specifically about rural contexts, and they do not look specifically at *how* online forums relate to resilience. In addition, by establishing resilience resources from online forums for rural people, our paper shows a causal pathway from an intervention to resilience. In doing so, it provides a theory for further testing—relating to *how* resilience can build through types of interventions, in specific contexts and for key groups.

Implications for Policy

The logic model assists in understanding what happens in the processes of resilience building. The logic model is useful for policy, as it summarizes the features of an intervention that needs to be in place to build resilience. It deconstructs the features of the studied intervention and suggests how these lead to the realization of resilience resources. Knowledge gained from this study could inform the creation of tools to support the development of other resilience-building interventions and to assess their impacts. The logic model is useful for policy makers whose time and budgets are limited, as it can be used as an evaluative framework to guide the measurement of the effects of forums as interventions.

The study shows that resilience can be built through an intentional intervention that is web-based and, thus, relatively cost-effective. Further testing of forums and other types of digital platforms as enablers of resilience generation is required. As in our study, such initiatives complement on-premises services and should not be viewed as a replacement for rural services, as we have highlighted the limits of the services they can provide. Mental health peer support forums are useful in providing 24/7 access and some therapeutic services and should be considered as a part of rural health service systems when these are examined. Importantly, our findings suggest online forums as a preventive service readily on hand to help rural residents to cope with daily struggles, support their well-being, and help to build resilience—all contributing to easing the pressure on clinical service infrastructure.

Implications for Practice

Resilience from using online forums can be regarded as a bottom-up phenomenon generated through the labor of many forum users and nonprofit organization staff, harnessing interactions and applying rules about conduct and communication. Through forum interactions, users harness connections and knowledge and play an active role in decision-making and in their own processes of health and recovery. Although having impacts on other users, they influence the course of their own life. Considering this intertwined relationship between forum users, the logic model helps to make sense of the processual nature of developing resilience, its feedback loops, and the interconnections between enabling features and resilience resources. The logic model systematizes factors and resources and their relatedness, identifying and describing critical aspects and showing how—in the case of our study—online forums help to enable changes for people. The abstracted but evidence-based logic model helps to highlight how the collective effects of forum interactions support change for individuals and, potentially, at the community level. This idea of stepping back to see the big picture of how interventions effect change is significant for practitioners in helping them to understand the outcomes of day-to-day practices.

Strengths, Limitations, and Future Studies

Findings result from analyzing data relating to 3 online forums and from 2 types of data—interviews and forum posts. The triangulation enabled verification of the findings and understanding of causality. On the basis of analysis of large data sets, our study suggests consistent findings with respect to resilience resources. In saying this, we limited our analysis to using 1 framing of resilience. Coding was conducted according to our interpretation of the resilience theory and the framework developed for this study. There are other interpretations and frameworks of resilience.

Forum users are anonymous; therefore, we cannot comment about their sociodemographic profile. Interviewees were a mix of ages and self-reported gender and from a mix of locations across Australia. There were more interviewees who used SANE forums compared with the others, potentially because SANE interviewees were sought during the COVID-19 lockdowns, whereas others were sought after the lockdowns ceased.

Although there are indications that forums help people to engage more in their geographical community, we are unable to say that they transfer their resilience resources to make their geographical community more resilient. So—in turn—although people on forums may “develop new trajectories” [22] for their and the web-based communities’ future, we cannot comment about the impact of the forum on the future of the geographical community.

Findings relate to resilience for a specific group of people in context; therefore, they are not transferable to other peer support forums and contexts. Future studies might examine the transferability of resilience built via mental health forums, for dealing with emergent challenges, for example, disaster preparedness.

Conclusions

The study aimed to explore, systematize, and show a pathway by which specialized online peer support forums can act as an intervention to assist resilience building for rural people experiencing mental ill-health.

By developing a new, evidence-based Theoretical Resilience Framework, we show the resilience resources built. Furthermore, via the mixed methods deployed, we reveal the features of forums that enable them to be contextually suitable. From this, we can see that, in using forums, the rural user group built resilience resources as a by-product of meeting their most immediate needs.

We used evidence to inform a logic model showing a pathway from forum as intervention to resilience resources built via the forum’s enabling features. This illuminates new understanding of the causal mechanisms of forums that can be used by forum host organizations and policy and can inform those developing rural mental health systems about the role of forums and about the gaps in formal public services they appear to fill.

Forums work by harnessing users’ collective labor, enabled by organizations that are attuned to peoples’ needs but also operate according to a set of clear guidelines and practices, enabling trust. There are lessons to be learned about the roles that web-based services can play in contemporary health systems and more widely about how health organizations can operate to facilitate services in responsive partnerships with users, rather than simply providing services according to outdated formulas.

Conflicts of Interest

None declared.

Multimedia Appendix 1
High-resolution version of Figure 1.
[PNG File , 693 KB - mental_v10i1e47459_app1.png]

References

1. Smith-Merry J, Goggin G, Campbell A, McKenzie K, Ridout B, Baylosis C. Social connection and online engagement: insights from interviews with users of a mental health online forum. JMIR Ment Health 2019 Mar 26;6(3):e11084 [FREE Full text] [doi: 10.2196/11084] [Medline: 30912760]
2. Latkovikj MT, Popovska MB. Online research about online research: advantages and disadvantages. E Methodol 2020 May 28;6(6):44-56. [doi: 10.15503/emet2019.44.56]
3. Prescott J, Hanley T, Ujhelyi K. Peer communication in online mental health forums for young people: directional and nondirectional support. JMIR Ment Health 2017 Aug 02;4(3):e29 [FREE Full text] [doi: 10.2196/mental.6921] [Medline: 28768607]
4. Repper J, Carter T. A review of the literature on peer support in mental health services. J Ment Health 2011 Aug;20(4):392-411. [doi: 10.3109/09638237.2011.583947] [Medline: 21770786]
5. Shalaby RA, Agyapong VI. Peer support in mental health: literature review. JMIR Ment Health 2020 Jun 09;7(6):e15572 [FREE Full text] [doi: 10.2196/15572] [Medline: 32357127]
6. Colomer-Lahiguera S, Ribi K, Dunnack HJ, Cooley ME, Hammer MJ, Miaskowski C, et al. Experiences of people affected by cancer during the outbreak of the COVID-19 pandemic: an exploratory qualitative analysis of public online forums. Support Care Cancer 2021 Sep 11;29(9):4979-4985 [FREE Full text] [doi: 10.1007/s00520-021-06041-y] [Medline: 33576878]
7. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
8. Saravanakumar M, SuganthaLakshmi T. Social media marketing. Life Sci J 2012;9(4):4444-4451.
9. Sillence E. Giving and receiving peer advice in an online breast cancer support group. Cyberpsychol Behav Soc Netw 2013 Jun;16(6):480-485. [doi: 10.1089/cyber.2013.1512] [Medline: 23659724]
10. Lovatt M, Bath PA, Ellis J. Development of trust in an online breast cancer forum: a qualitative study. J Med Internet Res 2017 May 23;19(5):e175 [FREE Full text] [doi: 10.2196/jmir.7471] [Medline: 28536093]
11. Tsai JC, Hung SY. Examination of community identification and interpersonal trust on continuous use intention: evidence from experienced online community members. Inform Manag 2019 Jun;56(4):552-569. [doi: 10.1016/j.im.2018.09.014]
12. Batenburg A, Das E. Emotional approach coping and the effects of online peer-led support group participation among patients with breast cancer: a longitudinal study. J Med Internet Res 2014 Nov 28;16(11):e256 [FREE Full text] [doi: 10.2196/jmir.3517] [Medline: 25474819]
13. Labrague LJ. Psychological resilience, coping behaviours and social support among health care workers during the COVID-19 pandemic: a systematic review of quantitative studies. J Nurs Manag 2021 Oct;29(7):1893-1905 [FREE Full text] [doi: 10.1111/jonm.13336] [Medline: 33843087]
14. El-Den S, Schneider C, Mirzaei A, Carter S. How to measure a latent construct: psychometric principles for the development and validation of measurement instruments. Int J Pharm Pract 2020 Aug;28(4):326-336. [doi: 10.1111/ijpp.12600] [Medline: 31944468]
15. Markantoni M, Steiner AA, Meador JE. Can community interventions change resilience? Fostering perceptions of individual and community resilience in rural places. Community Development 2019 Jan 15;50(2):238-255. [doi: 10.1080/15575330.2018.1563555]
16. Skerratt S. Enhancing the analysis of rural community resilience: evidence from community land ownership. J Rural Stud 2013 Jul;31:36-46. [doi: 10.1016/j.jrurstud.2013.02.003]
17. Currie M, McKee A, Glass J, Markantoni M, Pinker A, McMorran R, et al. Understandings and applications of rural community resilience amongst Scottish stakeholders: introducing dual discourses. Community Development 2022 Sep 07;54(2):187-205. [doi: 10.1080/15575330.2022.2120509]
18. Steiner A, Atterton J. Exploring the contribution of rural enterprises to local resilience. J Rural Stud 2015 Aug;40:30-45. [doi: 10.1016/j.jrurstud.2015.05.004]
19. Wilson GA. Community resilience, globalization, and transitional pathways of decision-making. Geoforum 2012 Nov;43(6):1218-1231. [doi: 10.1016/j.geoforum.2012.03.008]

20. Buikstra E, Ross H, King CA, Baker PG, Hegney D, McLachlan K, et al. The components of resilience-perceptions of an Australian rural community. J Community Psychol 2010 Oct 06;38(8):975-991. [doi: 10.1002/jcop.20409]
21. Brown K, Westaway E. Agency, capacity, and resilience to environmental change: lessons from human development, well-being, and disasters. Annu Rev Environ Resour 2011 Nov 21;36(1):321-342. [doi: 10.1146/annurev-environ-052610-092905]
22. Magis K. Community resilience: an indicator of social sustainability. Soc Natural Resources 2010 Apr 05;23(5):401-416. [doi: 10.1080/08941920903305674]
23. Berkes F, Ross H. Community resilience: toward an integrated approach. Soc Nat Resour 2013 Jan;26(1):5-20. [doi: 10.1080/08941920.2012.736605]
24. Roberts E, Anderson BA, Skerratt S, Farrington J. A review of the rural-digital policy agenda from a community resilience perspective. J Rural Stud 2017 Aug;54:372-385. [doi: 10.1016/j.jrurstud.2016.03.001]
25. Kelly D, Steiner A, Mazzei M, Baker R. Filling a void? The role of social enterprise in addressing social isolation and loneliness in rural communities. J Rural Stud 2019 Aug;70:225-236 [FREE Full text] [doi: 10.1016/j.jrurstud.2019.01.024] [Medline: 31787802]
26. Lawrence-Bourne J, Dalton H, Perkins D, Farmer J, Luscombe G, Oelke N, et al. What is rural adversity, how does it affect wellbeing and what are the implications for action? Int J Environ Res Public Health 2020 Oct 01;17(19):7205 [FREE Full text] [doi: 10.3390/ijerph17197205] [Medline: 33019735]
27. Duff C. Atmospheres of recovery: assemblages of health. Environ Plan A 2015 Sep 04;48(1):58-74. [doi: 10.1177/0308518x15603222]
28. Steiner AA, Farmer J. Engage, participate, empower: modelling power transfer in disadvantaged rural communities. Environ Plan C Politics Space 2017 Apr 06;36(1):118-138. [doi: 10.1177/2399654417701730]
29. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: meta-narrative reviews. J Adv Nurs 2013 May 29;69(5):987-1004 [FREE Full text] [doi: 10.1111/jan.12092] [Medline: 23356699]
30. Salkind NJ, editor. Encyclopedia of Research Design. Thousand Oaks, California, United states: SAGE Publications; 2010.
31. Snyder H. Literature review as a research methodology: an overview and guidelines. J Business Res 2019 Nov;104:333-339. [doi: 10.1016/j.jbusres.2019.07.039]
32. Norris FH, Stevens SP, Pfefferbaum B, Wyche KF, Pfefferbaum RL. Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. Am J Community Psychol 2008 Mar;41(1-2):127-150. [doi: 10.1007/s10464-007-9156-6] [Medline: 18157631]
33. Fletcher D, Sarkar M. Psychological resilience. Eur Psychologist 2013 Jan;18(1):12-23. [doi: 10.1027/1016-9040/a000124]
34. Hegney D, Buikstra E, Baker P, Rogers-Clark C, Pearce S, Ross H, et al. Individual resilience in rural people: a Queensland study, Australia. Rural Remote Health 2007 Oct 22. [doi: 10.22605/rrh620]
35. Valentine G. Tell me about . . .: using interviews as a research methodology. In: Methods in Human Geography. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 1997.
36. Maclean K, Cuthill M, Ross H. Six attributes of social resilience. J Environ Plan Manage 2013 May 14;57(1):144-156. [doi: 10.1080/09640568.2013.763774]
37. Leite M, Ross H, Berkes F. Interactions between individual, household, and fishing community resilience in southeast Brazil. Ecol Soc 2019;24(3). [doi: 10.5751/es-10910-240302]
38. Timmermans S, Tavory I. Theory construction in qualitative research. Sociol Theory 2012 Sep 10;30(3):167-186. [doi: 10.1177/0735275112457914]
39. Steiner A, Markantoni M. Unpacking community resilience through capacity for change. Community Development J 2013 Aug 01;49(3):407-425. [doi: 10.1093/cdj/bst042]
40. Australian Statistical Geography Standard (ASGS). Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/statistical-geography/australian-statistical-geography-standard-asgs [accessed 2021-10-23]
41. Nelson KS, Nguyen TD, Brownstein NA, Garcia D, Walker HC, Watson JT, et al. Definitions, measures, and uses of rurality: a systematic review of the empirical and quantitative literature. J Rural Stud 2021 Feb;82:351-365. [doi: 10.1016/j.jrurstud.2021.01.035]
42. Kamstra P, Farmer J, McCosker A, Gardiner F, Dalton H, Perkins D, et al. A novel mixed methods approach for integrating not-for-profit service data via qualitative geographic information system to explore authentic experiences of ill-health: a case study of rural mental health. J Mixed Methods Res 2022 Oct 21:155868982211352. [doi: 10.1177/15586898221135291]
43. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
44. Hills D. Logic mapping: hints and tips for better transport evaluations. The Tavistock Institute. 2010 Oct. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/3817/logicmapping.pdf [accessed 2023-04-27]
45. Kelly D, Steiner A, Mason H, Teasdale S. Men's Sheds: a conceptual exploration of the causal pathways for health and well-being. Health Soc Care Community 2019 Sep;27(5):1147-1157 [FREE Full text] [doi: 10.1111/hsc.12765] [Medline: 31206945]

46. Steiner A, McMillan CH, Hill O'Connor C. Investigating the contribution of community empowerment policies to successful co-production- evidence from Scotland. Public Manag Rev 2022 Feb 03:1-23. [doi: 10.1080/14719037.2022.2033053]
47. Steiner A, Calò F, Shucksmith M. Rurality and social innovation processes and outcomes: a realist evaluation of rural social enterprise activities. J Rural Stud 2023 Apr;99:284-292 [FREE Full text] [doi: 10.1016/j.jrurstud.2021.04.006] [Medline: 37089783]
48. Munoz SA, Steiner A, Farmer J. Processes of community-led social enterprise development: learning from the rural context. Community Development J 2014 Nov 14;50(3):478-493. [doi: 10.1093/cdj/bsu055]
49. Cheesmond NE, Davies K, Inder KJ. Exploring the role of rurality and rural identity in mental health help-seeking behavior: a systematic qualitative review. J Rural Mental Health 2019 Jan;43(1):45-59. [doi: 10.1037/rmh0000109]
50. Lyons A, Leonard W, Bariola E. Mental health and resilience among rural Australian lesbians and gay men. Rural Society 2015 Nov 19;24(3):244-265. [doi: 10.1080/10371656.2015.1099268]
51. Smith AJ, Hallum-Montes R, Nevin K, Zenker R, Sutherland B, Reagor S, et al. Determinants of transgender individuals' well-being, mental health, and suicidality in a rural state. Rural Ment Health 2018 Apr;42(2):116-132 [FREE Full text] [doi: 10.1037/rmh0000089] [Medline: 30333896]
52. Cousineau N, McDowell I, Hotz S, Hébert P. Measuring chronic patients' feelings of being a burden to their caregivers: development and preliminary validation of a scale. Med Care 2003 Jan;41(1):110-118. [doi: 10.1097/00005650-200301000-00013] [Medline: 12544548]
53. van Spijker BA, Salinas-Perez JA, Mendoza J, Bell T, Bagheri N, Furst MA, et al. Service availability and capacity in rural mental health in Australia: analysing gaps using an Integrated Mental Health Atlas. Aust N Z J Psychiatry 2019 Oct;53(10):1000-1012. [doi: 10.1177/0004867419857809] [Medline: 31250654]
54. Casant J, Helbich M. Inequalities of suicide mortality across urban and rural areas: a literature review. Int J Environ Res Public Health 2022 Feb 25;19(5):2669 [FREE Full text] [doi: 10.3390/ijerph19052669] [Medline: 35270369]
55. Haski-Leventhal D, Bargal D. The volunteer stages and transitions model: organizational socialization of volunteers. Human Relations 2008 Jan 01;61(1):67-102. [doi: 10.1177/0018726707085946]
56. Ullrich PM, Lutgendorf SK. Journaling about stressful events: effects of cognitive processing and emotional expression. Ann Behav Med 2002 Aug;24(3):244-250. [doi: 10.1207/S15324796ABM2403_10] [Medline: 12173682]
57. McCosker A, Maddox A, Albury K, Dietzel C, Barratt M. Digital cultures of care, safety and wellbeing. AoIR Selected Papers Internet Res 2020 Oct 05. [doi: 10.5210/spir.v2020i0.11138]
58. Kang YB, McCosker A, Kamstra P, Farmer J. Resilience in web-based mental health communities: building a resilience dictionary with semiautomatic text analysis. JMIR Form Res 2022 Sep 22;6(9):e39013 [FREE Full text] [doi: 10.2196/39013] [Medline: 36136394]
59. Goldstein BE. Skunkworks in the embers of the cedar fire: enhancing resilience in the aftermath of disaster. Hum Ecol 2007 Nov 7;36(1):15-28. [doi: 10.1007/s10745-007-9133-6]
60. Vanderhoven E, Steiner A, Teasdale S, Calò F. Can public venture capital support sustainability in the social economy? Evidence from a social innovation fund. J Business Venturing Insights 2020 Jun;13:e00166. [doi: 10.1016/j.jbvi.2020.e00166]
61. Vaneeckhaute LE, Vanwing T, Jacquet W, Abelshausen B, Meurs P. Community resilience 2.0: toward a comprehensive conception of community-level resilience. Community Development 2017 Sep 09;48(5):735-751. [doi: 10.1080/15575330.2017.1369443]

Abbreviations

ASGS: Australian Statistical Geography Standard

Edited by J Torous; submitted 21.03.23; peer-reviewed by M Mazzei, M Markantoni; comments to author 06.04.23; revised version received 09.05.23; accepted 11.05.23; published 28.06.23.

Please cite as:
Steiner A, Farmer J, Kamstra P, Carlisle K, McCosker A, Kilpatrick S
Online Mental Health Forums and Rural Resilience: Mixed Methods Study and Logic Model
JMIR Ment Health 2023;10:e47459
URL: https://mental.jmir.org/2023/1/e47459
doi:10.2196/47459
PMID:37379080

Original Paper

Patient Experience of Digitalized Follow-up of Antidepressant Treatment in Psychiatric Outpatient Care: Qualitative Analysis

Matilda Hamlin[1,2], MD; Joacim Holmén[2], RN, MSc; Elisabet Wentz[1,2], MD, PhD; Harald Aiff[1,2], MD, PhD; Lilas Ali[2,3], RN, PhD; Steinn Steingrimsson[1,2], BSc, MD, PhD

[1]Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[2]Department of Psychiatry, Sahlgrenska University Hospital, Region Västra Götaland, Gothenburg, Sweden

[3]Institute of Health and Care Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Corresponding Author:
Steinn Steingrimsson, BSc, MD, PhD
Institute of Neuroscience and Physiology
Sahlgrenska Academy
University of Gothenburg
Universitetsplatsen 1
Gothenburg, 405 30
Sweden
Phone: 46 31786 00 00
Email: steinn.steingrimsson@vgregion.se

Abstract

Background: Nonadherence to pharmaceutical antidepressant treatment is common among patients with depression. Digitalized follow-up (ie, self-monitoring systems through mobile apps) has been suggested as an effective adjunct to conventional antidepressant treatment to increase medical adherence, improve symptoms of depression, and reduce health care resource use.

Objective: The aim of this study was to determine patients' experience of digitalized follow-up using a mobile app as an adjunct to treatment concurrent with a new prescription, a change of antidepressant, or a dose increase.

Methods: This was a qualitative, descriptive study. Patients at 2 psychiatric outpatient clinics were recruited at the time of changing antidepressant medication. After using a mobile app (either a commercial app or a public app) for 4-6 weeks with daily registrations of active data, such as medical intake and questions concerning general mental health status, individual semistructured interviews were conducted. Recorded data were transcribed and then analyzed using content analysis.

Results: In total, 13 patients completed the study. The mean age was 35 (range 20-67) years, 8 (61.5%) were female, and all reported high digital literacy. Overall, the emerging themes indicated that the patients found the digital app to be a valuable adjunct to antidepressant treatment but with potential for improvement. Both user adherence and medical adherence were positively affected by a daily reminder and the app's ease of use. User adherence was negatively affected by the severity of depression. The positive experience of visually presented data as graphs was a key finding, which was beneficial for self-awareness, the patient-physician relationship, and user adherence. Finally, the patients had mixed reactions to the app's content and requested tailored content.

Conclusions: The patients identified several factors addressing both medical adherence and user adherence to a digital app when using it for digitalized follow-up concurrent with the critical time related to changes in antidepressant medication. The findings highlight the need for rigorous evidence-based empirical studies to generate sustainable research results.

(JMIR Ment Health 2023;10:e48843) doi:10.2196/48843

KEYWORDS

major depressive disorder; digital psychiatry; mobile app; adherence; antidepressant; antidepressants; depressive; depression; mHealth; mobile health; app; apps; application; applications; experience; interview; interviews; medication; prescribe; prescription; dose

Introduction

Depression is the most common mental disorder among adults and the second leading cause of the disease burden worldwide [1]. Standard pharmaceutical treatment is effective compared to a placebo among patients with major depressive disorder [2]. Even so, more than one-third of patients discontinue a new prescription within 3 months [2], and similar rates of discontinuation are also seen when increasing dosage, combining antidepressants, or changing antidepressants [3]. Nonadherence increases the risk of deterioration of psychiatric symptoms, with an increased number of emergency visits and hospitalizations, resulting in not only great suffering at an individual level but also an increased economic burden on the health care system [4,5].

Reasons for nonadherence to antidepressants are often multifactorial. Common factors are negative side effects or a lack of response to antidepressants, patient-related factors (eg, depressive symptoms themselves, psychiatric comorbidity, young age, and low socioeconomic status), and also environment-related factors (eg, shared decision-making and high availability and high continuity of specialized psychiatric care related to changes in medication) [6,7]. How patients handle the first weeks related to a newly prescribed antidepressant is a key factor, which is affected by guidance, information, and shared decision-making with the treating physician [8]. However, on-time high availability and high continuity between appointments to support, guide, and inform a patient are difficult factors to offer in a nonautomated manner. Novel and creative tools to increase self-care beyond the traditional health care setting may increase medical adherence to antidepressants at critical times.

The rapid implementation of digital psychiatry during the COVID-19 pandemic [9] will most likely continue to increase because of the potential postpandemic, pent-up demand for psychiatric care [10]. Beyond the broad implementation of digitalized health care and telecommunications, there is still potential for digitalized solutions with varying complexity and goals for health management to enhance psychiatric care, meaning that digital technology is used to fundamentally transform how health care is delivered and managed. A digital self-monitoring system can collect both active and passive data, whereby psychiatric care can be offered on demand, and it enables a patient to self-monitor the data collected, which can contribute to improved self-awareness [11]. However, it is important to develop sustainable systems suitable for both the health care provider and patient needs, while, at the same time, complying with existing regulations [12]. Self-monitored mobile apps as an adjunct to usual care can be an effective strategy in antidepressant treatment. Some studies report that a follow-up after changes in antidepressant treatment using mobile apps to report self-estimations of mood, sleep quality, and intake of medicine not only increases adherence to the pharmaceutical treatment [13,14] but also mitigates the severity of depression [15,16]. As an example, Corden et al [14] reported in a pilot study (N=11) how a mobile app together with a digital pill dispenser adjunct to initiation of psychopharmacological resulted in a mean medication adherence of 82%. The digital intervention included a reminder of medical intake in the absence of registration through the pill dispenser, weekly assessments of depression severity and medical side effects, graphical feedback of weekly assessments, and access to didactic lessons [14]. Furthermore, relatively easy interventions delivered via mobile phones, such as automatic reminders, can increase the user's adherence to antidepressant medication [17], and this is well tolerated by patients [18]. However, other studies imply that reminders alone do not have a significant effect on adherence to medications and that additional efforts may be necessary to further increase adherence [19].

Engagement with and adherence to using a mobile app itself over time is a key factor that must be considered. Most people stop using an unguided mental health app 2 weeks after download [20]. Additionally, engagement may be difficult, specifically among patients with depression, given the symptomatology of a lack of motivation, a lack of interest, and impaired cognition, such as loss of memory [21]. Previous research has shown that factors promoting increased adherence to digital apps among patients with depression include reminder notifications, ease of use, and increased self-awareness. The latter can be achieved by continuously following and reviewing their own progress in an ecological context, which can increase motivation [21,22]. Furthermore, by creating an extra communication channel with a health care provider, the user's engagement can increase and strengthen the therapeutic alliance between patient and clinician [22].

To summarize, digitalized follow-up is an innovative method to augment treatment at critical times among patients with depression. It may optimize health care resources by increasing medical adherence, improve symptoms of depression, and reduce health care resource use. However, several aspects must be considered before the dynamic and rapidly developing field of digitalized solutions in psychiatric care can be broadly implemented in clinical settings. The aim of our qualitative study is to elucidate patient experiences of using digitalized follow-up concurrently with a newly prescribed, increased dosage or a switch of antidepressant medication.

Methods

Study Design

This is a qualitative, descriptive study using individual semistructured interviews to collect data. Two different technical solutions were tested and interviews conducted thereafter. The method used for analyzing data is an exploratory qualitative content analysis design [23] with the intention to gain a deeper understanding of the topic and develop a clearer direction for future research.

Participants and Settings

Eligible patients at 2 public health care psychiatric outpatient clinics located in southwest Sweden were asked to participate in the study either by a treating nurse or by a physician. No further advertisement of the study was conducted.

Patients were included if they had a new prescription, a change of antidepressant, or a dose increase. They also had to have mastered both spoken and written Swedish language. Exclusion

criteria were patients in need of inpatient care due to depression, acute suicidality, or psychotic symptoms. All exclusion criteria were individually assessed by the patient's treating specialist in psychiatry.

Digital Intervention

Two different mobile apps were used to approach the study aim. One app was developed by a product company, and the other app used an existing e-service platform in public health care. The apps were tested and incorporated in the study design to minimize the risk of focusing on the experience of a specific mobile app and to, instead, broaden the research scope to examine the experience of using a mobile app as a method of follow-up.

Patients were assigned to use 1 of the 2 different mobile apps for 4-6 weeks prior to a semistructured follow-up interview. A research assistant, nurse, or treating physician trained in the use of the mobile apps instructed patients in how to use them. The health care provider was not obliged to monitor the patients' recorded data, concerning which the patients were informed when enrolling in the study.

Commercial Mobile App

The first mobile app is called SENO (Medilevel), which is specifically designed to provide safe initiation of psychopharmacological treatment by offering daily remote monitoring and self-management [24]. The app was authorized and accepted by the IT department of the Västra Götalandsregionen (VGR) health care region. The log-in procedure includes 2-factor authentication with a 4-digit number received in a text message as a code. The app includes a pop-up reminder function at a self-chosen time.

Patients recorded data actively every day concerning intake of medicine, side effects according to a preset list of common side effects, and a grading of 3 questions concerning their general mental health status (energy, interest in things, and anxiety). The 3 questions were presented as 3 time plots, enabling a visual presentation of change over time. In addition, the Montgomery-Åsberg Depression Rating Scale—Self-Assessment (MADRS-S) was completed once weekly [25]. One-way text communication was possible if a health care provider wished to contact a patient through the mobile app.

Public Mobile App and E-Service

The second mobile app is incorporated into the existing and well-established e-service, 1177.se [26], which is used by all public health care providers in Sweden. The e-service can at present offer various features, such as medical advice, personal information about prescriptions or medical records, and a possibility to message your health care provider. The 1177.se app was specifically created for this study and replicated from the commercial app as closely as possible. The log-in procedure is with digital personal identification. A pop-up reminder function is possible but only at 10:00 A.M. Similar to SENO, patients recorded data actively every day concerning intake of medication, side effects according to a preset list of common side effects, and a grading of 3 questions concerning their general mental health status (energy, interest in things, and anxiety). The 3 questions were also presented as 3 time plots, and the app also enabled a time plot of weekly MADRS-S scores. Two-way text communication has already been incorporated in the e-service.

To summarize, the apps differed in following areas: The commercial app enables 1-way text communication for health care providers to contact the user, whereas the public app enables 2-way text communication. The time for a reminder is self-chosen in the commercial app, whereas it is at set at 10:00 A.M. in the public app. A time plot of MADRS-S scores is only available in the public app, and the font also differs between the 2 apps. Due to the nature of the study, no comparison was conducted between the 2 apps.

Data Collection

Interviews were performed by either a female medical doctor and PhD student with minor experience in psychiatric care or a male nurse specialized in psychiatry. There was no established relationship between interviewers and patients prior to the study. The patients were interviewed individually after performing 4-6 weeks of digital intervention. In total, 3 interviews were conducted in person and the remaining 10 were conducted over the phone. This was primarily because of societal circumstances due to the COVID-19 pandemic.

The interview was semistructured whereby the patients answered prepared open-ended questions based on the study aim (Multimedia Appendix 1), with each question being followed by an open discussion. The questions were about patient experiences of the mobile app, such as "How did the app affect your medical adherence?" and "How did the app affect your appointment with your treating physician?" The mean duration of an interview was 22 (range 11-31) minutes.

Data Analysis

A qualitative, inductive content analysis with a hypothesis-generating approach was used to find patterns and interpret meaningful content in the data [27]. Descriptive content analysis makes it possible to reach a high level of reliability through the adequacy of the analysis themes and draw valid conclusions from interpreting the data [28]. Due to the nature of the topic, patient statements were expected to be short and without latent content. The data were analyzed manually after being transcribed into text by the first author (MH). The analysis started with MH repeatedly reading all interview texts, forming a sense of the whole, and discussing initial interpretations of the whole with another author (LA). Thereafter, each interview text was divided by choosing meaning units, followed by condensed meaning units linked to the study aim. Next, a code for each condensed meaning unit was completed with a few words that represent the content. Finally, the different codes were distributed into themes and subthemes that represent a meaning found in several codes by MH and LA. All authors discussed the final themes and subthemes of the analysis.

Ethical Considerations

An ethical committee at the VGR approved the study protocol (Dnr: 2020-01640). The study followed the Declaration of

Helsinki principles with respect to research ethics. At a doctor's appointment, patient consent was obtained for study participation after the patients were informed about the study both orally and in written form by a research assistant, treating physician, or treating nurse. All personal data were pseudonymized, and the audio data and transcribed data were stored in accordance with the Swedish General Data Protection Regulations.

Results

Participant Details

In total, 13 patients were interviewed between March 2021 and January 2023. The long recruitment period of nearly 2 years was partly related to the COVID-19 pandemic and partly due to reorganization at the outpatient psychiatric clinic. Two-thirds (n=9, 69.2%) of the patients used the commercial mobile app, and one-third (n=4, 30.8%) the public mobile app and e-service prior to the interviews. The difference between the type of app used was assumed not to affect the study aim; further participants were therefore not recruited to the public mobile app when data completion was fulfilled. The mean age of the patients was 35 (range 20-67) years. The demographic data of each patient is presented in Table 1. The most common reason for using the digital intervention was initiating new medical treatment either as monotherapy or as an add-on (n=6, 46.2%), followed by dose adjustment (n=5, 38.5%), while 1 (7.7%) patient used it for discontinuation and 1 (7.7%) for a switch between 2 medications. All patients were recruited from specialized psychiatric outpatient clinics and had undergone more than 1 previous medical treatment for depression.

We excluded 4 patients prior to the interview (n=3, 75%, did not answer multiple phone calls and text messages, and n=1, 25%, expressed in text an iatrogenic risk of using the app and participating in the study). None of the 13 (100%) patients who were included and completed the study expressed any deterioration in their mental illness due to the study.

Table 1. Characteristics of interviewed patients (N=13).

Participant	Gender	Age (years)	Digital intervention	Digital device	Digital literacy	Medication change
P1	Female	20	Commercial	iPhone	High	New prescription
P2	Female	26	Commercial	Android	High	Switch of medication
P3	Female	24	Commercial	iPhone	High	New prescription
P4	Male	27	Commercial	Android	High	New prescription
P5	Female	30	Commercial	Android	High	Change of dose
P6	Female	21	Commercial	iPhone	High	New prescription
P7	Female	32	Commercial	Android	High	New prescription
P8	Female	36	Commercial	iPhone	High	Change of dose
P9	Male	32	Commercial	NA	High	New prescription
P10	Female	61	Public	iPhone	High	Discontinuation
P11	Male	40	Public	iPhone	High	Change of dose
P12	Male	67	Public	iPhone	High	Change of dose
P13	Male	45	Public	Computer	High	Change of dose

Main Themes

Four main themes emerged during data analysis: *adherence to using a mobile app*, *insight into one's condition*, *support in a health care setting*, and *tailored content*. An overview of themes and subthemes is presented in Table 2, which is followed by a further description and illustration with quotations.

Table 2. Overview of themes and subthemes that emerged from the study data.

Themes	Subthemes
Adherence to using a mobile app	• Daily registration • Medical adherence • Degree of mental illness
Insight into one's condition	• Questions leading to reflection • Visual overview of patterns over time
Support in a health care setting	• Increased understanding for the health care provider • Evaluation of recorded data together
Tailored content	• None

Adherence to Using a Mobile App

This theme describes the challenges of adhering to the mobile app over 4-6 weeks. It highlights the importance of creating a routine and how specific features, such as a systematic reminder and the time consumed for registration, can affect adherence. It also highlights that it can be challenging to adhere to and engage with a daily task beyond everyday activities or routines in a situation of a change in antidepressant treatment.

Daily Registration

The anticipated time to log in, register, and answer the daily questions was from 1 minute up to a few minutes. The main reason for nonregistration was forgetfulness. Therefore, to register every day, it is important to create a routine that fits in with daily life. Three key factors emerged as important if the patients were to adhere to registration every day: a daily reminder, a simple log-in procedure, and a non-time-consuming register of the daily questions. The majority of patients did include the importance of a daily reminder as a key function for daily adherence. As the preferred time of the day to register varied and was inconsistent between patients, and also for the same patient, many requested extra reminders that could vary in number and time each day. A few patients who used the public app did not receive the reminder due to technical issues, and they all maintained that it affected adherence negatively. The same patients also expressed the importance of being able to individually choose the time for a reminder.

Patients described that the log-in procedure with 2-way authentication via a text message was too complicated. The log-in procedure that had to be repeated when closing the app temporarily was frustrating. A few maintained that a simple log-in outweighed the possible benefits of increased data security. Many of those who used the commercial app emphasized that the same digital personal identification system used for the public app would have been a simpler and less time-consuming solution for log-in. Furthermore, the patients preferred that the recording of answers to the daily questions take only up to a maximum of a few minutes in order to maintain daily adherence. Several found the multiple-choice options and the grading of questions as positive and reported that free-text questions were too time-consuming and could therefore affect adherence negatively.

The majority of patients only reflected on the perspective of how to maintain adherence every day during the whole study period. However, 1 (7.7%) participant expressed:

> *To be honest, I think every day is a bit too often. I would say maybe once or twice a week. These types of symptoms don't change very quickly, and one's daily form fluctuates, so I understand if someone wants an average over time. However, I'm not sure if it would be as effective to fill in every day. That's just my personal reflection. It could lead to worse adherence because it's a bit tedious, and I know from experience that it can be challenging.* [Patient 10 (P10), female, 61 years old]

Medical Adherence

The first question in the app every day was "Have you taken your medication?," with a response option of yes or no. A question to all patients therefore was "How has the use of the mobile app affected your medical adherence?" There were 4 (30.8%) patients who expressed a history of poor medical adherence for various reasons, such as forgetfulness and carelessness and also an active decision to not take a drug. A few of them expressed how an expectation when entering the study was to increase medical adherence. All these 4 (30.8%) patients showed increased medical adherence when using the mobile app, because of the notification to use the app and because the first question reminded them of their medication and whether they had or had not taken it. Even though medical adherence was not expressed as an issue among the other patients, an extra reminder was appreciated and not found to be unnecessary. No one reported worsened medical adherence due to the mobile app.

> *At times, I've been very careless with my medications...but I think this app has helped me...especially with this reminder to fill in the daily measurements, which also helps me remember my evening medication...there has been no carelessness during this period except for a few days when I genuinely forgot but that always happens.* [P3, female, 24 years old]

Degree of Mental Illness

Patients described adherence to the app as decreasing in situations in which they were feeling worse because of their depression. Using the app could feel like too big of an effort due to the loss of energy, and daily registration on the app, even taking only 1 or 2 minutes, could cause a setback in mood if one was having a bad day linked to their mental illness. Similarly, minor technical issues, such as a complicated log-in procedure or the patient getting logged out unexpectedly, could also cause frustration toward the app if the illness worsened. A few patients expressed that adherence to the app was good at first, but as their mental illness subsequently worsened over a few weeks, so too did the use of the app. The reasons included impaired memory and a feeling that nothing in the app would make them feel better or increase their motivation.

> *I know from my own issues that even the smallest things can feel difficult when you're struggling. Having to schedule and do something every day when you don't have the energy for anything can be overwhelming. Even though it's the smallest thing...* [P7, female, 32 years old]

One further aspect reported by a patient was that she chose the timing of registration based on how she felt during the day. Since mental illness was perceived as way worse in the morning, she actively chose not to register at that time. Another patient reported how the lack of motivation toward using the app was constant since there was no improvement in depression throughout the whole study period.

Insight Into One's Condition

This theme describes the experience of how the content can contribute to self-awareness and insight into one's depression. The gain in self-awareness also had a perceived positive impact on the adherence to the app. Daily registration and additional content contributed to daily reflection and an understanding of their depression in several cases but did not necessarily lead to their own sense of improvement in the severity of depression.

Questions Leading to Reflection

To designate a few minutes every day and reflect on their own mental health status by answering the questions with respect to grading of mood, interest, and anxiety was a new experience for several of the patients. Many preferred to record during the evening as it made it possible to reflect on and summarize the day. The questions contributed to a reflection on not only *how* they were feeling on a specific day but also *why* they were feeling better or worse that day. To ask themselves how they were feeling and mapping it through the app made the patients feel good about themselves. It contributed to a feeling of doing something for their own sake and investing in something that would help them. The recording of answers had to be honest for it to be useful to them.

A participant experienced increased body awareness by asking themselves simple daily questions of mood, interest, and anxiety. However, another participant had not reflected at all on the recorded data, and yet another participant reported that daily recording on a bad day could have a reinforcing negative effect.

> *I have gained more, in a way, understanding or that I can see a connection with maybe things that have happened or how it has been, or so. It's not strange that I'm completely exhausted when my mood goes one way or another.* [P6, female, 21 years old]

Visual Overview of Patterns Over Time

All patients who used the commercial app had a positive experience of the statistics function where the daily recordings were visualized as simple time series plots. The visual presentation of recorded data was described as interesting, meaningful, helpful, fun, valuable, supportive, and a way to solidify their depression. It was a valuable tool to help remember how they were feeling a week or several weeks ago and contributed to a concrete overview of patterns over the study period. The expectations to evaluate fluctuations in their depression over time increased motivation and adherence to the app, and several patients expressed how the statistics function was the reason they continued daily registration. Several requested more advanced diagrammatic functions, including additional self-selected variables.

The 4 (30.8%) patients who used the public app had difficulties in finding the statistics function and wished the function had lit up and been easier to find. However, they all maintained that such a visual function was something they had not experienced before and that it could be a potential help in their mental illness. In addition, 1 (7.7%) of the patients experienced the statistics function for the first time during the study interview. Interestingly, the self-perceived experience did not correspond to what the time series plot showed. When studying the time plot of MADRS-S scores, the patient also realized that he did not know how to interpret the score, even though he had filled out the form several times over previous years.

> *Let's see, okay. Well, it has actually changed more than I thought when I look at it now. It actually has. It's not how I experience it, but yes it has.* [P13, male, 45 years old]

Support in a Health Care Setting

This theme involves how the collected data were found to strengthen the patient-physician relationship. A potential use of the recorded data could be earlier contact with the patient and a strengthening of effort if the health care provider detects any abnormal recordings. The data recorded over time could also deepen understanding, contribute to a holistic view, and improve joint decision-making with a patient.

Increased Understanding for the Health Care Provider

Several patients expressed that the continuously recorded data were an improvement over usual treatment and could increase understanding by the health care provider. The statistics function was specifically emphasized as valuable to help the physician understand and obtain an overall picture of how depression had unfolded during the time between appointments. Eventual fluctuations of variables (mood, energy, and anxiety) could be presented visually without relying on the patient to remember, recall, and evaluate what information might be important for the treating physician. Furthermore, the physician's understanding of depression could be enhanced even further if tailored content was possible, where patients could continuously add individual information that they considered valuable.

Furthermore, 1 (7.7%) patient retold a previous experience at a psychiatric outpatient clinic when her treating physician was temporarily replaced by someone else. The 2 physicians had different opinions and treatment recommendations based on the side effects that the patient experienced. The patient therefore thought that continuously recorded data over time could be valuable to physicians when taking over the treatment of a patient from another physician. Data beyond medical journals that are recorded by the patient themselves could help a new physician understand how the mental illness had fluctuated previously and what side effects there had been in relation to the treatment administered.

> *And for the doctor in this aspect as well, being able to see how it has fluctuated over time.* [P11, male, 40 years old]

Evaluation of Recorded Data Together

A digital communication with booking of appointments and video consultations was already a natural part of the outpatient clinics' working approach. Even though there was already a way to message the outpatient clinics digitally (through 1177.se), the patients experienced difficulty in quickly getting in touch through this facility. Support and help from the outpatient clinics were sometimes not enough. The majority of patients believed that expanding the use of a mobile app as a standard when changing antidepressants could be a positive improvement, even

though increased human contact would be the best alternative. Furthermore, those who used the public app expressed the benefits of the app being incorporated in 1177.se, mainly because all digitally provided health care would be gathered in 1 app and because they were already familiar with that app.

Patients expressed how data could be used as a potential marker of deterioration and a reason for earlier contact with a patient. The fact that all data were available and potentially evaluated by a health care provider in a timely manner increased adherence to the mobile app. Patients who used SENO expressed that a 1-way message function as an additional way to maintain contact with the outpatient clinics was positive. Patients considered the 1-way communication as a tool for the health care provider to potentially remind the patient to register or to quickly respond to abnormal data. However, there was an understanding and acceptance that quick 2-way communication was probably not possible, given the current lack of personnel and resources in psychiatric care.

Even though no one reported they had actually evaluated and discussed the recorded data together with their treating physician, it was still emphasized as a potential asset to the patient-physician relationship. If the physician notices any abnormal data between appointments, they can take the initiative to discuss the data without depending on the patient to retell information. Furthermore, to use the statistics function as support when discussing the patient's mental illness could establish a more equal relationship, where the patient feels more involved and better understands decisions regarding treatment.

> *I think it's very positive because when you meet or have contact, it would be good to have some flesh on the bones, simply put.* [P12, male, 67 years old]

Tailored Content

Tailored content means requesting individual adaptation of functions and content in the app to maximize its potential usefulness. There was mixed experience concerning the app's content. The functions and content were in several cases too simple or too difficult and in some cases relatable or nonrelatable.

Most patients found both the content and graphics simple and clear. The simplicity contributed to the ease of use, with an acceptable use of time and with no difficulties in navigating the app's features. The weekly recording of MADRS-S took a few additional minutes, but the patients had filled out this form several times before and were quite familiar with its contents. Furthermore, they experienced MADRS-S as an important and a comprehensive way to capture their depression.

Despite the ease of use, the majority of patients found the content to be too simple to improve their mental illness and said that a lack of in-depth and detailed content could affect motivation negatively. For the content to improve their mental illness, it was important to be able to relate to and fully understand the meaning of not only the questions asked as a whole but also the specific words used. As an example, 1 (7.7%) patient expressed difficulty in relating to the opposites "a lot of anxiety" and "little anxiety" when grading anxiety on a scale from 1 to 5, assuming that they have anxiety every day.

Additional content was requested, such as more information about side effects, the ability to grade the severity of each side effect, physical status, workload, sleep quality, and unexpected external events that affect mental status. Furthermore, many requested the additional content to be tailor-made, considering what are important and not important factors to a user when mapping their mental illness. For example, 1 (7.7%) patient expressed that all the collected data should be able to act as a visual self-written diary.

> *I would actually like to have it in a way that I could keep my own diary and track other factors that could affect my well-being with medications, such as how physically active I have been and how much I work. That's what I mean when I say I want to make changes...yes. It would be really valuable to track things that one believes affect their well-being, like how often one has had psychotherapy, so that it's not just about the medications…and maybe how one sleeps.* [P10, female, 61 years old]

Discussion

Principal Findings

The main finding of our study resulted in 4 themes that elucidate how patients at psychiatric outpatient clinics experienced digitalized follow-up using a mobile app concurrently with a new prescription, a change of antidepressant, or a dose increase. Adherence to both the mobile app and the antidepressant itself was positively affected by a daily reminder and relatively non-time-consuming daily registration. However, poor adherence to the app was attributed to the severity of mental illness in several cases. Positive experiences with the visual display of collected data in diagrams was a primary finding, which was beneficial for the patients' self-awareness and reflection over their ongoing situation with mental illness. Furthermore, most patients believed that the mobile app, especially the visual display of data, can be a supplement to the patient-physician relationship and contribute to the physician's understanding of their depression. However, none of the patients had actually discussed the content of the app with their treating physician. Finally, the patients had mixed reactions to the app's content: some appreciated the ease of use and simplicity, but the majority requested more advanced and tailored content. There was no obvious deterioration in any patient's mental illness linked to study participation.

The design and content of a mobile app can potentially affect the behavior of patients with depression. To understand the factors affecting engagement with and adherence to a mobile app is therefore a key factor for future implementation in clinical practice. The importance of both reminder notifications [29] and ease of use is in accordance with previous research [30]. However, even though a few patients preferred an easy log-in procedure at the expense of data security, trust in data security is reported as an important factor associated with a positive attitude toward digital health care [31]. Furthermore, studies have reported that several participants linked a higher severity of depression with reduced adherence to the mobile app [32,33]. A large-scale multinational study investigating the utility of

smartphone apps to monitor depression by collecting both active and passive data showed that a higher severity of depression at baseline was associated with less contribution of both active and passive data to a mobile app [33]. The aspect that the severity of depression may affect the use of a mobile app is not surprising, given the core symptoms of depression. Elevated depressive symptoms are a common reason for nonattendance of scheduled appointments in health care; therefore, a mobile app with tailored support strategies might facilitate reestablishment of contact with a health care provider [34].

All patients expressed a willingness to help in the development of psychiatric care as a key factor for participation in the study; however, not many found it to be a potential tool to help themselves. An imbalance between intrinsic and extrinsic motivation has been previously reported as a barrier to user adherence, meaning that users must be informed and well aware of an app's value for themselves and for health care providers [29].

The commercial app in this study was primarily designed to increase medical adherence to antidepressants, and interestingly, there were 4 patients using the commercial app who subjectively experienced increased medical adherence during the study. This is in accordance with previous smaller clinical trials, indicating that mobile apps with a daily reminder and collection of active data can increase medical adherence to antidepressants [14]. However, despite our study results, further large-scale clinical trials with objective measurements are still needed to fully understand how a mobile app can affect medical adherence.

A primary finding of the analysis was the positive experience of the feature with visual time series plots, how it contributed to self-awareness, and how it contributed to patient-physician communications. This is consistent with previous research [35,36]. First, a systematic review of experiences concerning data visualization through digital apps among patients with chronic neurological and mental health conditions reported how visualization of collected data can increase self-awareness, develop communication and understanding with the health care provider, and increase the user's adherence to a mobile app [36]. Furthermore, Scheuer and Torous [35] recently investigated patient perceptions on how useful digitally collected data concerning their mental illness can be if presented visually in various graphs. Simple graphs were reported as being more useful (eg, survey scores); however, by having access to graphs with more complex data, the participants perceived an increased comfort in sharing them with a health care provider. In addition, 89% (25/28) of the participants in that study reported that visual graphs of both active and passive data would contribute to better communication with their treating physician. These results are substantiated by the emerging themes in our study, such as *insight into one's* condition and *support in a health care setting*. If we were to add an extra finding, the majority of our patients found the graphs to be not only helpful in the patient-physician alliance but also as a tool for the physician to gain a deeper understanding and a holistic view of an individual's depression. However, even though most patients experienced the app as a potential support in a health care setting, no one actually discussed the content of the app with the treating physician. It is reasonable to believe that this primarily depended on the nature of the study, where all patients were informed when enrolling in the study that the health care provider was not obliged to monitor the recorded data.

Depression is a heterogenous illness with diverse prominent symptoms and comorbidities, affecting people of all ages with varying digital literacy. Not surprisingly, a theme emerged of the patients desiring tailored content based on their own characteristics and preferences. Especially, there was an urge to tailor the content that was visually presented in graphs. A survey study [37] with open-ended questions investigated which outcome domains are important to patients (n=1912), informal caregivers (n=464), and health care providers (n=627) when evaluating antidepressant treatment. A large number of outcome domains (n=80) emerged from the study results, where many of those, such as daily functioning (eg, workload, active family life, or social life), are not measured in the questionnaires usually assessed in both clinical practice and clinical trials when evaluating antidepressant treatment. Furthermore, a systematic review mentioned previously [36] also reported the importance of customizing visually presented data and that this can be a key factor for long-term user adherence. Promising results of self-tailored content in internet-delivered cognitive behavioral therapy among patients with major depressive disorder have also been reported [38]. Overall, research indicates that customized treatment is important in all phases of treatment in this group.

To summarize, the results we found in our study agree in many ways with previous research concerning the digital use of mobile apps among patients with depression. However, the results contributed 2 additional perspectives we believe can be valuable in future research of using mobile apps as part of antidepressant treatment. First, our findings substantiate previous research that all data collected should be available and visually presented to both patient and physician. A well-informed patient with easy access to all collected data can increase both intrinsic and extrinsic motivation for adherence to treatment, even if the data are actively collected and rather simple, as in our study. The second perspective is that our study procedure reflects the dynamic and somewhat problematic nature of implementing digitalized solutions, such as mobile apps, in a clinical setting. The study aim was based on real-world observations with a supply-demand gap of health care at a vulnerable time of treatment among patients with depression. The patients highlighted several important areas of improvement, and their overall experience was that a digitalized follow-up has the potential to enhance and strengthen the delivery of health care. However, given the dynamic field of psychiatry with digitalized follow-up using mobile apps, the results may become outdated if modifications to the app are made explicitly based on participant experiences before conducting a controlled clinical trial. With endless new possibilities to collect data and create digital biomarkers through patient smartphones, how can we avoid a never-ending rat race in the research of digital psychiatry before implementation in a clinical setting? Studies of patient experiences are important, but likewise, studies of efficacy are important to reflect on how specific elements of digitalized solutions can affect patient behavior. Future research and implementation in clinical practice should rely on rigorous

evidence-based empirical studies, considering the perspectives of multiple stakeholders with the aim to generate sustainable research results.

Limitations

The interviews were conducted by 2 researchers separately (MH and JH) with different credentials and experiences in psychiatric care. The qualitative content analysis was thereafter only conducted by 1 researcher (MH), which can affect the intercoder credibility and trustworthiness of the results. However, the study results can be assumed to be trustworthy and credible, given that the analysis was continuously discussed together with a supervisor with great experience in qualitative research (LA) and subsequently discussed with all authors. Furthermore, since the study collected relatively insensitive data, the risk of recall bias was relatively low and the need for multiple researchers to conduct the analysis was therefore not crucial.

The fact that patients were recruited from specialized psychiatric outpatient clinics must be considered. Previous experiences with and therefore attitudes toward medication and adherence may vary wildly, which may also affect their usage and thoughts about the app. Furthermore, the majority of the patients expressed no previous problems to medication adherence, even though a specific goal of the digital intervention was to improve medical adherence, which could have affected study results.

One further limitation is that the study population was a relatively homogenous group with respect to age and gender. Furthermore, all patients reported high digital literacy, and potentially important information, such as the education level and residential status, was not collected. Lastly, no data were gathered on how many people declined to participate. All the factors mentioned here can affect the variation of the phenomenon and transferability. That is, the generalizability of our results is mainly relevant to patients willing to use digitalized follow-up, which is of importance since the broad implementation of such solutions may make health care harder to access for those who are unwilling to use digitalized follow-up. The implementation of digitalized solutions has a long way to go to supersede regular care, but the potential is enormous [39].

Conclusions

Patients recruited from psychiatric outpatient clinics experienced digital follow-up using a self-monitoring mobile app as a valuable adjunct to antidepressant treatment at the time of medication change. A key finding that emerged from the derived themes was that easy access and visual presentation of collected data may improve medical adherence by promoting self-awareness and improved patient-physician relationship.

There are endless new possibilities to collect data and create digital biomarkers through mobile apps. Future research and implementation in clinical practice rely on rigorous evidence-based empirical studies considering the perspectives of multiple stakeholders, including patient experiences, with the aim of generating sustainable research results.

Acknowledgments

The authors would like to thank Linda Wesén for developing the public mobile app. They also thank Michelle Karlsson and Sara Nylén for recruiting patients. The study was funded by the Västra Götalandsregionen Innovation Fund (VGRINN-941 256). Access to use the commercial mobile app SENO was free of charge. We would also like to thank Peter Todd (Tajut Ltd, Kaiapoi, New Zealand) for third-party editorial assistance in drafting of this manuscript, for which he received financial compensation. MH received financial support from the Gothenburg Society of Medicine (GLS-987104), and SS received funding from the Söderström König Foundation (SLS-974247). The funders had no influence on the conduct or reporting of the study.

Data Availability

Due to local regulations, we are not able to share the data freely, but raw data can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest

SS has received a salary as a scientific consultant for Mindforce but has no commercial interest to declare.

Multimedia Appendix 1
Interview guide.
[DOCX File , 15 KB - mental_v10i1e48843_app1.docx]

References

1. World Mental Health Report: transforming mental health for all. World Health Organization. 2022 Jun 16. URL: https://www.who.int/teams/mental-health-and-substance-use/world-mental-health-report [accessed 2023-04-28]
2. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018 Apr 07;391(10128):1357-1366 [FREE Full text] [doi: 10.1016/S0140-6736(17)32802-7] [Medline: 29477251]

3. Sanglier T, Milea D, Saragoussi D, Toumi M. Increasing escitalopram dose is associated with fewer discontinuations than switch or combination approaches in patients initially on escitalopram 10 mg. Eur Psychiatry 2012 May 15;27(4):250-257. [doi: 10.1016/j.eurpsy.2010.08.009] [Medline: 21036548]
4. Ho SC, Jacob SA, Tangiisuran B. Barriers and facilitators of adherence to antidepressants among outpatients with major depressive disorder: a qualitative study. PLoS One 2017 Jun 14;12(6):e0179290 [FREE Full text] [doi: 10.1371/journal.pone.0179290] [Medline: 28614368]
5. Ho SC, Chong HY, Chaiyakunapruk N, Tangiisuran B, Jacob SA. Clinical and economic impact of non-adherence to antidepressants in major depressive disorder: a systematic review. J Affect Disord 2016 Mar 15;193:1-10. [doi: 10.1016/j.jad.2015.12.029] [Medline: 26748881]
6. Solmi M, Miola A, Croatto G, Pigato G, Favaro A, Fornaro M, et al. How can we improve antidepressant adherence in the management of depression? A targeted review and 10 clinical recommendations. Braz J Psychiatry 2021 Apr;43(2):189-202 [FREE Full text] [doi: 10.1590/1516-4446-2020-0935] [Medline: 32491040]
7. Pompili M, Venturini P, Palermo M, Stefani H, Seretti ME, Lamis DA, et al. Mood disorders medications: predictors of nonadherence - review of the current literature. Expert Rev Neurother 2013 Jul 09;13(7):809-825. [doi: 10.1586/14737175.2013.811976] [Medline: 23898852]
8. Anderson C, Kirkpatrick S, Ridge D, Kokanovic R, Tanner C. Starting antidepressant use: a qualitative synthesis of UK and Australian data. BMJ Open 2015 Dec 30;5(12):e008636 [FREE Full text] [doi: 10.1136/bmjopen-2015-008636] [Medline: 26719312]
9. Kinoshita S, Cortright K, Crawford A, Mizuno Y, Yoshida K, Hilty D, et al. Changes in telepsychiatry regulations during the COVID-19 pandemic: 17 countries and regions' approaches to an evolving healthcare landscape. Psychol Med 2020 Nov 27;52(13):2606-2613. [doi: 10.1017/s0033291720004584]
10. Hamlin M, Ymerson T, Carlsen HK, Dellepiane M, Falk Ö, Ioannou M, et al. Changes in psychiatric inpatient service utilization during the first and second waves of the COVID-19 pandemic. Front Psychiatry 2022 Feb 17;13:829374 [FREE Full text] [doi: 10.3389/fpsyt.2022.829374] [Medline: 35250673]
11. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Feb;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
12. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
13. Lauritsen L, Andersen L, Olsson E, Søndergaard SR, Nørregaard LB, Løventoft PK, et al. Usability, acceptability, and adherence to an electronic self-monitoring system in patients with major depression discharged from inpatient wards. J Med Internet Res 2017 Apr 21;19(4):e123 [FREE Full text] [doi: 10.2196/jmir.6673] [Medline: 28432040]
14. Corden ME, Koucky EM, Brenner C, Palac HL, Soren A, Begale M, et al. MedLink: a mobile intervention to improve medication adherence and processes of care for treatment of depression in general medicine. Digit Health 2016 Aug 03;2:2055207616663069 [FREE Full text] [doi: 10.1177/2055207616663069] [Medline: 29942564]
15. Kordy H, Wolf M, Aulich K, Bürgy M, Hegerl U, Hüsing J, et al. Internet-delivered disease management for recurrent depression: a multicenter randomized controlled trial. Psychother Psychosom 2016 Jan;85(2):91-98 [FREE Full text] [doi: 10.1159/000441951] [Medline: 26808817]
16. Hunkeler EM, Hargreaves WA, Fireman B, Terdiman J, Meresman JF, Porterfield Y, et al. A web-delivered care management and patient self-management program for recurrent depression: a randomized trial. Psychiatr Serv 2012 Nov;63(11):1063-1071. [doi: 10.1176/appi.ps.005332011] [Medline: 22983558]
17. Hammonds T, Rickert K, Goldstein C, Gathright E, Gilmore S, Derflinger B, et al. Adherence to antidepressant medications: a randomized controlled trial of medication reminding in college students. J Am Coll Health 2015 Feb 06;63(3):204-208. [doi: 10.1080/07448481.2014.975716] [Medline: 25338175]
18. Sahm L, MacCurtain A, Hayden J, Roche C, Richards HL. Electronic reminders to improve medication adherence--are they acceptable to the patient? Pharm World Sci 2009 Dec 3;31(6):627-629. [doi: 10.1007/s11096-009-9327-7] [Medline: 19731069]
19. Choudhry NK, Krumme AA, Ercole PM, Girdish C, Tong AY, Khan NF, et al. Effect of reminder devices on medication adherence: the REMIND randomized clinical trial. JAMA Intern Med 2017 May 01;177(5):624-631 [FREE Full text] [doi: 10.1001/jamainternmed.2016.9627] [Medline: 28241271]
20. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
21. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, RADAR-CNS Consortium. Barriers to and facilitators of engagement with mHealth technology for remote measurement and management of depression: qualitative analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]

22. de Angel V, Lewis S, White KM, Matcham F, Hotopf M. Clinical targets and attitudes toward implementing digital health tools for remote measurement in treatment for depression: focus groups with patients and clinicians. JMIR Ment Health 2022 Aug 15;9(8):e38934 [FREE Full text] [doi: 10.2196/38934] [Medline: 35969448]
23. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
24. SENO: Medilevels system för insättning av antidepressiva läkemedel. Medilevel. URL: https://medilevel.se/produkter/seno [accessed 2023-10-05]
25. Svanborg P, Åsberg M. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. Acta Psychiatr Scand 1994 Jan;89(1):21-28. [doi: 10.1111/j.1600-0447.1994.tb01480.x] [Medline: 8140903]
26. 1177 - secure about your health and care. 1177. URL: https://www.1177.se/ [accessed 2023-10-05]
27. Krippendorff K. Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: SAGE Publishing; 2018.
28. Montgomery BM, Duck S. Studying Interpersonal Interaction. New York, NY: Guilford Press; 1991.
29. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
30. Dinkel D, Harsh Caspari J, Fok L, Notice M, Johnson D, Watanabe-Galloway S, et al. A qualitative exploration of the feasibility of incorporating depression apps into integrated primary care clinics. Transl Behav Med 2021 Sep 15;11(9):1708-1716. [doi: 10.1093/tbm/ibab075] [Medline: 34231855]
31. Hamlin M, Steingrimsson S, Cohen I, Bero V, Bar-Tl A, Adini B. Attitudes of the public to receiving medical care during emergencies through remote physician-patient communications. Int J Environ Res Public Health 2020 Jul 20;17(14):5236 [FREE Full text] [doi: 10.3390/ijerph17145236] [Medline: 32698481]
32. Zhang Y, Pratap A, Folarin AA, Sun S, Cummins N, Matcham F, RADAR-CNS consortium. Long-term participant retention and engagement patterns in an app and wearable-based multinational remote digital depression study. NPJ Digit Med 2023 Feb 17;6(1):25 [FREE Full text] [doi: 10.1038/s41746-023-00749-3] [Medline: 36806317]
33. Matcham F, Leightley D, Siddi S, Lamers F, White KM, Annas P, RADAR-CNS Consortium. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022 Feb 21;22(1):136 [FREE Full text] [doi: 10.1186/s12888-022-03753-1] [Medline: 35189842]
34. Bruehlman-Senecal E, Aguilera A, Schueller SM. Mobile phone-based mood ratings prospectively predict psychotherapy attendance. Behav Ther 2017 Sep;48(5):614-623 [FREE Full text] [doi: 10.1016/j.beth.2017.01.002] [Medline: 28711112]
35. Scheuer L, Torous J. Usable data visualization for digital biomarkers: an analysis of usability, data sharing, and clinician contact. Digit Biomark 2022 Sep 12;6(3):98-106 [FREE Full text] [doi: 10.1159/000525888] [Medline: 36471766]
36. Polhemus A, Novak J, Majid S, Simblett S, Morris D, Bruce S, et al. Data visualization for chronic neurological and mental health condition self-management: systematic review of user perspectives. JMIR Ment Health 2022 Apr 28;9(4):e25249 [FREE Full text] [doi: 10.2196/25249] [Medline: 35482368]
37. Chevance A, Ravaud P, Tomlinson A, Le Berre C, Teufer B, Touboul S, et al. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. Lancet Psychiatry 2020 Aug;7(8):692-702. [doi: 10.1016/s2215-0366(20)30191-7]
38. Andersson G, Käll A, Juhlin S, Wahlström C, de Fine Licht E, Färdeman S, et al. Free choice of treatment content, support on demand and supervision in internet-delivered CBT for adults with depression: a randomized factorial design trial. Behav Res Ther 2023 Mar;162:104265 [FREE Full text] [doi: 10.1016/j.brat.2023.104265] [Medline: 36791537]
39. Torous J, Myrick K, Aguilera A. The need for a new generation of digital mental health tools to support more accessible, effective and equitable care. World Psychiatry 2023 Feb 14;22(1):1-2 [FREE Full text] [doi: 10.1002/wps.21058] [Medline: 36640397]

Abbreviations

MADRS-S: Montgomery-Åsberg Depression Rating Scale—Self-Assessment
VGR: Västra Götalandsregionen

Edited by J Torous; submitted 09.05.23; peer-reviewed by H Bowers, M Gasmi , M Richter; comments to author 22.06.23; revised version received 07.09.23; accepted 08.09.23; published 11.10.23.

Please cite as:
Hamlin M, Holmén J, Wentz E, Aiff H, Ali L, Steingrimsson S
Patient Experience of Digitalized Follow-up of Antidepressant Treatment in Psychiatric Outpatient Care: Qualitative Analysis
JMIR Ment Health 2023;10:e48843
URL: https://mental.jmir.org/2023/1/e48843
doi:10.2196/48843
PMID:37819697

Original Paper

Participants' Engagement With and Results From a Web-Based Integrative Population Mental Wellness Program (CHAMindWell) During the COVID-19 Pandemic: Program Evaluation Study

Joseph A Rosansky[1,2], PhD; Kayley Okst[1,3], BA; Miriam C Tepper[1,4], MD; Ana Baumgart Schreck[1], LICSW; Carl Fulwiler[1,2], MD, PhD; Philip S Wang[1,2,5], MD, DrPH; Zev Schuman-Olivier[1,2], MD

[1]Department of Psychiatry, Cambridge Health Alliance, Cambridge, MA, United States
[2]Department of Psychiatry, Harvard Medical School, Boston, MA, United States
[3]Department of Psychology, New York University, New York, NY, United States
[4]New York State Psychiatric Institute, Columbia University, New York, NY, United States
[5]Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, United States

Corresponding Author:
Zev Schuman-Olivier, MD
Department of Psychiatry
Cambridge Health Alliance
1493 Cambridge St
Cambridge, MA, 02139
United States
Phone: 1 617 591 6132
Email: zschuman@cha.harvard.edu

Abstract

Background: The COVID-19 pandemic involved a prolonged period of collective trauma and stress during which substantial increases in mental health concerns, like depression and anxiety, were observed across the population. In this context, CHAMindWell was developed as a web-based intervention to improve resilience and reduce symptom severity among a public health care system's patient population.

Objective: This program evaluation was conducted to explore participants' engagement with and outcomes from CHAMindWell by retrospectively examining demographic information and mental health symptom severity scores throughout program participation.

Methods: We examined participants' symptom severity scores from repeated, web-based symptom screenings through Computerized Adaptive Testing for Mental Health (CAT-MH) surveys, and categorized participants into symptom severity-based tiers (tier 1=asymptomatic to mild; tier 2=moderate; and tier 3=severe). Participants were provided tier-based mindfulness resources, treatment recommendations, and referrals. Logistic regressions were conducted to evaluate associations between demographic variables and survey completion. The McNemar exact test and paired sample *t* tests were performed to evaluate changes in the numbers of participants in tier 1 versus tier 2 or 3 and changes in depression, anxiety, and posttraumatic stress disorder severity scores between baseline and follow-up.

Results: The program enrolled 903 participants (664/903, 73.5% female; 556/903, 61.6% White; 113/903, 12.5% Black; 84/903, 9.3% Asian; 7/903, 0.8% Native; 36/903, 4% other; and 227/903, 25.1% Hispanic) between December 16, 2020, and March 17, 2022. Of those, 623 (69%) completed a baseline CAT-MH survey, and 196 completed at least one follow-up survey 3 to 6 months after baseline. White racial identity was associated with completing baseline CAT-MH (odds ratio [OR] 1.80, 95% CI 1.14-2.84; *P*=.01). Participants' odds of having symptom severity below the clinical threshold (ie, tier 1) were significantly greater at follow-up (OR 2.60, 95% CI 1.40-5.08; *P*=.001), and significant reductions were observed across symptom domains over time.

Conclusions: CHAMindWell is associated with reduced severity of mental health symptoms. Future work should aim to address program engagement inequities and attrition and compare the impacts of CHAMindWell to a control condition to better characterize its effects.

(JMIR Ment Health 2023;10:e48112) doi:10.2196/48112

KEYWORDS

COVID-19 pandemic; digital psychiatry; early identification; integrative medicine; mental wellness; mindfulness; population mental health; prevention; stratified care

Introduction

The COVID-19 pandemic involved a prolonged period of collective trauma and stress during which mental health concerns, such as depression and anxiety, increased across the population [1,2]. During and after collective trauma, the need for acute psychological services is widespread [3], but those affected by disasters (eg, Hurricane Katrina) are often undertreated due to a lack of access to services and stigma associated with psychiatric treatment [4]. Additionally, during the last major pandemic before COVID-19, the Spanish flu pandemic of 1918, increased social distancing was associated with increased suicide rates, irrespective of actual flu mortality rates [5].

These gaps were also observed during the COVID-19 pandemic, with substantial increases in depression and anxiety and a simultaneous reduction in the use of outpatient mental health treatment at the start of the pandemic [2,6]. Furthermore, a meta-analysis of 54 investigations found that rates of suicidal ideation, suicide attempts, and self-harm increased during the pandemic compared to the year before the pandemic began [7]. A variety of factors may have contributed to this, including reduced access to traditional behavioral health treatments due to pandemic precautions and provider burnout, challenges implementing and accessing remote care, and ongoing stigma related to using behavioral health care [8,9]. In this context, there was an urgent need to develop alternative approaches to providing population behavioral health support.

Previous research has identified stratified mental health care as an approach that can maximize care system resources while effectively treating patients' symptoms [10,11]. This approach uses decision tools to assign patients to the interventions best suited to address their specific symptoms. Stepped-care approaches initially assign all patients to similar low-intensity treatments and then increase treatment intensity for those who do not respond, which also appears to reduce costs and promote recovery and has been more commonly used than stratified care [12]. However, stepped-care approaches appear to be less effective for patients with more complex presentations (eg, greater symptom severity, disability, younger age, and unemployment), and a recent cluster-randomized trial of 951 patients in the United Kingdom found that stratified care was associated with improved outcomes [10,13]. These findings could reflect that stratified care approaches avoid delays in getting individuals to their appropriate levels of treatment and do not require patients to fail lower levels of care first.

While stratified care models can help lower costs, the expenses associated with providing behavioral health support to large groups of people can be prohibitive unless cost-effective treatment approaches are used. Providing individuals with educational materials about ways to improve their mental health (ie, passive psychoeducation) is one low-cost, evidence-based approach that has been shown to reduce symptoms of depression [14]. Self-guided behavioral health treatment apps that use cognitive behavioral- and mindfulness-based treatment approaches are another low-cost approach that has been shown to promote symptom reduction [15,16]. In addition, regular remote symptom assessments and personalized feedback after stratification-based treatment assignments appear to be promising tools for improving patient outcomes in mental health care settings, and many aspects of these types of interventions can be automated and incorporated into a measurement-based care approach [17-19]. We, therefore, attempted to incorporate these elements into CHAMindWell, an integrative population mental wellness program using a semiautomated approach incorporating elements of stratified and measurement-based care that was developed in response to the COVID-19 pandemic at a large public safety net health care system in Massachusetts serving a catchment area of 450,000 people. In this program evaluation, we sought to assess participants' engagement with the CHAMindWell program as well as the impacts of program participation on participants' mental health symptoms.

Methods

Program Design and Implementation

Overview

We designed CHAMindWell to prevent the onset or worsening of mental health problems among healthy participants and facilitate early identification and treatment to reduce symptom severity among those with clinical concerns. To access mental health care resources before this program's development, patients needed to be referred to the hospital system's outpatient psychiatry department or primary care behavioral health integration and complete a psychiatric diagnostic assessment in a stepped-care model. To reduce the burden that a population mental wellness program focused on early identification might place on traditional mental health services (eg, individual therapy) while still meeting the treatment needs of participants with more complex presentations, CHAMindWell used a stratified care model. The program included components of symptom monitoring, motivational interviewing, psychoeducation, coaching, app recommendations, safety assessments, referrals to mindfulness training and mindfulness-based group therapies, and referrals to other internal and external treatment sources. We made several mindfulness-based group services and resources available to CHAMindWell participants because these interventions have been shown to effectively reduce stress, anxiety, and depression and increase resilience while being acceptable for web-based delivery [20-22].

Within 1 week after the first COVID-19 lockdown in Massachusetts, we initiated preparations for a pandemic-related mental health surge. We targeted this program toward the hospital system's primary care population and patients who had been placed on waitlists for behavioral health services. To recruit these individuals, we periodically sent them emails and messages

through the electronic health record's patient portal (MyChart) with information about CHAMindWell. We also educated primary care providers about the program and developed a website where potential participants could learn about and register for CHAMindWell [23].

Participants in CHAMindWell completed recurring surveys assessing mental health symptom severity using Computerized Adaptive Testing for Mental Health (CAT-MH) modules [24]. The modules assessed for symptoms of depression, anxiety, posttraumatic stress disorder (PTSD), psychosis, mania or hypomania, and the risk of developing a substance use disorder. CAT-MH used multidimensional item response theory procedures to adaptively select questions from each module's item bank based on a participant's previous responses to efficiently obtain a precise estimate of symptom severity, quantified by a severity score between 0 and 100 [25]. Although a module assessing suicide risk was initially included, we discontinued that module because it was designed to predict the risk of developing suicidality over the next 6 months rather than describe immediate risk [26,27].

To ensure confidentiality, meet hospital information security requirements, and increase efficiency in screening large numbers of individuals, we developed an open-access Research Electronic Data Capture (REDCap) external module to distribute CAT-MH surveys and paired it with a coordinator dashboard to manage daily monitoring of results [28,29]. The coordinator dashboard included data such as completion time, missing assessments, and CAT-MH results for each participant. We also collected participant feedback, both formal (eg, Patient and Family Advisory Committee) and informal (eg, through phone contact with participants), to iteratively improve the user experience and overall program design.

Once enrolled, participants in CHAMindWell were asked to complete CAT-MH survey batteries at baseline and months 1, 2, 3, 4, 6, 9, and 12. In response to participant feedback about the frequency of survey invitations, in June 2022, we reduced the survey invitation frequency to baseline and months 2, 4, 6, 9, and 12. Project coordinators reviewed CAT-MH survey results on a rolling basis and assigned each participant to 1 of 3 tiers based on their most recent results. Symptom severity cut points for each tier varied by CAT-MH module, and participants were assigned an overall tier equal to their highest tiered domain (eg, depression or anxiety). Participants' tiers were updated each time they completed a CAT-MH survey, and they were provided individually tailored, tier-based services.

Tier 1 Interventions

Participants with low-level symptom severity scores that did not meet clinical thresholds (eg, severity score <50; specific cut points varied by module) were assigned to tier 1. These participants were offered phone support from a coordinator trained in motivational interviewing, including guidance on using a clinician-reviewed list of self-help apps and websites, information about free access to 8-week mindfulness courses (5-10 free spots available in each of 12 groups per year), and community-based mindfulness programs offered through the hospital's Center for Mindfulness and Compassion. Participants were also sent monthly email newsletters that included information about CHAMindWell services and psychoeducation on rotating mental wellness topics, such as seasonal changes in affect, impacts of diet on mental health, behavioral changes to improve sleep, and related subjects.

Tier 2 Interventions

Participants with moderate to moderately severe symptoms (eg, severity score between 50 and 75) were assigned to tier 2. They received all interventions offered to those in tier 1 and were also offered a 20-minute web-based mental wellness check-in with a clinician. These mental wellness check-ins were used to confirm symptoms indicated by CAT-MH scores (eg, anxiety, depression, psychosis, and mania) and to further evaluate participants' treatment needs in the context of their unique circumstances. Clinicians evaluated for safety concerns, reviewed participants' most recent CAT-MH survey results, discussed current symptoms, and provided brief coaching and app recommendations. Based on participants' needs and interests, clinicians were able to provide referrals to relevant clinical trials and group-based mental health services. Clinicians were encouraged to refer these participants to group-based services unless symptoms indicated a clear need for higher levels of care or they voiced strong preferences for individual treatment.

Tier 3 Interventions

Participants endorsing severe mental health symptoms (eg, severity score ≥75) were assigned to tier 3. These participants were provided all the interventions offered to those in tiers 1 and 2, but clinicians were encouraged to also offer tier 3 participants referrals to traditional behavioral health services (eg, individual therapy and psychopharmacological treatment). Participants in need of urgent or specialty care (eg, suicidal ideation, psychosis, severe substance use, or severe mania) were referred to appropriate hospital-based treatment programs.

This program evaluation assessed participants' engagement with and outcomes from the CHAMindWell program. We evaluated (1) whether demographic variables were associated with different levels of program engagement, (2) whether participants' symptoms improved over the course of program participation, and (3) which types of referrals were placed for participants with clinically relevant mental health symptoms.

Study Design

We retrospectively analyzed demographic and mental health symptom data collected from program participants who enrolled between December 16, 2020, and March 17, 2022, through the CHAMindWell consent and registration form and completed CAT-MH surveys. To preserve the methodological integrity of relevant clinical trials, participants who elected to join a trial were removed from CHAMindWell until their research participation concluded.

Participants

Individuals were eligible to enroll in CHAMindWell if they were 18 years or older, a primary care patient or staff member at Cambridge Health Alliance, and spoke English fluently. English fluency was required for enrollment because participants

needed to be able to comprehend and complete web-based surveys in English.

To examine the effects of CHAMindWell on mental health symptoms, we analyzed data from participants who completed both a baseline and a follow-up survey between 3 and 6 months later. Because some participants did not complete certain scheduled CAT-MH surveys and others did not open and complete surveys until weeks or months after receiving invitations, we used responses completed at least 75 days after baseline and as close to 180 days after baseline as possible for each participant's follow-up assessment.

Measures

Demographic Information

Participants were asked to provide their ages, gender identities, and racial and ethnic backgrounds. Participants were able to select more than one racial and ethnic background from a provided list (Table 1).

Table 1. Participant characteristics by program engagement status. Participants could select more than one racial and ethnic background. Besides the association noted in the table, logistic regressions did not reveal any significant associations between demographic characteristics and follow-up status.

Variable	Enrolled (n=903)	Completed baseline (n=623)	Remained enrolled after baseline (n=536)	Completed 3-6-month follow-up (n=196)
Gender, n (%)				
Female	664 (73.5)	460 (73.8)	393 (73.3)	144 (73.5)
Male	196 (21.7)	131 (21)	120 (22.4)	43 (21.9)
Genderqueer	20 (2.2)	17 (2.7)	11 (2.1)	3 (1.5)
Prefer not to say	14 (1.6)	6 (9.6)	4 (0.7)	2 (1)
Missing	9 (1)	9 (1.4)	8 (1.5)	4 (2)
Racial and ethnic background, n (%)				
White[a]	556 (61.6)	413 (66.3)	350 (65.3)	139 (70.9)
Black, African American, or Haitian	113 (12.5)	63 (10.1)	57 (10.6)	18 (9.2)
Asian or Indian	84 (9.3)	60 (9.6)	50 (9.3)	19 (9.7)
Native American or Alaskan Native	7 (0.8)	3 (0.5)	2 (0.4)	1 (0.5)
Other	36 (4)	22 (3.5)	19 (3.5)	5 (2.6)
Hispanic	227 (25.1)	152 (24.4)	128 (23.9)	41 (20.9)
Missing	41 (4.5)	24 (3.9)	22 (4.1)	5 (2.6)
Age (years), mean (SD)	39.4 (13.0)	39.6 (12.7)	39.6 (12.8)	42.0 (13.2)

[a]The variable was significantly associated with increased odds of completing baseline CAT-MH; $P<.05$.

Framework of CAT-MH

Each CAT-MH [30,31] module included a large, domain-specific item bank from which items were adaptively selected using multidimensional item response theory procedures. Symptom severity scores ranging from 0 (asymptomatic) to 100 (greatest severity) were calculated for each module. Most modules assessed symptom severity within the past 2 weeks. However, the substance use disorder module measured the risk of having or developing a substance use disorder over the next 6 months rather than current symptoms of substance use [32]. This program evaluation examined participants' baseline and follow-up scores for the depression, anxiety, and PTSD modules because we felt these domains were most likely to be impacted by mental wellness support.

Data Analysis

To evaluate whether participants' demographic characteristics were associated with degrees of program engagement, we conducted 2 logistic regressions. The completion of at least one CAT-MH survey was the dependent variable for the first regression, and the completion of at least one CAT-MH survey 3-6 months after baseline was the dependent variable for the second. Both dependent variables were dummy-coded (0=not completed and 1=completed). For both regressions, age, gender, and racial and ethnic identities were included as independent variables.

Overall, 19.7% (178/903) of participants were missing data for age or gender. Missing data were due to participants not answering questions and changes made to the enrollment form as the program evolved (eg, participants who enrolled before March 16, 2021, were not asked to provide their ages). Therefore, we performed each regression twice, first using only complete cases and then imputing missing data using the *mice* (version 3.14.8; multivariate imputation by chained equations) package in R statistics (R Core Team) [33]. This package used fully conditional specification to separately impute missing values for each independent variable used in the regressions. We used predictive mean matching to impute missing values for age (173/903, 19.2% missing) and polytomous logistic regression for gender (9/903, 1% missing). We repeated this procedure to create 5 complete data sets, performed logistic regressions with each data set, and pooled the results for final analyses [34].

To evaluate the mental health impacts of participating in CHAMindWell, we performed a McNemar exact test estimating the change in participants' odds of having tier 1 (nonclinically relevant) symptoms between baseline and 3-6-month follow-up. We also conducted paired sample *t* tests to evaluate within-subject changes in anxiety, depression, and PTSD symptom severity scores between baseline and follow-up because we felt those symptom domains were the most likely to be improved through participation in CHAMindWell. Because participants in tiers 2 and 3 had more acute needs for symptom reduction and received more interventions than those in tier 1, we repeated these *t* tests for the subsample of participants assigned to tiers 2 and 3 at baseline.

To evaluate whether symptom reductions may have been caused by the natural easing of pandemic-related restrictions and stressors over time, we tested for associations between symptom severity scores and time since the COVID-19 lockdown that surveys were completed among baseline (n=623) and follow-up surveys (n=196). To that end, we conducted 6 regression analyses (3 symptom domains × 2 survey populations) using CAT-MH severity scores for depression, anxiety, and PTSD as the dependent variables and the number of days after COVID-19 lockdown that surveys were completed as the independent variables. We also examined outcomes from a 6-month sample of mental wellness check-ins conducted between September 2020 and March 2021 and calculated the proportions of referrals made to different types of services. All analyses were conducted in RStudio (version 2022.07.2; R Core Team) [35].

Ethical Considerations

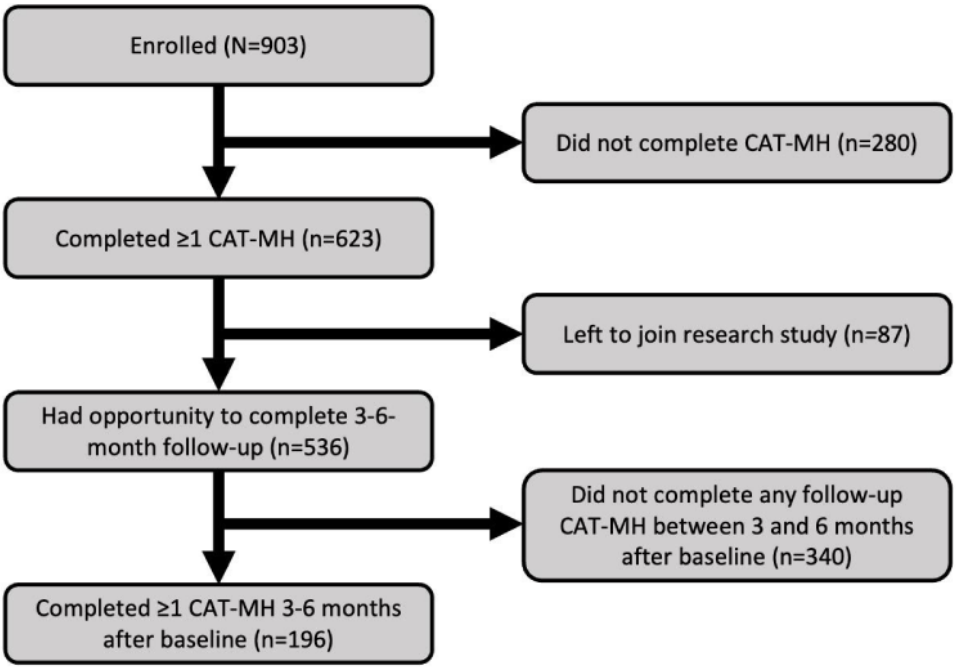
The Cambridge Health Alliance institutional review board office indicated that this program evaluation met the criteria for a quality improvement initiative that did not require informed consent.

Results

Overview

As Figure 1 illustrates, 623 (69%) of the 903 participants who enrolled in CHAMindWell completed at least 1 CAT-MH assessment, and 87 of those transitioned out of CHAMindWell to join a research study before they had the opportunity to complete a 3-month follow-up survey. Of the 536 who had the opportunity, 196 (36.6%) completed a 3-6-month follow-up CAT-MH survey. Table 1 displays the demographic characteristics of participants by program engagement status.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram for CHAMindWell program evaluation. CAT-MH: Computerized Adaptive Testing for Mental Health.



Program Engagement

Logistic regressions revealed that White racial identity was the only variable significantly associated with completing a baseline CAT-MH assessment in both complete case analysis (odds ratio [OR] 1.87, 95% CI 1.13-3.10; *P*=.02) and after imputing missing data (OR 1.80, 95% CI 1.14-2.84; *P*=.01). Using complete case analysis, age was the only variable associated with completing follow-up, and the effect size was notably small (OR 1.03, 95% CI 1.01-1.04; *P*=.003). After imputing missing data, no demographic variables were significantly associated with completing follow-up.

Overall Symptomatology

At baseline, 108 (55.1%) of the 196 participants who went on to complete a 3-6-month follow-up survey had clinically significant symptoms in at least one domain and were assigned to tier 2 or 3. Mean baseline severity scores for each domain were: $mean_{depression}$ 49.4 (SD 17.0), $mean_{anxiety}$ 40.8 (SD 19.3), $mean_{PTSD}$ 41.7 (SD 17.4), $mean_{mania/hypomania}$ 24.6 (SD 20.6), $mean_{substance\ use\ disorder}$ 44.0 (SD 15.0), and $mean_{psychosis}$ 30.9 (SD 14.2). Of the 108 in tier 2 or 3 at baseline, 39 (36.1%) no longer had clinically significant symptom severity in any of the domains measured at follow-up. A significant McNemar exact

test revealed that participants' odds of having mental health symptom severity below the clinical threshold were improved by 2.6 times at follow-up (OR 2.60, 95% CI 1.40-5.08; *P*=.001).

Changes in Anxiety, Depression, and PTSD Symptomatology

As Table 2 illustrates, there were significant reductions in depression, anxiety, and PTSD symptom severity across the entire sample and among those reporting clinically significant symptoms at baseline (ie, tiers 2 and 3). Regression analyses revealed no significant associations between symptom severity and survey completion date among baseline or follow-up surveys for depression (baseline *P*=.31; follow-up *P*=.12), anxiety (baseline *P*=.47; follow-up *P*=.97), or PTSD (baseline *P*=.99; follow-up *P*=.10).

Table 2. Changes in Computerized Adaptive Testing for Mental Health (CAT-MH) score between baseline and follow-up among the full sample (n=196) and those in tiers 2 and 3 at baseline (n=108).

CAT-MH scale	Baseline severity score, mean (SD)	Follow-up severity score, mean (SD)	Percent change, %	Effect size, Cohen *d* (95% CI)	*P* value
Overall sample (n=196)					
Depression	49.4 (17.0)	43.7 (20.5)	–11.6	0.30 (0.19-0.41)	<.001
Anxiety	40.7 (19.4)	35.3 (20.5)	–13.4	0.27 (0.15-0.39)	<.001
PTSD[a]	41.7 (17.5)	35.9 (19.9)	–13.9	0.31 (0.19-0.43)	<.001
Baseline tiers 2 and 3 (n=108)					
Depression	60.2 (14.3)	52.4 (20.3)	–13	0.42 (0.26-0.59)	<.001
Anxiety	52.6 (16.1)	48.6 (18.9)	–15.9	0.44 (0.24-0.64)	<.001
PTSD[a]	51.3 (15.3)	44.6 (19.7)	–13	0.37 (0.18-0.56)	<.001

[a]PTSD: posttraumatic stress disorder.

Mental Wellness Check-In Outcomes

Between September 2021 and March 2022, a total of 95 individuals in tier 2 or 3 completed a check-in with a CHAMindWell mental wellness clinician. Of those 95 participants, 37 (38.9%) were referred to mindfulness-based group therapy, 9 (9.5%) were referred to other group therapy services, 17 (17.9%) were referred to individual-level services (ie, psychotherapy or psychopharmacology), and 32 (33.7%) did not receive any additional services besides those offered by CHAMindWell (eg, brief coaching, psychoeducation, and guidance on use of self-help apps).

Discussion

Overview

CHAMindWell used a stratified care approach to population mental wellness that was associated with reduced severity of mental health symptoms and reduced prevalence of moderate to severe symptom levels among program participants. To our knowledge, this was the first program evaluation to examine the effects of a population mental wellness program designed to promote recovery among those with clinically significant mental health symptoms while preventing the onset of clinically significant symptoms among healthy participants. This program integrated regular mental health screenings with mindfulness-informed mental wellness coaching and treatment referrals. Participants were significantly less likely to report clinically relevant symptom severity after 3-6 months of program engagement and reported significant reductions in symptoms of anxiety, depression, and PTSD at follow-up. We also observed greater symptom reductions among those who reported more severe symptoms at baseline. Furthermore, participants were infrequently referred to individual-level mental health treatments, suggesting that CHAMindWell was associated with improved patient outcomes while minimizing the burden placed on mental health service systems.

While the observed findings could reflect natural symptom reductions caused by the easing of pandemic-related stressors and a return to prepandemic symptomatology, we found no significant associations between dates of survey completion and depression, anxiety, or PTSD symptom severity among baseline or follow-up surveys. These null findings suggest that the observed reductions were not primarily driven by the easing of pandemic-related stressors over time. On the other hand, the lack of linear associations could also reflect multiple rolling periods of peak stress intensity during the 15-month data collection period.

With respect to program engagement, the demographic characteristics of individuals who enrolled in CHAMindWell appear similar to those of the patient population Cambridge Health Alliance (CHA) serves [36]. This suggests that CHAMindWell is, at least in theory, appealing to CHA's diverse patient population. However, we found that slightly under one-third of enrolled participants (280/903, 31%) did not complete a baseline survey, and White identity was associated with increased odds of baseline survey completion.

There are a variety of factors that may have contributed to this finding. For example, mistrust of health providers may impact the willingness of some in historically underserved communities to self-report mental health symptoms through web-based surveys [37]. Alternatively, some participants may have found the surveys too burdensome or impersonal, and others may have had mental health symptoms that interfered with their ability to complete surveys [38]. In addition, previous research has

found that participants who are enrolled digitally may be less likely to complete follow-up than those who are enrolled in person [39]. Finally, a majority of the patients in our hospital system are multilingual or do not speak English, so some may have had difficulty interacting with English-based emails and surveys.

Limitations

As a program evaluation, this project was not designed to establish causality, for which a more rigorous experimental design, including a control group, would be needed. In addition, the program experienced substantial attrition in the follow-up sample, and the majority of participants identified as White, which may impact the generalizability of our findings to regions with greater racial diversity.

There are also practical limitations associated with implementing a population mental wellness program like CHAMindWell. For example, this program relied on participants self-reporting their mental health symptoms through web-based surveys. As such, individuals with limited access to the internet or internet-connected devices may not have been able to participate. Furthermore, all aspects of this program, including surveys and communications, were conducted in English, which may have limited some participants' abilities to effectively engage with the program and self-report their symptoms. These factors could impact the generalizability of our findings and the usefulness of this intervention when working with certain populations. To begin addressing these concerns, our team has implemented and is beginning to evaluate a newly developed version of CHAMindWell in Spanish (CHAMindWell en Español).

Finally, this program evaluation used CAT-MH completion as a measure of engagement in CHAMindWell, which does not capture engagement with the recommended resources within each tier. Measuring engagement with and adherence to specific resources and clinician recommendations in the future will help us better understand the applied effects of CHAMindWell.

Conclusions

CHAMindWell is an integrative population mental wellness program that is associated with reductions in the severity of mental health symptoms among participants while limiting referrals to individual-level services. Such programs may help mental health care systems meet increased treatment needs among their patients without overburdening their providers. Future work should aim to better characterize and address the factors that contribute to program nonengagement and attrition and compare the impacts of CHAMindWell to a control condition.

Acknowledgments

Work on this project was funded by the National Center for Complementary and Integrative Health (NCCIH; grants 3R33AT010125-03S1 [ZSO, CF co-PIs] and T32AT000051 [JAR]). The funding sources were not involved in study design, data collection, analysis, interpretation, manuscript preparation, or publication. The authors are solely responsible for all aspects of this manuscript, and its contents do not necessarily represent the official views of the NCCIH or the National Institutes of Health.

The authors would like to acknowledge the important work of Robert Gibbons, Leah Howard, Lexie Comeau, Todd Griswold, Mark Albanese, Christina Luberto, Stephanie Cohen, Gareth Parry, Esteban Da Cruz, Fiona Kate Rice, Rocio Hernandez Chavez, Saul Granados, Clare Bumpus, and Gabriella Conversano in assisting with designing and implementing various aspects of the CHAMindWell program. We also want to acknowledge the important work of the CHAMindWell mental wellness clinicians: Sagal Ahmed, Jessica Dudzinska, David Estey, Esther Estey, Thomas Le, Maria Neizvestnaya, Stephanie Palacios, Erika Pruett, Ilana Sichel, Lia Smith, Quang Tran, Blanche Wright, and Teresa Yeh.

Portions of these data were presented as 1 of 3 selected presentations focused on impactful clinical trials at the NCCIH Advisory Council Meeting on May 13, 2022, and as posters during the 2022 Integrative Medicine Network Forum that took place virtually on November 10 and 11, 2022, and the 2023 International Society of Contemplative Research conference in San Diego, CA, on February 3, 2023.

Conflicts of Interest

None declared.

References

1. Stanley BL, Zanin AC, Avalos BL, Tracy SJ, Town S. Collective emotion during collective trauma: a metaphor analysis of the COVID-19 pandemic. Qual Health Res 2021;31(10):1890-1903 [FREE Full text] [doi: 10.1177/10497323211011589] [Medline: 33980096]
2. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
3. Luszczynska A, Benight CC, Cieslak R. Self-efficacy and health-related outcomes of collective trauma. Eur Psychol 2009;14(1):51-62 [FREE Full text] [doi: 10.1027/1016-9040.14.1.51]

4. Wang PS, Gruber MJ, Powers RE, Schoenbaum M, Speier AH, Wells KB, et al. Mental health service use among hurricane Katrina survivors in the eight months after the disaster. Psychiatr Serv 2007;58(11):1403-1411 [FREE Full text] [doi: 10.1176/ps.2007.58.11.1403] [Medline: 17978249]
5. Stack S, Rockett IRH. Social distancing predicts suicide rates: analysis of the 1918 flu pandemic in 43 large cities, research note. Suicide Life Threat Behav 2021;51(5):833-835 [FREE Full text] [doi: 10.1111/sltb.12729] [Medline: 33565144]
6. TrendWatch: the impacts of the COVID-19 pandemic on behavioral health. American Hospital Association. 2022. URL: https://www.aha.org/guidesreports/2022-05-31-trendwatch-impacts-covid-19-pandemic-behavioral-health [accessed 2023-10-03]
7. Dubé JP, Smith MM, Sherry SB, Hewitt PL, Stewart SH. Suicide behaviors during the COVID-19 pandemic: a meta-analysis of 54 studies. Psychiatry Res 2021;301:113998 [FREE Full text] [doi: 10.1016/j.psychres.2021.113998] [Medline: 34022657]
8. Alkhamees AA, Assiri H, Alharbi HY, Nasser A, Alkhamees MA. Burnout and depression among psychiatry residents during COVID-19 pandemic. Hum Resour Health 2021;19(1):46 [FREE Full text] [doi: 10.1186/s12960-021-00584-1] [Medline: 33823857]
9. Cornwell BL, Szymanski BR, McCarthy JF. Impact of the COVID-19 pandemic on primary care-mental health integration services in the VA health system. Psychiatr Serv 2021;72(8):972-973 [FREE Full text] [doi: 10.1176/appi.ps.202000607] [Medline: 33397142]
10. Delgadillo J, Ali S, Fleck K, Agnew C, Southgate A, Parkhouse L, et al. Stratified care vs stepped care for depression: a cluster randomized clinical trial. JAMA Psychiatry 2022;79(2):101-108 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.3539] [Medline: 34878526]
11. Fletcher S, Spittal MJ, Chondros P, Palmer VJ, Chatterton ML, Densley K, et al. Clinical efficacy of a decision support tool (Link-me) to guide intensity of mental health care in primary practice: a pragmatic stratified randomised controlled trial. Lancet Psychiatry 2021;8(3):202-214 [FREE Full text] [doi: 10.1016/S2215-0366(20)30517-4] [Medline: 33571453]
12. Ho FYY, Yeung WF, Ng THY, Chan CS. The efficacy and cost-effectiveness of stepped care prevention and treatment for depressive and/or anxiety disorders: a systematic review and meta-analysis. Sci Rep 2016;6:29281 [FREE Full text] [doi: 10.1038/srep29281] [Medline: 27377429]
13. Delgadillo J, Moreea O, Lutz W. Different people respond differently to therapy: a demonstration using patient profiling and risk stratification. Behav Res Ther 2016;79:15-22 [FREE Full text] [doi: 10.1016/j.brat.2016.02.003] [Medline: 26937855]
14. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
15. Gál É, tefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021;279:131-142 [FREE Full text] [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
16. Rathbone AL, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. J Med Internet Res 2017;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
17. Goldberg SB, Buck B, Raphaely S, Fortney JC. Measuring psychiatric symptoms remotely: a systematic review of remote measurement-based care. Curr Psychiatry Rep 2018;20(10):81 [FREE Full text] [doi: 10.1007/s11920-018-0958-z] [Medline: 30155749]
18. Berrouiguet S, Perez-Rodriguez MM, Larsen M, Baca-García E, Courtet P, Oquendo M. From eHealth to iHealth: transition to participatory and personalized medicine in mental health. J Med Internet Res 2018;20(1):e2 [FREE Full text] [doi: 10.2196/jmir.7412] [Medline: 29298748]
19. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. Cogn Behav Pract 2015;22(1):49-59 [FREE Full text] [doi: 10.1016/j.cbpra.2014.01.010] [Medline: 27330267]
20. Krusche A, Cyhlarova E, Williams JMG. Mindfulness online: an evaluation of the feasibility of a web-based mindfulness course for stress, anxiety and depression. BMJ Open 2013;3(11):e003498 [FREE Full text] [doi: 10.1136/bmjopen-2013-003498] [Medline: 24293203]
21. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. J Consult Clin Psychol 2010;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
22. Segal ZV, Dimidjian S, Beck A, Boggs JM, Vanderkruik R, Metcalf CA, et al. Outcomes of online mindfulness-based cognitive therapy for patients with residual depressive symptoms: a randomized clinical trial. JAMA Psychiatry 2020;77(6):563-573 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.4693] [Medline: 31995132]
23. CHAMindWell. Center for Mindfulness Compassion. 2023. URL: https://www.chacmc.org/chamindwell [accessed 2023-10-03]
24. Gibbons RD, Alegría M, Cai L, Herrera L, Markle SL, Collazos F, et al. Successful validation of the CAT-MH scales in a sample of Latin American migrants in the United States and Spain. Psychol Assess 2018;30(10):1267-1276 [FREE Full text] [doi: 10.1037/pas0000569] [Medline: 29792502]

25. Gibbons RD, deGruy FV. Without wasting a word: extreme improvements in efficiency and accuracy using Computerized Adaptive Testing for Mental Health Disorders (CAT-MH). Curr Psychiatry Rep 2019;21(8):67 [FREE Full text] [doi: 10.1007/s11920-019-1053-9] [Medline: 31264098]
26. Gibbons RD, Kupfer D, Frank E, Moore T, Beiser DG, Boudreaux ED. Development of a Computerized Adaptive Test Suicide Scale-the CAT-SS. J Clin Psychiatry 2017;78(9):1376-1382 [FREE Full text] [doi: 10.4088/JCP.16m10922] [Medline: 28493655]
27. Brenner LA, Betthauser LM, Penzenik M, Bahraini N, Gibbons RD. Validation of a Computerized Adaptive Test Suicide Scale (CAT-SS) among United States Military Veterans. PLoS One 2022;17(1):e0261920 [FREE Full text] [doi: 10.1371/journal.pone.0261920] [Medline: 35061760]
28. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
29. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
30. Achtyes ED, Halstead S, Smart L, Moore T, Frank E, Kupfer DJ, et al. Validation of computerized adaptive testing in an outpatient nonacademic setting: the VOCATIONS trial. Psychiatr Serv 2015;66(10):1091-1096 [FREE Full text] [doi: 10.1176/appi.ps.201400390] [Medline: 26030317]
31. Gibbons RD, Weiss DJ, Frank E, Kupfer D. Computerized adaptive diagnosis and testing of mental health disorders. Annu Rev Clin Psychol 2016;12:83-104 [FREE Full text] [doi: 10.1146/annurev-clinpsy-021815-093634] [Medline: 26651865]
32. Gibbons RD, Alegria M, Markle S, Fuentes L, Zhang L, Carmona R, et al. Development of a computerized adaptive substance use disorder scale for screening and measurement: the CAT-SUD. Addiction 2020;115(7):1382-1394 [FREE Full text] [doi: 10.1111/add.14938] [Medline: 31943486]
33. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-67 [FREE Full text] [doi: 10.18637/jss.v045.i03]
34. Austin PC, White IR, Lee DS, van Buuren S. Missing data in clinical research: a tutorial on multiple imputation. Can J Cardiol 2021;37(9):1322-1331 [FREE Full text] [doi: 10.1016/j.cjca.2020.11.010] [Medline: 33276049]
35. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
36. Diversity. Cambridge Health Alliance. 2022. URL: https://www.challiance.org/about-cha/diversity-at-cha [accessed 2022-08-11]
37. Benkert R, Cuevas A, Thompson HS, Dove-Meadows E, Knuckles D. Ubiquitous yet unclear: a systematic review of medical mistrust. Behav Med 2019;45(2):86-101 [FREE Full text] [doi: 10.1080/08964289.2019.1588220] [Medline: 31343961]
38. Lewis CC, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing measurement-based care in behavioral health: a review. JAMA Psychiatry 2019;76(3):324-335 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
39. Bajardi P, Paolotti D, Vespignani A, Eames K, Funk S, Edmunds WJ, et al. Association between recruitment methods and attrition in internet-based studies. PLoS One 2014;9(12):e114925 [FREE Full text] [doi: 10.1371/journal.pone.0114925] [Medline: 25490045]

Abbreviations

CAT-MH: Computerized Adaptive Testing for Mental Health
CHA: Cambridge Health Alliance
OR: odds ratio
PTSD: posttraumatic stress disorder
REDCap: Research Electronic Data Capture

Edited by J Torous; submitted 11.04.23; peer-reviewed by J Thornton, M Hamlin; comments to author 22.05.23; revised version received 31.08.23; accepted 03.09.23; published 26.10.23.

Please cite as:
Rosansky JA, Okst K, Tepper MC, Baumgart Schreck A, Fulwiler C, Wang PS, Schuman-Olivier Z
Participants' Engagement With and Results From a Web-Based Integrative Population Mental Wellness Program (CHAMindWell) During the COVID-19 Pandemic: Program Evaluation Study
JMIR Ment Health 2023;10:e48112
URL: https://mental.jmir.org/2023/1/e48112
doi:10.2196/48112
PMID:37883149

Original Paper

Predicting Patients' Satisfaction With Mental Health Drug Treatment Using Their Reviews: Unified Interchangeable Model Fusion Approach

Yi Wang[1], BSc; Yide Yu[1], MSc; Yue Liu[1], PhD; Yan Ma[1,2], PhD; Patrick Cheong-Iao Pang[1], PhD

[1]Faculty of Applied Sciences, Macao Polytechnic University, Macao, Macao

[2]School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China

Corresponding Author:
Yue Liu, PhD
Faculty of Applied Sciences, Macao Polytechnic University
Rua de Luís Gonzaga Gomes
Macao, 999078
Macao
Phone: 853 85996433
Email: yue.liu@mpu.edu.mo

Abstract

Background: After the COVID-19 pandemic, the conflict between limited mental health care resources and the rapidly growing number of patients has become more pronounced. It is necessary for psychologists to borrow artificial intelligence (AI)–based methods to analyze patients' satisfaction with drug treatment for those undergoing mental illness treatment.

Objective: Our goal was to construct highly accurate and transferable models for predicting the satisfaction of patients with mental illness with medication by analyzing their own experiences and comments related to medication intake.

Methods: We extracted 41,851 reviews in 20 categories of disorders related to mental illnesses from a large public data set of 161,297 reviews in 16,950 illness categories. To discover a more optimal structure of the natural language processing models, we proposed the Unified Interchangeable Model Fusion to decompose the state-of-the-art Bidirectional Encoder Representations from Transformers (BERT), support vector machine, and random forest (RF) models into 2 modules, the encoder and the classifier, and then reconstruct fused "encoder+classifer" models to accurately evaluate patients' satisfaction. The fused models were divided into 2 categories in terms of model structures, traditional machine learning–based models and neural network–based models. A new loss function was proposed for those neural network–based models to overcome overfitting and data imbalance. Finally, we fine-tuned the fused models and evaluated their performance comprehensively in terms of F_1-score, accuracy, κ coefficient, and training time using 10-fold cross-validation.

Results: Through extensive experiments, the transformer bidirectional encoder+RF model outperformed the state-of-the-art BERT, MentalBERT, and other fused models. It became the optimal model for predicting the patients' satisfaction with drug treatment. It achieved an average graded F_1-score of 0.872, an accuracy of 0.873, and a κ coefficient of 0.806. This model is suitable for high-standard users with sufficient computing resources. Alternatively, it turned out that the word-embedding encoder+RF model showed relatively good performance with an average graded F_1-score of 0.801, an accuracy of 0.812, and a κ coefficient of 0.695 but with much less training time. It can be deployed in environments with limited computing resources.

Conclusions: We analyzed the performance of support vector machine, RF, BERT, MentalBERT, and all fused models and identified the optimal models for different clinical scenarios. The findings can serve as evidence to support that the natural language processing methods can effectively assist psychologists in evaluating the satisfaction of patients with drug treatment programs and provide precise and standardized solutions. The Unified Interchangeable Model Fusion provides a different perspective on building AI models in mental health and has the potential to fuse the strengths of different components of the models into a single model, which may contribute to the development of AI in mental health.

(JMIR Ment Health 2023;10:e49894) doi:10.2196/49894

KEYWORDS

artificial intelligence; AI; mental disorder; psychotherapy effectiveness; deep learning; machine learning; natural language processing; NLP; data imbalance; model fusion

Introduction

Background

According to the World Health Organization, the number of people who have had anxiety and depressive illnesses has greatly increased since 2020 owing to the COVID-19 pandemic and modern high-paced lifestyles. The early projections indicate a 26% and 28% increase in anxiety and severe depressive disorders, respectively, in 2020 [1]. Moreover, a study has shown that people with mental illnesses may be socially ostracized, stigmatized, or discriminated against [2]. Mental health has become a crucial issue for global development as it affects millions of people worldwide and has a significant social and economic impact. Therefore, there is an urgent need to address the challenges and gaps in mental health care and to promote efficient treatment. The treatment of mental illness implies a long-term process for some patients with mental illnesses. During this process, a psychologist must frequently judge the effectiveness or satisfaction of psychotherapy and drug therapy at a certain stage based on the patient's feedback and their professional skills. The judgment can help psychologists to monitor the changes in the patient's condition for adjusting the treatment plan accordingly. Moreover, it is necessary to identify and repair any ruptures or conflicts that may arise in the therapeutic relationship. Psychologists can then empower the patient to take an active role in their own recovery and enhance their motivation and satisfaction with the therapy. For instance, psychologists often use patient feedback to measure the effectiveness of the treatment approach [3-6]. However, such a manual analysis conducted by psychologists is sometimes inefficient and inconsistent across different psychologists. Moreover, a large amount of jumbled information provided by patients with mental illness may lead to neglected critical information, possibly negatively affecting the reliability of the analysis process. The analysis process requires a large amount of human resources. Coupled with the fact that medical resources are insufficient in many locations, there may be many patients with mental illnesses who do not receive effective treatment and whose feedback cannot be analyzed in a timely manner. Deep learning (DL) and machine learning (ML) in natural language processing (NLP) can handle large amounts of electronic health records to extract information quickly [7]. Hence, they have the potential to improve the efficiency of mental health care decision-making, alleviate insufficient medical resources, and shed light on solutions to address these issues [7].

Related Work

In NLP, there are massive methods for input presentations that generate information containing suitable context, dimensionality, and feature types to ensure the accurate prediction of a classification model [8]. Patient review analysis can seek the opinions or perspectives of patients with mental disorders by extracting specific emotional expressions from their comments [9]. This is regarded as an effective way to evaluate the state of patients. Simple NLP methods such as bag-of-words (BOW) [10] convert text into fixed-length vectors by counting the frequency of each word. Its variant, term frequency–inverse document frequency (TF-IDF) measure [11], is a weighting scheme that assigns a score to each word in a document based on its term frequency and inverse document frequency. TF-IDF enhances the term frequency BOW vectors by assigning more weight to relevant words and less weight to common words [12], and it can be effectively used in different applications. These methods offer significant advantages in terms of speed. However, they disregard the context, and their performance may be unsatisfactory for the tasks in which contextual information is significant for prediction [13,14]. Recently, transformer-based [15] models were introduced, and the state-of-the-art method in this series is Bidirectional Encoder Representations from Transformers (BERT) [16]. They can embed the contextual information of the text and have the potential to evaluate patients' feedback.

In addition, sentiment classification NLP models can be divided into several main categories: symbolic artificial intelligence (AI) and rule-based systems, traditional ML based, and DL based. For traditional ML-based models, support vector machine (SVM) [17-19] and tree models such as random forest (RF) [20] and decision trees are representative. In contrast, popular DL-based models include but are not limited to recurrent neural network (RNN) [21], convolutional neural network (CNN) [22], and attention-based network [15,23]. These methods can address the challenges and gaps in the field of mental health care.

The recent development of thriving NLP technologies has been applied to many medical-related tasks. For instance, a medical opinion lexicon deals with the health care problem of patients [24], a health care analysis–based study on medicines and services [25], an analysis of electronic health records and chatbots for patient communication [26], and NLP methods for extracting information from radiology reports [27]. According to the study by Le Glaz et al [28], traditional ML methods and classical neural networks have been heavily used in mental health domain in previous studies and the informality of text data on mental health was exposed in previous studies. ML-based and DL-based methods have been proven effective and used extensively for mental disorder detection, especially for depression and suicide [29]. A long short-term memory (LSTM)–based RNN [30] and a hybrid CNN and LSTM model [31] were used to detect depression. Besides this, Shah-Mohammadi et al [32] measured treatment effectiveness by discharge records and discharge status. Clinical Language Annotation, Modeling and Processing was used to extract entities from the patients' notes, and RF and logical regression were applied to predict treatment effectiveness. In addition, Zhang et al [33] used 8000 attention-deficit/hyperactivity disorder medication prescriptions to train 2 dense neural network models: one for identifying noninformative prescription texts and the other for predicting prescribed daily dosage and

treatment duration. The metrics used to evaluate the models were accuracy, precision, recall, and F_1-score [29]. However, they do not consider the large variance in model performance for multiclass classification owing to data imbalance. The DL or ML models trained on drug reviews in the health care field mainly deal with text on physiological aspects such as relief of pain and breathing difficulty, rather than focusing on patients' psychological states. Moreover, the way in which patients with mental disorders express themselves might be very different from that of patients with physical illnesses. For example, for depression, the analysis should focus on whether a patient feels cheerful after taking the drug; however, for gastric colic, it is more about whether the pain is relieved. As such, we cannot use the same standard to measure the reviews of physiological treatment and psychological treatment.

Significance of the Study

This paper pilots applying burgeoning AI technologies to analyze patients' feedback to help psychologists to evaluate patients' satisfaction with drug treatment to rapidly adapt the medication regimen. In this paper, we proposed a model fusion method called Unified Interchangeable Model Fusion (UIMF) to decompose SVM, RF, and BERT models and recompose 6 fused models. Then, we trained them on an open data set consisting of the reviews and scores of the drug treatment provided by patients with mental disorders. A novel loss function was proposed for all neural network–based models to alleviate overfitting and data imbalance problems. In addition, we adopted the κ coefficient [34-36], a metric that accounts for chance agreement in imbalanced data, along with other general metrics, to comprehensively evaluate the model performance. The results revealed that the fused transformer bidirectional encoder+RF model performs the best performance while consuming more computing resources and shows a high degree of reliability and indicates that the prediction results of the model are almost in perfect agreement with the random choice of patients' subjective satisfaction. The word-embedding encoder+RF model requires the least computing resources and exhibits relatively good performance.

Methods

Data

Overview

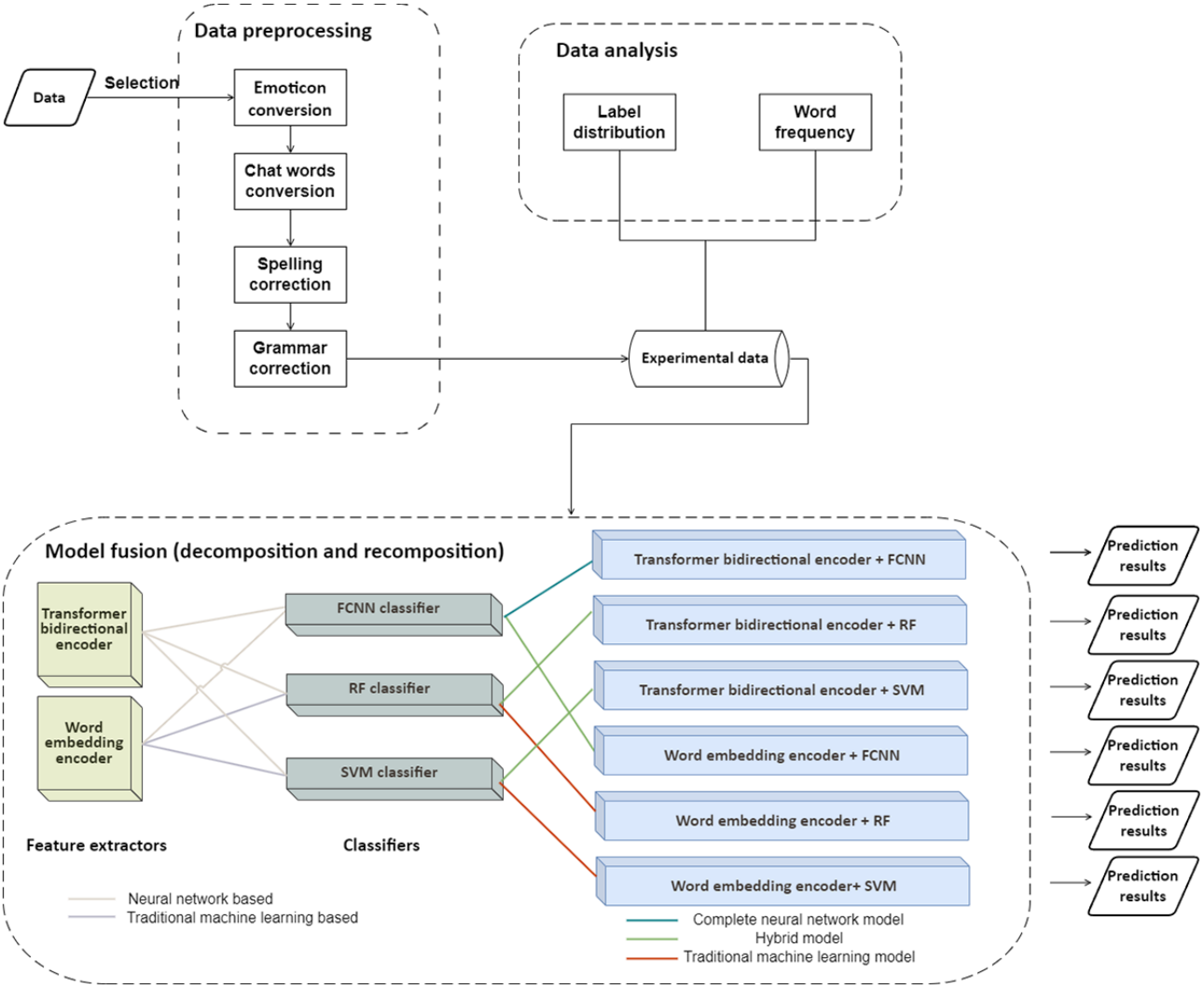
We combined 2 open medical-related text data sets on Kaggle [37,38] to form the raw data set. It contained 16,950 categories of physical and mental illnesses. For each review, 4 attributes, including patient ID, subjective scores (1-10), condition, and drug name are provided. Table 1 provides several sample data.

Figure 1 illustrates the workflow of the operations consisting of data handling and model fusion. Data handling includes 3 main tasks: data selection, data preprocessing, and data analysis.

Table 1. Sample data of the 2 open medical-related text data sets on Kaggle.

Unique ID	Drug name	Condition	Review	Score
206,461	Valsartan	Left ventricular dysfunction	"It has no side effect, I take it in combination of Bystolic 5 Mg and Fish Oil"	9
92,703	Keppera	Epilepsy	"I Ve had nothing but problems with the Keppera: constant shaking in my arms & legs & pins & needles feeling in my arms & legs severe light headedness no appetite & etc."	1
121,333	Venlafaxine	Depression	"My go started me on venlafaxine yesterday to help with depression and the change, a hour after taking them I was feeling very sick couldn't stomach food or fluids, thought keep it up as she told me they did come with some side effects which would get better, took another one last night and was so ill I couldn't stand, being sick sweating shaking thought I was going to pass out. Did get some sleep hoping to feel better this morning, took another one and felt so spaced-out dry mouth shaking, sick, so booked in to see go again to make sure I should be feeling like this, only to find out she had put me on the wrong dose should have been on 37.5mg was put on 150mg, now on right dose hope this will be better"	4

Figure 1. Overview of our workflow: data selection, data preprocessing, data analysis, and Unified Interchangeable Model Fusion. FCNN: fully connected neural network; RF: random forest; SVM: support vector machine.



Data Selection

To create the target data set, we extracted 19 categories of mental disorders defined by the World Health Organization [39] and insomnia which often coincides with the diagnosis of mental disorders [40] from the Kaggle data sets. The target data set included bipolar disorder, depression, panic disorder, psychosis, schizophrenia, anxiety, major depressive disorder, obsessive compulsive disorder, generalized anxiety disorder, autism spectrum disorder, paranoid disorder, performance anxiety, schizoaffective disorder, agitated depression, social anxiety disorder, postpartum depression, dissociative identity disorder, persistent depressive disorder, intermittent explosive disorder, hyperekplexia, and insomnia. The target data set comprised 41,851 reviews.

Data Preprocessing

When writing the review, patients would inevitably use emojis and web-based chat buzzwords and make grammatical and spelling errors, which can result in the introduction of irrelevant symbols and noise that interfere with the important emotional information. In addition, some garbled codes or errors might be introduced owing to human or irresistible factors in data collection. Therefore, we applied a data preprocessing process that sequentially performed emoticon conversion, chat words conversion, spelling correction, and grammar correction. The process aimed to reduce the influence of these factors on the classification performance and to reduce the complexity and dimensionality of the original data. The details of the data preprocessing are listed in subsequent sections.

Emoticons conversion transforms emoticons into their textual meanings. Unlike normal punctuations with weaker emotional polarity, emoticons always represent strong emotions. To preserve the emotional information represented by the emoticons, a Python dictionary of emoticons EMOTICONS_EMO [41] was used to translate emoticons into textual meanings.

Chat words conversion transforms slang, informal, or nonstandard language into standard text. These words can convey various emotions and always have a clear sentiment polarity. However, some NLP corpus or lexicons may not include or only partially include them, leading to the missing emotional state of the patient. In addition, a consistent and clear style of writing is required in the patients' sentiment analysis. We constructed a Python dictionary based on the slang data [42] from GitHub, which consists of popular slang and abbreviations to convert chat words.

The NLP analysis of the health care domain is sensitive to spelling and grammar mistakes, as these errors can significantly affect the quality and credibility of AI-based health care prediction methods [43]. Confusion and misunderstanding may occur between the patient and psychological therapist, which can result in potential errors or risks in diagnosis and treatment. In a worse-case scenario, training a model with data that contain a large number of spelling and grammar errors may introduce bias and noise in the results, which can affect critical health care decision-making. Therefore, we have applied spelling and grammar corrections to this study. For spelling correction, the algorithm by Peter [44] was used to adjust the filtering conditions for candidate words. Two editing distances [45] were used to restrict the difference between the misspelt word and candidate word. The Python API [46] for grammar correction was chosen after conducting a feasibility assessment.

The label distribution analysis in Figure 2 shows the distribution of data across different scores. It can be seen that the data set with 10 scores encounters a more severe data imbalance problem. Data imbalance normally affects the variance in classification performance. It will lead the model to produce higher accuracy for scores with more data and vice versa. To alleviate this problem while considering the actual need for classification, we converted the 10 scores into 3 classes with labels 0, 1, and 2 for poor, fair, and good comments, respectively. As shown in Figure 2, the data distribution of the 3 classes became more balanced after the conversion. Table 2 shows the mapping between the 10 scores and 3 classes.

Figure 2. Data amount for 10 scores and for the 3 classes.

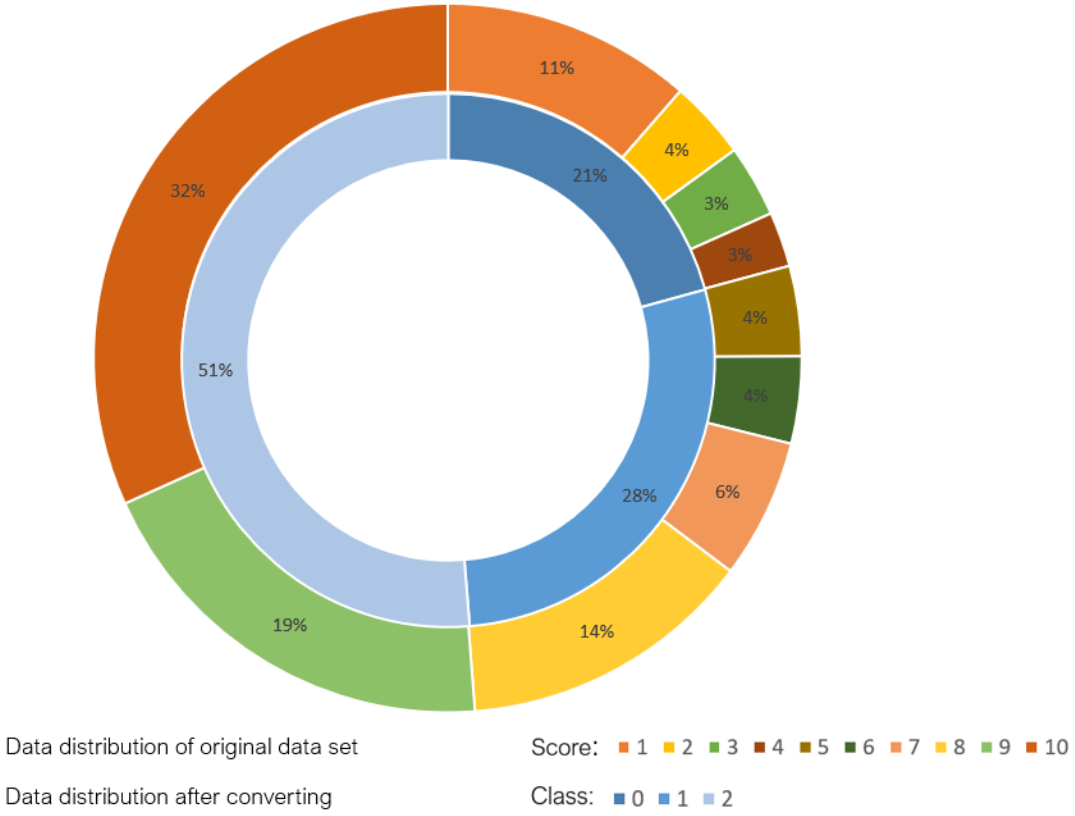


Table 2. The rules of label mapping and the meaning of each class label.

Class label	Score range	Proportion of data (%)	Meaning
0	1-4	20.82	A poor satisfaction of drug treatment and case needs urgent rediagnosis.
1	5-8	27.96	A fair satisfaction of drug treatment and cases needs to be monitored.
2	9-10	51.22	A good satisfaction of drug treatment and no further action needed.

For the target data set, we extracted 30% of the data for testing, 63% for training, and 7% for validation by stratified sampling and used a 10-fold validation strategy for training.

Problem Formulation

Overview

For any patient, the review data were , the score given by the patient for the treatment was , where , , and were the review record and the corresponding score provided by the patient during the j^{th} treatment, respectively. Here we regarded

each review and score pair as independent. The score was only affected by the corresponding review.

For each review the loss function was defined as , was the prediction result of a prediction function *f*, and was the ground truth of . As the prediction function is decomposed into encoder function and classification function, it is denoted as follows:

where , and *F* are the set of fused prediction functions of *G*; *H*, *G*={RF, SVM, fully CNN (FCNN)} is the set of classification functions; and *H*={transformer bidirectional encoder, word-embedding encoder} is the set of encoder functions.

The objective of this study was to find the optimal recomposed prediction function to minimize the empirical risk.

where $R_{emp}(f)$ is the empirical risk, which is a concept in statistical learning theory that measures how well a learning algorithm performs on a given data set [47]:

Preliminary Data Analysis

In Textbox 1, we show the top 30 most frequent words in class 2 (a good satisfaction), class 1 (a fair satisfaction), and class 0 (a poor satisfaction). It shows that the high-frequency words in the 3 classes are extremely similar. If using the BOW method that transforms a given text into a vector based on the word frequency, the vectors of those similar high-frequency words will have higher values in more dimensions and vice versa. In this situation, the vectors of high-frequency words become more critical influencing the classification than the low-frequency words. However, low-frequency words do not mean that they are always irrelevant or useless to the text; they may capture some specific or rare information. Hence, the term frequency BOW method could result in text vectors from every class being similar to each other for classifiers, and the distinctive information of text vectors from different classes is buried. Using TF-IDF helps distinguish the text from each class by reducing the vector similarity and highlighting the unique features. Therefore, we applied TF-IDF to the vectorizer to measure the importance of words in a text review and weighted the embedding of text based on word importance. The details of this are discussed in the following section.

Textbox 1. Top 30 most frequent words in 3 classes (the high-frequency words shared by all 3 classes are in italics).

Top 30 most frequent words in class 0

- also
- anxiety
- day
- depression
- doctor
- drug
- effect
- feel
- felt
- first
- get
- help
- hour
- like
- made
- medication
- month
- night
- one
- side
- sleep
- started
- take
- taking
- time
- took
- week
- work
- would
- year

Top 30 most frequent words in class 1

- also
- anxiety
- better
- day
- depression
- doctor
- effect
- feel
- first
- get

- good
- help
- like
- medication
- medicine
- month
- much
- night
- side
- sleep
- started
- still
- take
- taking
- time
- week
- weight
- work
- would
- year

Top 30 most frequent words in class 2

- anxiety
- attack
- back
- better
- day
- depression
- doctor
- effect
- feel
- first
- get
- help
- life
- like
- medication
- medicine
- month
- much
- night
- panic
- side
- sleep

- started
- take
- taking
- time
- week
- work
- would
- year

Uniform Interchangeable Model Fusion

Overview

In this study, we not only applied the end-to-end RF [48], SVM [18], and BERT [16] to classify the reviews but also we attempted to optimize the structure of these end-to-end models. We were motivated by the low interpretability of most ML models, which makes it impossible to determine which part of the model plays the most critical role in the model's performance and which part hinders it. By decomposing these models into standard functional modules and analyzing the performance of each module of those models, we can shed light on the reasons behind their impressive performance for specific NLP tasks. It is then possible to combine the best modules to construct the optimal models. Therefore, we proposed UIMF to fuse different models and displace the structure of the models. It addresses the difference of the underlying algorithmic logic of the neural network and traditional ML end-to-end models and the compatibility between them to fuse functional modules from different types of models. UIMF includes 2 phases: model decomposition inspired by the modularization technique proposed by Kingetsu et al [49], and model recomposition. It decomposes a model into manageable modules based on their functionality. Model recomposition is designed with a compatible model structure that chains the encoder and the classifier sequentially, and the modules are recomposed into fused models.

Model Decomposition

In NLP, both traditional ML and neural network models use the encoders to extract and vectorize text features and use the classifiers to perform classification. However, traditional ML models have more modular and independent components than neural network models. Traditional ML models and neural network models have their own strengths: traditional ML models are more robust, interpretable, and computationally efficient, whereas neural network models are better at fitting continuous functions, extracting features, and adapting to different tasks. To the best of our knowledge, no research has proven that the model structure of SVM, RF, or BERT is optimal. Therefore, we attempted to fuse them to combine their advantages from a structural perspective. However, because the structure of the model is contextually linked, randomly splicing and overlaying different model structures may cause internal inconsistencies. Therefore, we applied UIMF to decompose a model into 2 functional modules: the encoder and the classifier. For RF and SVM models, the encoder was a word-embedding algorithm, and the classifiers were the trees and support vectors, respectively. For BERT, the encoder was the multihead attention layers and the feed-forward network layers, and the classifier was the subsequent neural network classification layer. After model decomposition, we obtained 2 encoders—the word-embedding encoder and the transformer bidirectional encoder—and 3 classifiers—RF classifier, SVM classifier, and FCNN classifier—as shown in Figure 1.

For the word-embedding encoder, all the nonrepeating words in the data set were extracted and mapped to a vector space, and the words in the paragraph were replaced with the corresponding vectors. As mentioned earlier, we found that the most frequent words in class 2 were similar to those in class 0 and class 1. However, they lack words with emotional polarity, which may indicate the weak importance of context. To select and weigh the significant words rather than just high-frequency words, we used TF-IDF [11], a statistical measure of word importance in documents. The TF-IDF is combined with word embeddings to create document embeddings, which are vector representations of sentences or documents that preserve the semantic and syntactic properties of words. We multiplied the word's TF-IDF score by its embedding vector and averaged the result over all the words in the sentence. Thus, we avoided ignoring words with high emotional polarity and weakening their influence on the classification results.

Transformer bidirectional encoders are multihead attention layers and feed-forward network layers extracted from lightBERT. Compared with BERT, lightBERT can meet the requirements of a realistic production environment with edge devices. LightBERT was pretrained on Wikipedia using TensorFlow [50], and we fine-tuned it using the target data set.

As for classifiers, the FCNN layer, RF, and SVM are decomposed as the classifiers. The RF classifier can handle high-dimensional data and nonlinear relationships, perform feature selection and importance ranking, and be robust to noise from few grammar and spelling errors. The SVM classifier can work well with small data sets and be robust to outliers. While the FCNN can learn complex patterns and nonlinear relationships from data and be flexible to different architectures [51]. These classifiers use the decision tree ensemble method, kernel method, and approximation theorem, respectively, for classification, and they are well contrasted with each other.

These encoder and classifier modules are in preparation for the model structure reorganization fusion in the next section to

explore whether substituting and recomposing the internal structure of the end-to-end traditional ML and neural network models results in better models.

Model Recomposition

In model decomposition, the 2 encoders and 3 classifiers are cross-combined separately to be fused into 6 models. Three of them are existing models from the perspective of model structure: BERT (transformer bidirectional encoder+FCNN), RF (word-embedding encoder+RF), and SVM (word-embedding encoder+SVM). The other 3 models are real fused models including the transformer bidirectional encoder+RF classifier, transformer bidirectional encoder+SVM, and word-embedding encoder+FCNN; and their model structure are shown in Figures 3-5.

Figure 3. Architecture of transformer bidirectional encoder+random forest (RF) model.

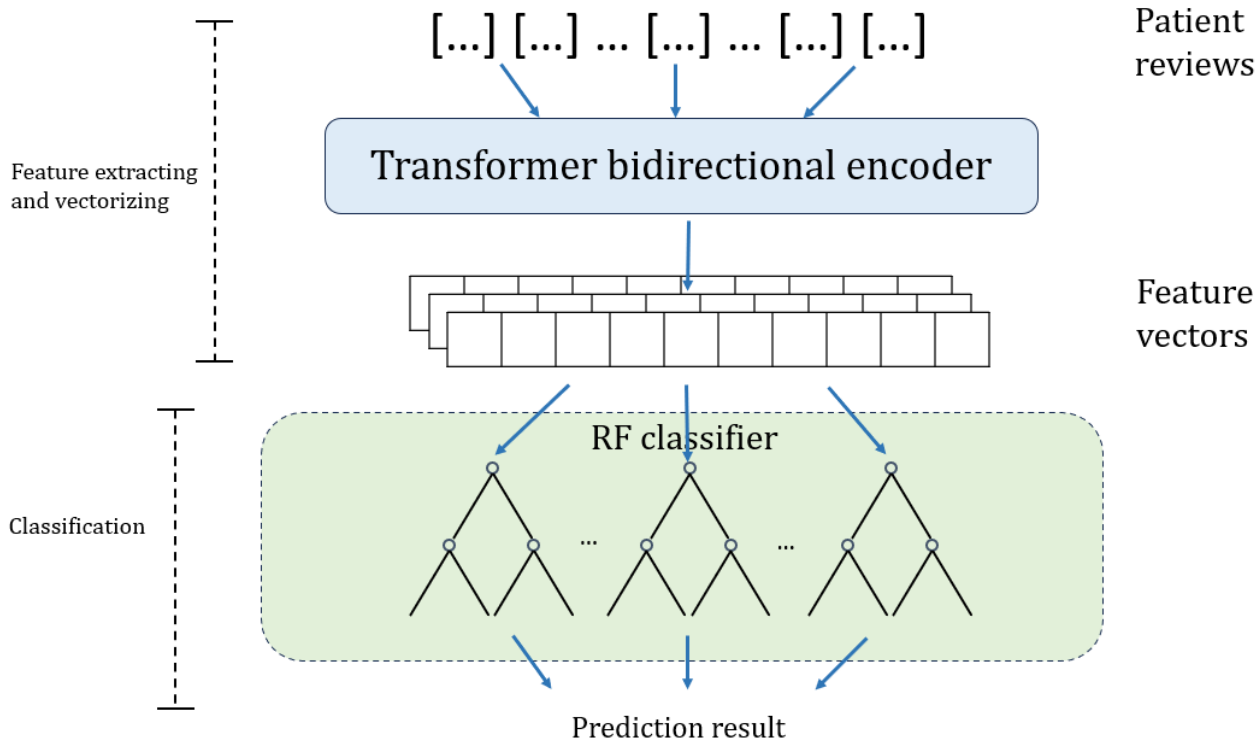


Figure 4. Architecture of transformer bidirectional encoder+support vector machine (SVM) model.

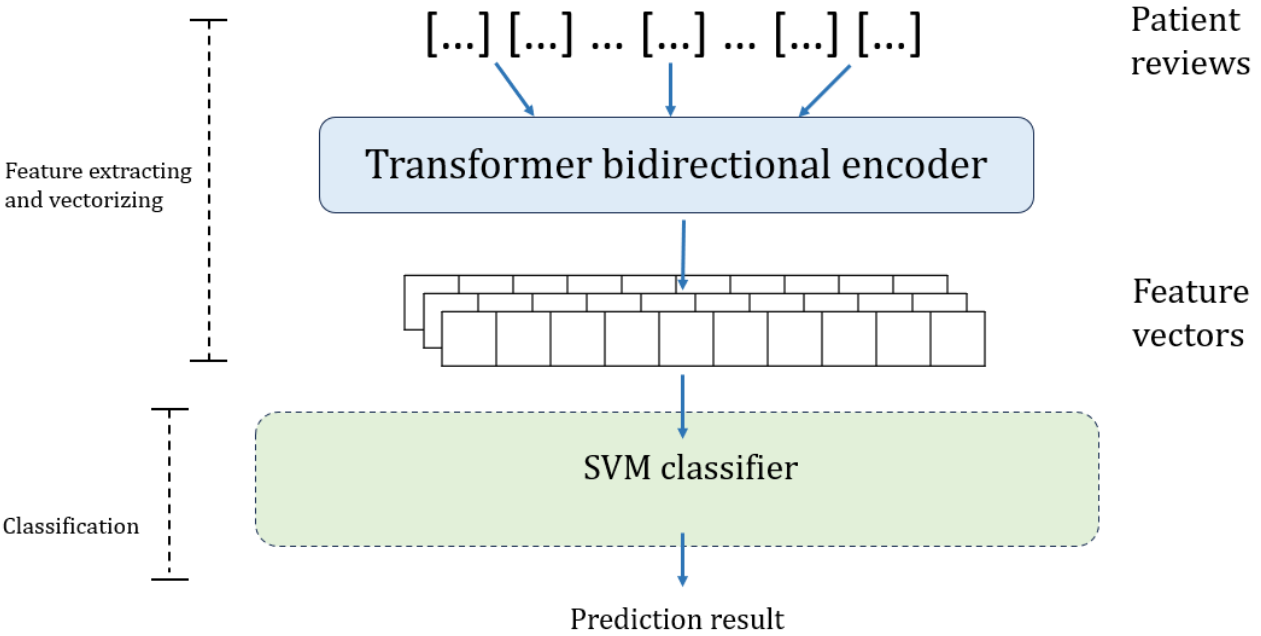
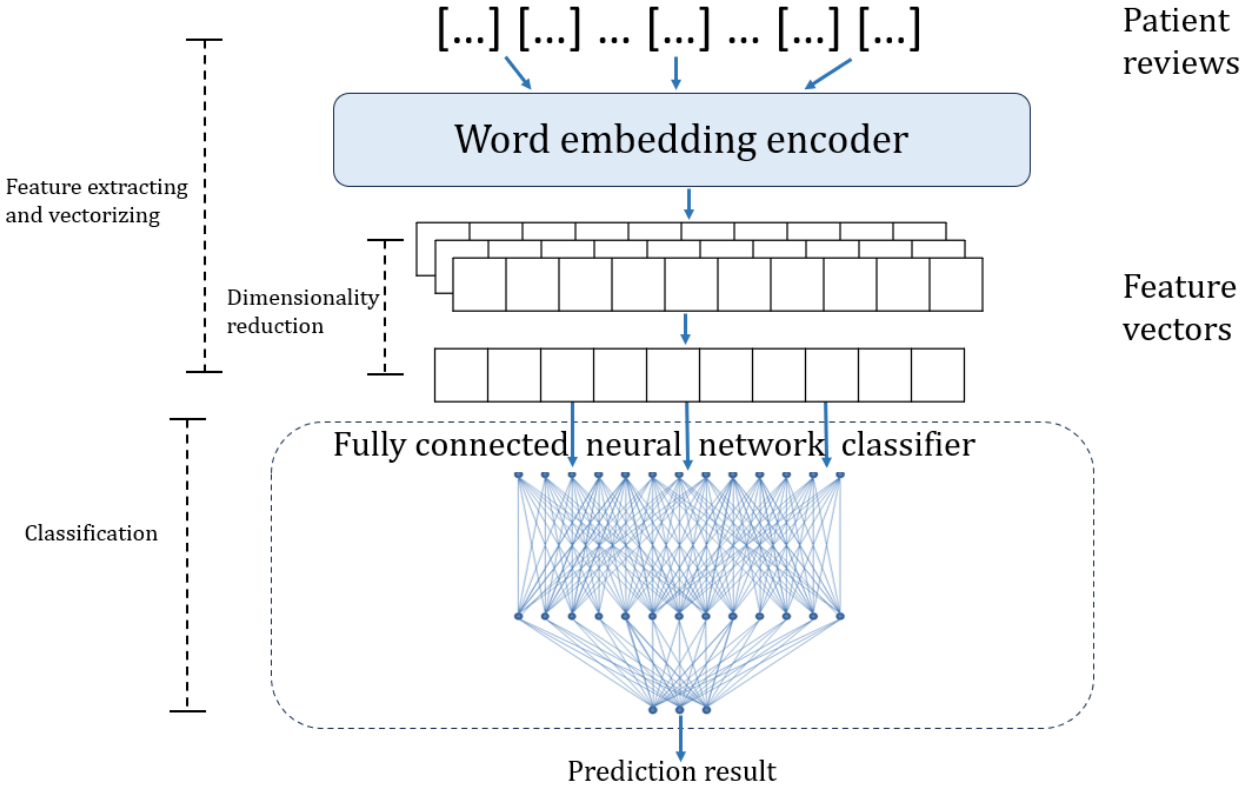


Figure 5. Architecture of the word-embedding encoder+fully connected neural network model. RF: random forest.



As less significant words were not filtered in the word-embedding encoder, which leads to a large number of features that have little impact on the classification result being embedded. Dealing with those high-dimensional feature vectors consumes a large amount of computational resources. To reduce dimensionality, we conducted feature engineering, which scored the importance of feature dimensions based on the TF-IDF and filtered the 6000 most significant dimensions for classification. On the basis of the model structure, the fused models are classified into 2 categories: traditional ML models and neural network models, which consist of hybrid models and pure neural network models.

The traditional ML models include word-embedding encoder+SVM and word-embedding encoder+RF classifier. Imbalanced data will have little effect on the performance of the word-embedding encoder. The SVM classifier attempts to find the optimal hyperplane that maximizes the margins between different classes of feedback from patients. This means that it focuses on patient instances close to the decision boundary (support vector) and ignores those far from the boundary. Consequently, it is less affected by the majority class. The RF classifier is also less affected by noise and outliers, which may be present in the majority classes. It builds multiple decision trees based on different classes of feedback from patients and their features to reduce the correlation and variability between trees and to increase the diversity and robustness of the set of trees. Overall, these 2 models are robust to imbalanced data sets.

The neural network models include the word-embedding encoder+FCNN classifier, transformer bidirectional encoder+SVM classifier, transformer bidirectional encoder+RF classifier, and transformer bidirectional encoder+FCNN classifier. The difference between BERT and our transformer bidirectional encoder+FCNN model is that the loss function of BERT model is cross-entropy, whereas the loss function of our transformer bidirectional encoder+FCNN is the focal flooding (FF) loss. In general, the loss function that the neural network models share is cross-entropy loss. Thus, for gradient-based optimization methods, the errors from the majority class (class 2) will predominate over the errors from the minority classes (class 0 and 1) and have a more significant impact on the parameter adjustment. To add insult to injury, because too many parameters and layers are used to boost the complexity and adaptability of these models, they can easily overfit to the majority class and underfit to the minority class. This situation can lead to poor generalization and low recall for the minority classes, which is more important in the target data set. We address this problem by modifying the loss function. The flooding loss [52] and focal loss [53] that are with high profile from computer vision are borrowed for this task. A new loss function, called FF loss, is defined as follows:

where b is a hyperparameter to control the minimum of the loss function against overfitting, is used to change the weight for negative examples (majority samples) to release the data imbalance problem, and γ is used to reduce the weight of easily classified samples to improve the performance of the model on data that are difficult to classify.

Hence, the neural network models with FF loss adjust the weights to focus more on the particular class of patients' reviews and reviews that are difficult to classify to prevent overfitting. We applied FF loss to all our fused neural network models. Finally, the Adamw [54] optimizer was used to optimize the

parameters of the models. In addition, all these models were trained using an 11th Gen Intel Core i9-11900KF@3.50GHz CPU and an NVIDIA GeForce RTX 3080 GPU.

Results

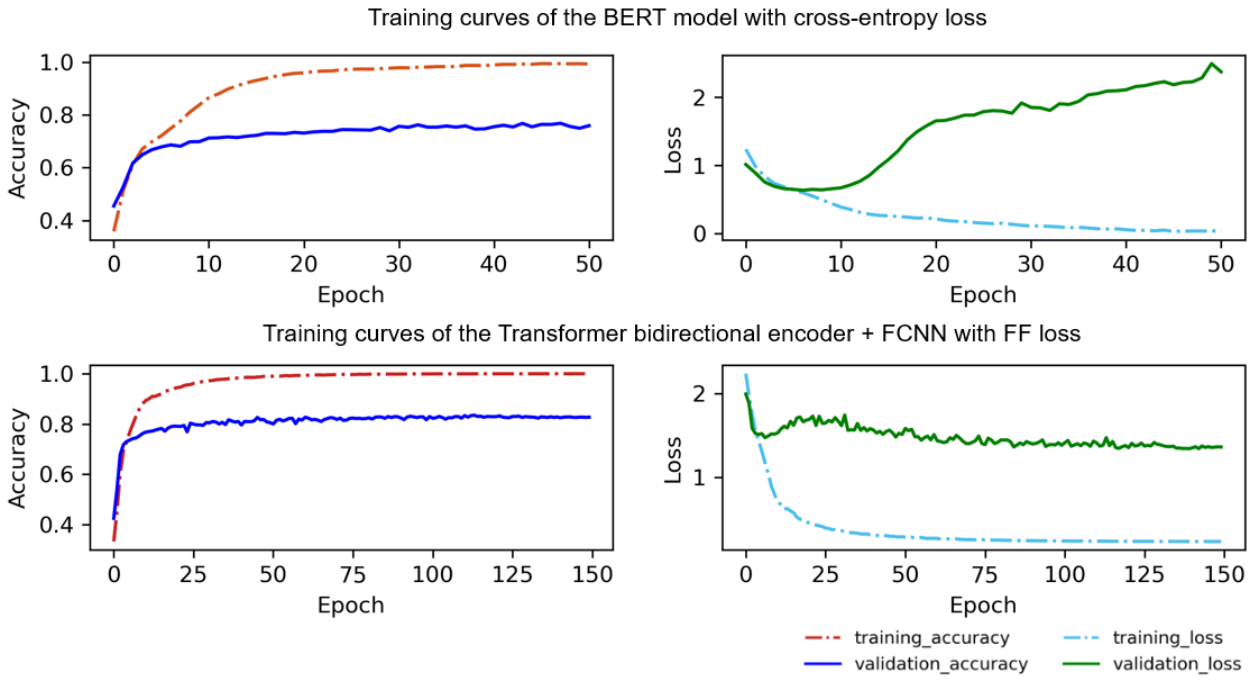
Extensive experiments were conducted for two main purposes: (1) fine-tuning all the fused models and (2) performance comparison between the fused models.

Optimization of the Transformer Bidirectional Encoder+FCNN Model

Loss Function Adjustment

As a benchmark, BERT was fine-tuned. The first 2 subfigures in Figure 6 show the training curve of the model with the cross-entropy. An overfitting problem appears even with the early stopping [55] regularization method. It is obvious that after decreasing at the start of the training, the validation loss continues to increase at the later phase of the training, although the other 3 lines continue to behave normally. This could be because the model training is in the early stage of overfitting, and it learns the specific patterns and noise in the training data too well but fails to generalize to new and unseen data. Moreover, the bias is temporarily insufficient to have a significant effect on the other curves for the time being. Hence, the remainder of the curve remains stable. However, as the number of training epoch increases, the model will inevitably learn a large amount of noise, which leads to significant increase in the validation loss and significant decrease in the validation accuracy.

Figure 6. Training curves of the Bidirectional Encoder Representations From Transformers (BERT) model with cross-entropy loss and the transformer bidirectional encoder+fully connected neural network (FCNN) with focal flooding (FF) loss.



Therefore, we trained the transformer bidirectional encoder+FCNN model (same structure as a BERT model [16]) but with FF loss function. The hyperparameters are set as follows: α=1.0, γ=1.0, b=0.125, and 2 hidden layers of the FCNN are used. The third and fourth subfigures of Figure 6 show that the validation loss rises briefly near the 25th epoch and then declines steadily until it stabilizes, whereas the validation accuracy increases steadily. Moreover, the accuracy of the transformer bidirectional encoder+FCNN model improved by approximately 3% to 6% compared with the BERT model. This was because limits the minimum value of the training loss, which can prevent the model from overlearning to the noise in the data set. It is significant proof that the FF loss function can effectively stop overfitting from occurring.

Hyperparameter Adjustment in Loss Function

To optimize the performance, hyperparameter adjustment experiments were conducted on the model with FF loss. The optimization focusing on adjusting *a*, γ, *b* is just needed to be given a suitable value because is essentially a regularization method, which is mainly for against overfitting. If the overfitting does not exist, regularization may not significantly improve model performance because it indicates that the model already has a good balance between bias and variance.

In these experiments, α of class 1 was adjusted while keeping α of class 0 and class 2 unchanged because the prediction accuracy for this class was much lower than in previous experiments. α of class 1 was set to be 0.5, 1.0, 2.0, and 4.0 and α of class 0 and class 2 were 1.0. In addition, γ was set to 0.25, 0.50, 0.75, 1.00, 1.25, and 1.50, which controls the weight of the data that are easy to classify. We formed 24 sets of α and γ pairs by permutation and experimented with each pair to generate Figure 7.

Figure 7. The κ coefficient of the transformer bidirectional encoder+fully connected neural network model with 2 hidden layers.

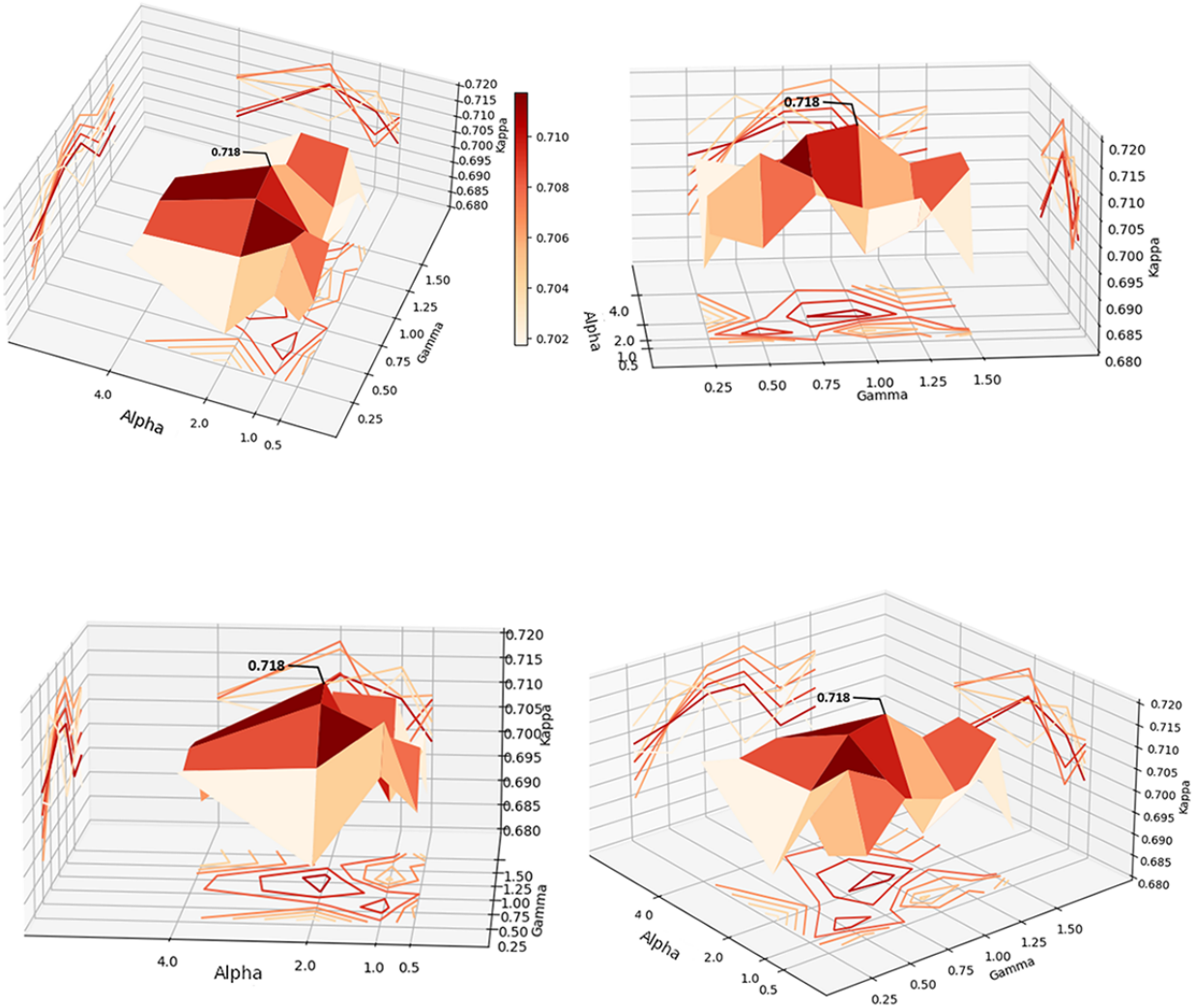


Figure 7 shows the trend in the κ coefficients of the model with respect to α and γ. The surface is flatter when α is larger than 2.0. The chart has a valley when α=1. As γ increased to 1.0, the surface rises. There is a ridge at γ=1. The peak of the κ coefficient appears when α=2 and γ=1.00−1.25. As we mentioned in FF loss function of model recomposition, is used to change the weight for negative examples (majority samples) to release the data imbalance problem, and is used to reduce the weight of easily classified samples. Hence, optimal model performance can be achieved by setting the weight of class 1 samples to 2 and the weight of difficultly classified samples to a range of 1.0 to 1.25. The number of hidden layers is also tuned; however, it does not significantly affect the κ coefficient.

Optimization of the Transformer Bidirectional Encoder+RF Model

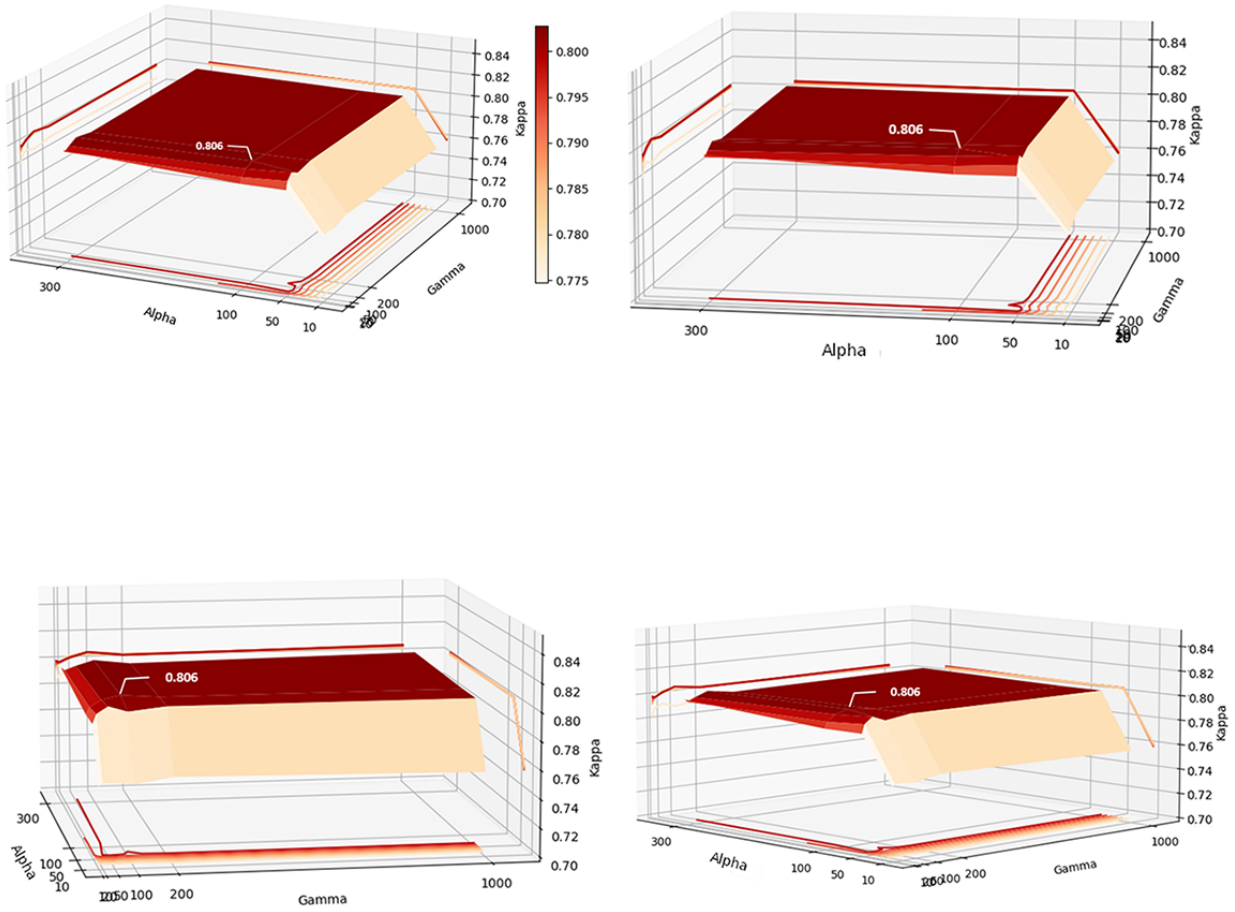
Table 3 shows the average κ coefficient of the models with different maximum depths and different numbers of subtrees. The κ coefficient varied in the range from 0.754 to 0.809. The trend of the κ coefficient with respect to the number of subtrees (γ) and maximum depth (α) is shown in Figure 8. It can be observed that the performance increases when increasing both gamma and alpha before they reach 100. After 100, increasing both alpha and gamma significantly increased the computational cost and reduced the generalizability of the model. The optimal model is with the maximum depth of 100 and 100 subtrees.

Table 3. The κ coefficient of the transformer bidirectional encoder+random forest with different number of subtrees and maximum depth.

Maximum depth	Number of trees					
	10	20	50	100	200	1000
10	0.754	0.756	0.756	0.755	0.759	0.756
50	0.791	0.798	0.803	0.799	0.802	0.802
100	0.791	0.796	0.801	*0.809* [a]	0.802	0.802
300	0.799	0.797	0.800	0.804	0.801	0.804

[a]The highest κ coefficient is italicized.

Figure 8. The trend of the κ coefficient of the transformer bidirectional encoder+random forest.



Optimization of the Transformer Bidirectional Encoder+SVM Model

We compared the performance of the models with linear and Gaussian kernels on the validation set. The linear kernel showed better results than the Gaussian kernel. The linear kernel showed better results than the Gaussian kernel in terms of F_1-score, accuracy, precision, recall, and κ coefficient. The linear kernel achieved an F_1-score of 0.833, an accuracy of 0.837, a precision of 0.833, a recall of 0.837, and a κ coefficient of 0.731. The Gaussian kernel, on the other hand, achieved an F_1-score of 0.826, an accuracy of 0.829, a precision of 0.825, a recall of 0.829, and a κ coefficient of 0.720. Besides, the model with a linear kernel achieves faster training, faster prediction, and lower cost, but its prediction accuracy for linearly indistinguishable data is much lower than that with a Gaussian kernel. The results indicate that the patients' reviews processed by the transformer bidirectional encoder are highly linearly separable.

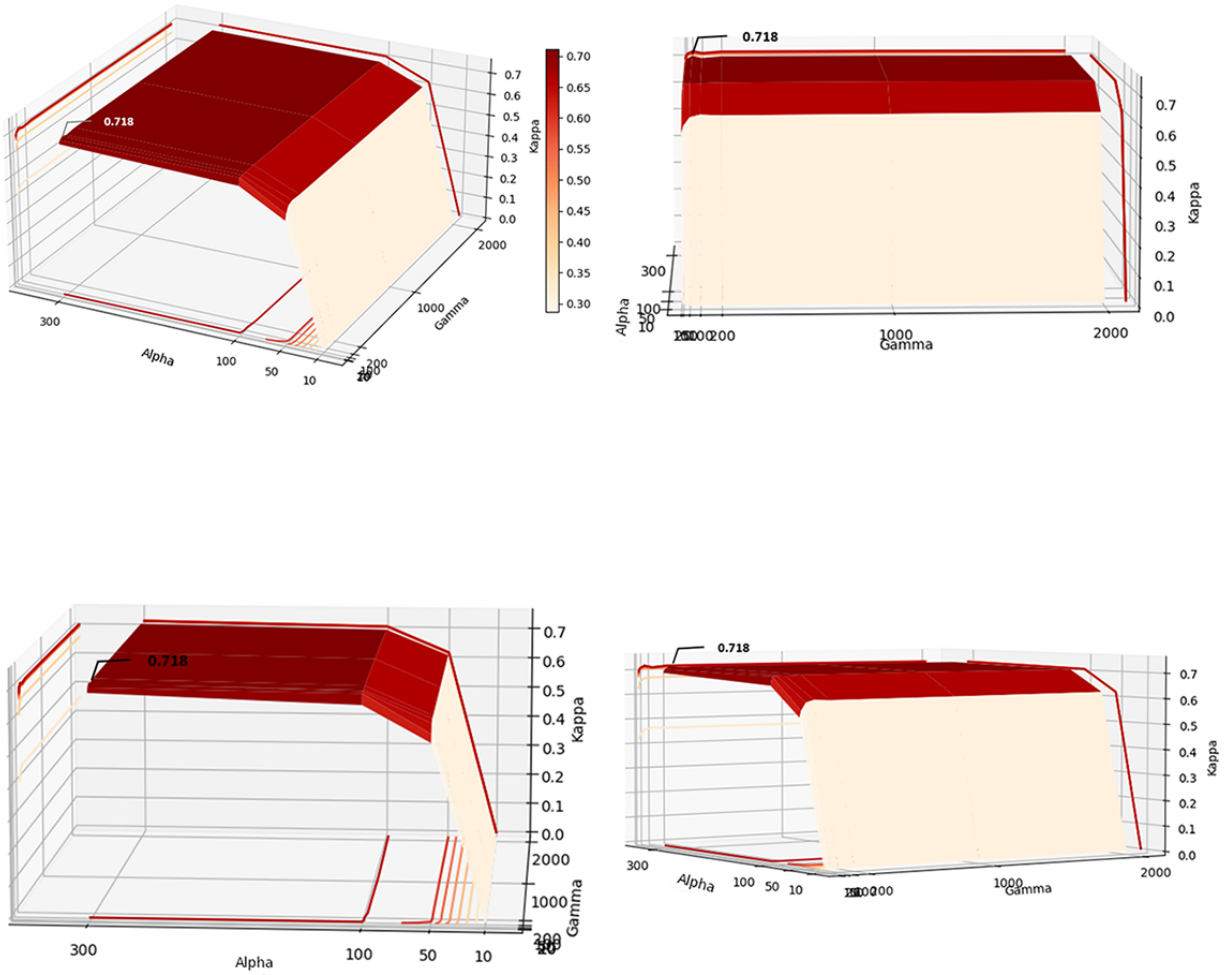
Optimization of the Word-Embedding Encoder+FCNN Model

The FCNN classifier was extracted from the fine-tuned transformer bidirectional encoder+FCNN model and then recomposed with the word-embedding encoder. Because the output of patients' reviews processed by the word-embedding encoder is up to 6000 dimensions, the number of input features of FCNN classifier increases significantly. Therefore, we increased the number of neurons in each hidden layer of the FCNN classifier. The best model achieved a κ coefficient of 0.573 when α=2 and γ=1.5. This model performs much worse than the other fused models.

Optimization of the Word-Embedding Encoder+RF Model

Because the input of this model was up to 6000 dimensions, the maximum number of subtrees in the experiments was increased to 2000. The trend of the κ coefficient with respect to number of subtrees (gamma) and maximum depth (alpha) is shown in Figure 9, as increasing the number of subtrees after reaching 50 and the maximum depth after reaching 300 does not significantly improve κ coefficient, but significantly increases the computational cost and reduces the generalizability of the model, the optimal word-embedding encoder+RF model is the model with a maximum depth of 300 and 50 subtrees, reaching a κ coefficient of 0.718.

Figure 9. The κ coefficient of the word-embedding encoder+random forest model.



Optimization of the Word-Embedding Encoder+SVM Model

We compared the performance of the model using the linear and Gaussian kernels on the validation set. The linear kernel achieved an F_1-score of 0.701, an accuracy of 0.708, a precision of 0.700, a recall of 0.708, and a κ coefficient of 0.514. The Gaussian kernel, on the other hand, achieved an F_1-score of 0.824, an accuracy of 0.824, a precision of 0.825, a recall of 0.824, and a κ coefficient of 0.715. It can be observed that all metrics of the Gaussian kernel are higher than those of the linear kernel. The results may also indicate that the patients' reviews processed by the word-embedding encoder are highly linearly indivisible.

Model Comparisons

A comprehensive comparison is made among all the optimized and fused models in terms of the average graded F_1-scores, accuracy, κ coefficient, and training time in Table 4. The models were evaluated using 10-fold cross-validation. In addition, results were obtained based on the average value of 10 times testing. The graded F_1-score is a metric derived from a knowledge-aware assessment of the severity of suicide risk for early intervention. The weights of the 3 classes are 2, 1 and 1, because class 0 is more important than class 2. Figure 10 shows the confusion matrices of all the optimized fused models. The accuracy for predicting each class varies in different models. In general, the transformer bidirectional encoder+RF model achieved the best overall performance, reaching a graded F_1-score of 0.872 and a κ coefficient of 0.806 at the expense of more training time. According to Figure 10, its accuracy for the 3 classes is balanced where class 1 is with the lowest accuracy of 0.77 and class 0 and class 2 are with high accuracy of 0.89 and 0.94, respectively. The word-embedding encoder+RF is the fastest model, which only required 97.635 seconds for training, and its performance is reasonably acceptable. It achieves a graded F_1-score of 0.801 and the κ coefficient of 0.695. The corresponding confusion matrix in Figure 10 shows that the model has a particularly good accuracy for class 2 but poor accuracy for class 1.

Table 4. Comparison of the optimized models.

Fused models	Graded F_1-score	Accuracy	κ	Train times
Transformer bidirectional encoder+RF[a]	*0.872* [b]	*0.873*	*0.806*	45,357.947
Transformer bidirectional encoder+SVM[c]	0.861	0.863	0.775	45,311.353
Transformer bidirectional encoder+FCNN[d]	0.857	0.858	0.769	45,297.559
Word-embedding encoder+RF	0.801	0.812	0.695	*97.635*
Word-embedding encoder+SVM	0.797	0.806	0.687	5233.945
Word-embedding encoder+FCNN	0.734	0.735	0.577	24,196.99
BERT[e] (benchmark) [16]	0.846	0.846	0.746	45,382.829
MentalBERT	0.867	0.867	0.785	110,669.089

[a]RF: random forest.

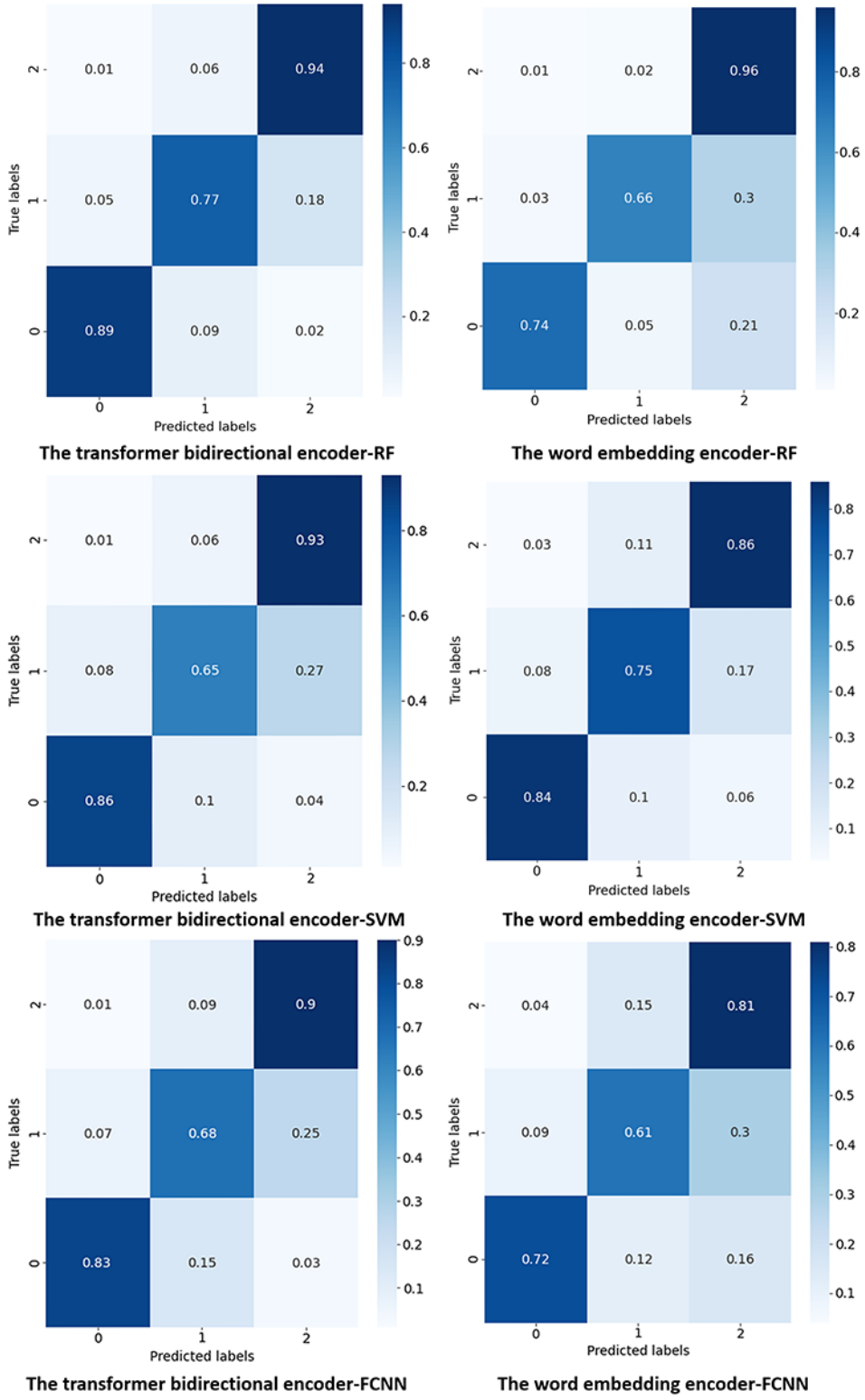
[b]Optimum value.

[c]SVM: support vector machine.

[d]FCNN: fully connected neural network.

[e]BERT: Bidirectional Encoder Representations From Transformers.

Figure 10. The confusion matrix of recomposed models. FCNN: fully connected neural network; RF: random forest; SVM: support vector machine.



Discussion

Principal Findings

In Table 4, regarding graded F_1-scores, accuracy, and κ coefficient, the 3 fused models using the transformer bidirectional encoder have better overall performance than those using the word-embedding encoder. The models with the transformer bidirectional encoder have more stable and balanced performance. This is because the transformer bidirectional encoder can map patients' reviews in a low-dimensional but more linearly separable vector space. Among the classifiers, the RF classifier is the best classifier. The reason behind this is it averages the results of multiple decision trees, each trained on a different subset of the data, to produce a more robust and accurate prediction. RF exhibits better performance, especially with the transformer bidirectional encoder.

The transformer bidirectional encoder+RF model achieves the best average performance, which far exceeds the state-of-the-art BERT model. The corresponding confusion matrix shows that the model performed the best accuracy in class 2, reaching 0.94. The lowest accuracy of 0.77 is for class 1. The reason is that the amount of data in class 2 is much larger than that in class 1, and the language style of the reviews in class 1 is more ambiguous without distinctive features. In addition, the transformer bidirectional encoder+RF model showed better performance on every metric and only 2 out of 5 of the training time compared with the current state-of-the-art MentalBERT model in the mental health domain. The variation in accuracy for all classes was relatively small compared with the other models. It is worth mentioning that the κ coefficient reaching 0.81 indicates the prediction results of the model are almost in perfect agreement with the random choice of patients' subjective satisfaction [34]. It means our model can precisely evaluate the satisfaction of the treatment. However, this method especially consumes graphics processing units, storage, and time resources. It may require the psychotherapy and mental drug treatment providers to have sufficient computing resources to train their patient reviews.

However, for the word-embedding+RF model, the training time decreases significantly at the expense of decreasing the graded F_1-score by 0.071 and accuracy by 0.061. If considering a practical scenario with limited resources available for training, such as small clinics or small hospitals, the word-embedding encoder+RF is also a good choice for a fast implementation.

Implications for Clinical Use

This pilot study is an in-depth exploration of AI methods for assisting psychologists in developing or optimizing mental medication regimens by analyzing the patients' satisfaction with mental drug treatment and reducing the impact of mental disorders on global development and alleviating the problem of strained mental health care resources in the post–COVID-19 pandemic era. The transformer bidirectional encoder+RF model outperformed the state-of-the-art BERT model and MentalBERT model in most aspects, and in terms of training time, it is comparable to BERT and significantly better than MentalBERT. Hence, we showed that the transformer bidirectional encoder+RF model fused via UIMF is one of the best models for predicting patients' satisfaction. In addition, for environments with constrained computational resources, the word-embedding encoder+RF could be used for a much faster training time. Both have the potential to assist psychologists in analyzing patients' satisfaction with drug treatment, while increasing their efficiency in the clinical setting. Moreover, it also has the potential to serve as a significant tool for the training and advancement of trainee psychologists.

Limitations

There are still some limitations to be addressed in future studies. For instance, the spelling and grammar corrections in this study were slightly deficient in terms of performance, which might have inevitably introduced noises. Moreover, the best model relied on a large pretrained language model that might contain biases or errors, and it may negatively affect the quality of the prediction. Hence, we suggest that future work can focus on exploring the potential of the GPT-4 model for spelling and grammar correction during data preprocessing. In addition, future attempts could be made to simulate patients using the GPT-4 model to generate more comprehensive textual data of the mental disorders health care domain based on existing limited data to train language models and improve the robustness of pretrained language models. Moreover, future work can attempt parallel fusion of multiple feature extractors, including BERT, BioMegatron, and GPT models, whose outputs are separate modalities and allow all modalities to be fused at the input level of the predictor, which is then learned by the predictive model.

Comparison With Prior Work

After a thorough search of previous studies, there are not many relevant studies. A study [56] investigated the performance of the n-gram, RNN, and BERT on their data set related to psychotherapy. Their results showed that the BERT model performed better than the other models. They did not attempt to optimize the performance of the present classifier, and their task only focused on DL models and simply classified the sentiment into 3 classes. However, in this study, we classified data with a more reasonable protocol, proposed a new loss function, and fine-tuned both ML and DL models. In another study [57], a new framework for evaluating a sentiment analysis model was developed. They trained CNN, LSTM, and gated recurrent units on a data set of movie reviews. The gated recurrent unit shows more explainable results related to psychological states. In contrast to this study, our models were trained on the data set of reviews from patients with mental illness, so the prediction results were more interpretable and plausible. Besides, a dictionary of medical opinions was built in the study by Asghar et al [24]. They were based on the corpus of health reviews and a medical polarity lexicon. The sentiment score was then generated by computing the word polarity score of the text data. This is a traditional and less-efficient approach, whereas our research explored advanced AI-based methods. Overall, they mechanically used various existing models without attempting to reorganize the model structure and optimize the model performance, and they did not relate the patient's emotional state to the treatment outcome. Our proposed approaches are better because we not only explore existing models but also design the UIMF approach and the FF loss function to construct fused models for optimization. Moreover, through our investigation, we provide practical suggestions on where to apply those models in clinical scenarios using different computing resources.

Conclusions

This study aimed to provide effective and accurate classification models to evaluate mental drug treatment satisfaction using reviews of patients with a mental disorders. The data set consisting of reviews of 20 disorders related to mental health was extracted from the Kaggle data sets and then preprocessed. We proposed the UIMF method, which decomposes state-of-the-art BERT models and traditional ML models into corresponding encoders and classifiers, followed by recomposing them to form 6 fused models. On the basis of our experiments, an optimal model with the highest accuracy and

a suboptimal model with fast training and fair accuracy are obtained and these can be applied in different scenarios.

Our work bridges this gap in this field, and its results can be considered as a valuable reference for psychologists. In addition, it opens up new possibilities for alleviating the conflict between the large increase in patients with mental illness and insufficient medical resources. By combining RF and transformer models, the contextual information of patients' reviews can be used to achieve better accuracy. Meanwhile, the model that combines RF and the word-embedding encoder meets the requirements of the production environment, in which the speed and efficiency of devices are limited, and the time resource is insufficient.

We believe that this research will contribute to the advancement of AI as a core method for improving mental illness treatment. DL and ML methods can analyze patients' feedback and measure their satisfaction with mental drug treatment, which can significantly improve the productivity of psychologists as well as treatment outcomes.

Acknowledgments

The authors would like to thank Kaggle for allowing access to the data, as well as all patients who contributed to the Kaggle data set, without whom this research would not have been possible. The valuable and constructive insights provided by the reviewers are greatly appreciated. PC-IP obtained funding from the Macao Polytechnic University (funding ID RP/FCA-10/2022) and the Macao Science and Technology Development Fund (funding ID 0048/2021/APD). The article processing fee of this paper was funded by the Macao Polytechnic University.

Conflicts of Interest

None declared.

References

1. Mental health and COVID-19: early evidence of the pandemic's impact: scientific brief, 2 March 2022. World Health Organization. 2022 Mar 2. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1 [accessed 2023-03-01]
2. Stigma, prejudice and discrimination against people with mental illness. American Psychiatric Association. URL: https://www.psychiatry.org/patients-families/stigma-and-discrimination [accessed 2023-03-01]
3. Syzdek BM. Client and therapist psychotherapy sentiment interaction throughout therapy. Psychol Stud 2020 Oct 09;65(4):520-530. [doi: 10.1007/s12646-020-00567-7]
4. Birjali M, Kasri M, Beni-Hssane A. A comprehensive survey on sentiment analysis: approaches, challenges and trends. Knowl Based Syst 2021 Aug 17;226:107134. [doi: 10.1016/j.knosys.2021.107134]
5. Acharya T, Agius M. The importance of hope against other factors in the recovery of mental illness. Psychiatr Danub 2017 Sep;29(Suppl 3):619-622 [FREE Full text]
6. DeAngelis T. Better relationships with patients lead to better outcomes. Monitor on Psychology. 2019 Nov 1. URL: https://www.apa.org/monitor/2019/11/ce-corner-relationships [accessed 2022-11-09]
7. Jain K, Prajapati V. NLP/deep learning techniques in healthcare for decision making. Prim Health Care 2021;11(3):373. [doi: 10.35248/2167-1079.21.11.373]
8. Brauwers G, Frasincar F. A survey on aspect-based sentiment classification. ACM Comput Surv 2022 Nov 21;55(4):1-37. [doi: 10.1145/3503044]
9. Pang B, Lee L. Using very simple statistics for review search: an exploration. 2008 Presented at: 22nd International Conference on Computational Linguistics; August 18-22, 2008; Manchester, United Kingdom.
10. Mohey El-Din D. Enhancement bag-of-words model for solving the challenges of sentiment analysis. Int J Adv Comput Sci Appl 2016;7(1):244-252. [doi: 10.14569/IJACSA.2016.070134]
11. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. J Doc 1972;28(1):11-21. [doi: 10.1108/eb026526]
12. Krouska A, Troussas C, Virvou M. The effect of preprocessing techniques on Twitter sentiment analysis. 2016 Presented at: 7th International Conference on Information, Intelligence, Systems & Applications (IISA); July 13-15, 2016; Chalkidiki, Greece. [doi: 10.1109/iisa.2016.7785373]
13. Das B, Chakraborty S. An improved text sentiment classification model using TF-IDF and next word negation. arXiv. Preprint posted online June 17, 2018 2023 [FREE Full text]
14. Yan D, Guo S. Leveraging contextual sentences for text classification by using a neural attention model. Comput Intell Neurosci 2019 Aug 01;2019:8320316 [FREE Full text] [doi: 10.1155/2019/8320316] [Medline: 31467518]
15. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. 2017 Presented at: 31st Conference on Neural Information Processing Systems; December 4-9, 2017; Long Beach, CA.
16. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1; June 2-7, 2019; Minneapolis, MN URL: https://aclanthology.org/N19-1423/

17. Burges CJ. A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 1998;2:121-167 [FREE Full text] [doi: 10.1023/A:1009715923555]
18. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/bf00994018]
19. Varghese R, Jayasree M. Aspect based sentiment analysis using support vector machine classifier. 2013 Presented at: 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI); August 22-25, 2013; Mysore, India URL: https://ieeexplore.ieee.org/document/6637416?denied= [doi: 10.1109/icacci.2013.6637416]
20. Breiman L. Classification and Regression Trees. New York, NY: Routledge; 1984.
21. Bai Q, Zhou J, He L. PG-RNN: using position-gated recurrent neural networks for aspect-based sentiment classification. J Supercomput 2021 Aug 20;78(3):4073-4094. [doi: 10.1007/s11227-021-04019-5]
22. Kim H, Jeong YS. Sentiment classification using convolutional neural networks. Appl Sci 2019 Jun 07;9(11):2347. [doi: 10.3390/app9112347]
23. Daniluk M, Rocktäschel T, Welbl J, Riedel S. Frustratingly short attention spans in neural language modeling. arXiv. Preprint posted online February 15, 2017 2023 [FREE Full text]
24. Asghar MZ, Khan A, Kundi FM, Qasim M, Saddozai FK, Ullah R, et al. Medical opinion lexicon: an incremental model for mining health reviews. Int J Acad Res 2014 Jan 30;6(1):295-302. [doi: 10.7813/2075-4124.2014/6-1/A.39]
25. Khan MT, Khalid S. Sentiment analysis for health care. In: Big Data: Concepts, Methodologies, Tools, and Applications. Hershey, PA: IGI Global; 2016.
26. Locke S, Bashall A, Al-Adely S, Moore J, Wilson A, Kitchen GB. Natural language processing in medicine: a review. Trends Anaesth Crit Care 2021 Jun;38:4-9. [doi: 10.1016/j.tacc.2021.02.007]
27. Wang S, Ren F, Lu H. A review of the application of natural language processing in clinical medicine. 2018 Presented at: 13th IEEE Conference on Industrial Electronics and Applications (ICIEA); May 31-June 02, 2018; Wuhan, China URL: https://ieeexplore.ieee.org/document/8398172 [doi: 10.1109/iciea.2018.8398172]
28. Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021 May 04;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
29. Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. NPJ Digit Med 2022 Apr 08;5(1):46 [FREE Full text] [doi: 10.1038/s41746-022-00589-7] [Medline: 35396451]
30. Uddin MZ, Dysthe KK, Følstad A, Brandtzaeg PB. Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Comput Appl 2021 Aug 27;34(1):721-744. [doi: 10.1007/S00521-021-06426-4]
31. Vandana, Marriwala N, Chaudhary D. A hybrid model for depression detection using deep learning. Meas Sens 2023 Feb;25:100587. [doi: 10.1016/j.measen.2022.100587]
32. Shah-Mohammadi F, Cui W, Bachi K, Hurd Y, Finkelstein J. Using natural language processing of clinical notes to predict outcomes of opioid treatment program. Annu Int Conf IEEE Eng Med Biol Soc 2022 Jul;2022:4415-4420 [FREE Full text] [doi: 10.1109/EMBC48229.2022.9871960] [Medline: 36085896]
33. Zhang L, Lagerberg T, Chen Q, Ghirardi L, D'Onofrio BM, Larsson H, et al. Prediction of treatment dosage and duration from free-text prescriptions: an application to ADHD medications in the Swedish prescribed drug register. Evid Based Ment Health 2021 Nov 01;24(4):146-152 [FREE Full text] [doi: 10.1136/ebmental-2020-300231] [Medline: 33795353]
34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310]
35. Kraemer HC. Kappa coefficient. In: Ruggeri F, Balakrishnan N, Longford NT, Kenett RS, Piegorsch WW, editors. Wiley StatsRef: Statistics Reference Online. Hoboken, NJ: John Wiley & Sons; Sep 29, 2014.
36. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005 Mar 1;85(3):257-268. [doi: 10.1093/ptj/85.3.257]
37. Li J, Rachel. UCI ML drug review dataset. Kaggle. URL: https://www.kaggle.com/datasets/jessicali9530/kuc-hackathon-winter-2018 [accessed 2022-09-01]
38. Harode R. WebMD drug reviews dataset. Kaggle. URL: https://www.kaggle.com/datasets/rohanharode07/webmd-drug-reviews-dataset [accessed 2022-09-01]
39. VizHub - GBD results. Institute for Health Metrics and Evaluation. URL: https://vizhub.healthdata.org/gbd-results/ [accessed 2023-03-01]
40. Sivertsen B, Hysing M, Harvey AG, Petrie KJ. The epidemiology of insomnia and sleep duration across mental and physical health: the SHoT study. Front Psychol 2021 Jun 14;12:662572 [FREE Full text] [doi: 10.3389/fpsyg.2021.662572] [Medline: 34194368]
41. emot. GitHub. URL: https://github.com/NeelShah18/emot/blob/master/emot/emo_unicode.py [accessed 2022-11-09]
42. sms_slang_translator. GitHub. URL: https://github.com/rishabhverma17/sms_slang_translator/blob/master/slang.txt [accessed 2022-11-09]
43. Cho MK. Rising to the challenge of bias in health care AI. Nat Med 2021 Dec 10;27(12):2079-2081. [doi: 10.1038/s41591-021-01577-2] [Medline: 34893774]
44. How to write a spelling corrector. Peter Norvig. 2016. URL: https://norvig.com/spell-correct.html [accessed 2023-11-22]

45. Brill E, Moore RC. An improved error model for noisy channel spelling correction. 2000 Presented at: 38th Annual Meeting of the Association for Computational Linguistics; October 3-6, 2000; Hong Kong, China. [doi: 10.3115/1075218.1075255]
46. GingerIt-Spell-Corrector-and-Grammar-Checker-using-Python. GitHub. URL: https://github.com/IshanSingh611/GingerIt-Spell-Corrector-and-Grammar-Checker-using-Python [accessed 2022-11-09]
47. Vapnik V. Principles of risk minimization for learning theory. In: Proceedings of the 4th International Conference on Neural Information Processing Systems. 1991 Presented at: NIPS'91; December 2-5, 1991; Denver, CO URL: https://dl.acm.org/doi/10.5555/2986916.2987018
48. Breiman L. Random forests. Mach Learn 2001;45:5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
49. Kingetsu H, Kobayashi K, Suzuki T. Neural network module decomposition and recomposition with superimposed masks. 2023 Presented at: 2023 International Joint Conference on Neural Networks (IJCNN); June 18-23, 2023; Gold Coast, Australia URL: https://ieeexplore.ieee.org/document/10191417
50. ALBERT: a lite BERT for self-supervised learning of language representations. TensorFlow Hub. URL: https://tfhub.dev/tensorflow/albert_en_base/3 [accessed 2022-10-01]
51. Lopez-Pacheco M, Yu W. Complex valued deep neural networks for nonlinear system modeling. Neural Process Lett 2022 Sep 23;54(1):559-580 [FREE Full text] [doi: 10.1007/s11063-021-10644-1] [Medline: 34580573]
52. Ishida T, Yamane I, Sakai T, Niu G, Sugiyama M. Do we need zero training loss after achieving zero training error? In: Proceedings of the 37th International Conference on Machine Learning. 2020 Presented at: ICML'20; July 13-18, 2020; Virtual Event.
53. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. 2017 Presented at: 2017 IEEE International Conference on Computer Vision (ICCV); October 22-29, 2017; Venice, Italy URL: https://ieeexplore.ieee.org/document/8237586 [doi: 10.1109/iccv.2017.324]
54. Loshchilov I, Hutter F. Decoupled weight decay regularization. In: 7th International Conference on Learning Representations, ICLR 2019. 2019 Presented at: ICLR 2019; May 6-9, 2019; New Orleans, LA URL: https://dblp.org/rec/conf/iclr/LoshchilovH19.html
55. Hassibi B, Stork DG, Wolff GJ. Optimal brain surgeon and general network pruning. 1993 Presented at: IEEE International Conference on Neural Networks; March 28-April 01, 1993; San Francisco, CA URL: https://ieeexplore.ieee.org/document/298572 [doi: 10.1109/icnn.1993.298572]
56. Tanana MJ, Soma CS, Kuo PB, Bertagnolli NM, Dembe A, Pace BT, et al. How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. Behav Res Methods 2021 Oct 22;53(5):2069-2082 [FREE Full text] [doi: 10.3758/s13428-020-01531-z] [Medline: 33754322]
57. Jo H, Kim SM, Ryu J. What we really want to find by sentiment analysis: the relationship between computational models and psychological state. arXiv. Preprint posted online April 11, 2017 2023 [FREE Full text]

Abbreviations

AI: artificial intelligence
BERT: Bidirectional Encoder Representations From Transformers
BOW: bag-of-words
CNN: convolutional neural network
DL: deep learning
FCNN: fully connected neural network
FF: focal flooding
LSTM: long short-term memory
ML: machine learning
NLP: natural language processing
RF: random forest
RNN: recurrent neural network
SVM: support vector machine
TF-IDF: term frequency–inverse document frequency
UIMF: Unified Interchangeable Model Fusion

Edited by J Torous; submitted 12.06.23; peer-reviewed by T Zhang, B Ofoghi; comments to author 30.06.23; revised version received 23.08.23; accepted 03.10.23; published 05.12.23.

Please cite as:
Wang Y, Yu Y, Liu Y, Ma Y, Pang PCI
Predicting Patients' Satisfaction With Mental Health Drug Treatment Using Their Reviews: Unified Interchangeable Model Fusion Approach
JMIR Ment Health 2023;10:e49894
URL: https://mental.jmir.org/2023/1/e49894
doi:10.2196/49894
PMID:38051580

Original Paper

Leveraging Symptom Search Data to Understand Disparities in US Mental Health Care: Demographic Analysis of Search Engine Trace Data

Ben Rochford[1*]; Sachin Pendse[1*], MS; Neha Kumar[1], PhD; Munmun De Choudhury[1], PhD

School of Interactive Computing, College of Computing, Georgia Institute of Technology, Atlanta, GA, United States

[*]these authors contributed equally

Corresponding Author:
Ben Rochford
School of Interactive Computing
College of Computing
Georgia Institute of Technology
756 W Peachtree St NW
Atlanta, GA, 30308
United States
Phone: 1 404 385 8603
Email: brochford3@gatech.edu

Abstract

Background: In the United States, 1 out of every 3 people lives in a mental health professional shortage area. Shortage areas tend to be rural, have higher levels of poverty, and have poor mental health outcomes. Previous work has demonstrated that these poor outcomes may arise from interactions between a lack of resources and lack of recognition of mental illness by medical professionals.

Objective: We aimed to understand the differences in how people in shortage and nonshortage areas search for information about mental health on the web.

Methods: We analyzed search engine log data related to health from 2017-2021 and examined the differences in mental health search behavior between shortage and nonshortage areas. We analyzed several axes of difference, including shortage versus nonshortage comparisons, urban versus rural comparisons, and temporal comparisons.

Results: We found specific differences in search behavior between shortage and nonshortage areas. In shortage areas, broader and more general mental health symptom categories, namely *anxiety* (mean 2.03%, SD 0.44%), *depression* (mean 1.15%, SD 0.27%), *fatigue* (mean 1.21%, SD 0.28%), and *headache* (mean 1.03%, SD 0.23%), were searched significantly more often (Q<.0003). In contrast, specific symptom categories and mental health disorders such as *binge eating* (mean 0.02%, SD 0.02%), *psychosis* (mean 0.37%, SD 0.06%), and *attention-deficit/hyperactivity disorder* (mean 0.77%, SD 0.10%) were searched significantly more often (Q<.0009) in nonshortage areas. Although suicide rates are consistently known to be higher in shortage and rural areas, we see that the rates of suicide-related *searching* are lower in shortage areas (mean 0.05%, SD 0.04%) than in nonshortage areas (mean 0.10%, SD 0.03%; Q<.0003), more so when a shortage area is rural (mean 0.024%, SD 0.029%; $Q<2 \times 10^{-12}$).

Conclusions: This study demonstrates differences in how people from geographically marginalized groups search on the web for mental health. One main implication of this work is the influence that search engine ranking algorithms and interface design might have on the kinds of resources that individuals use when in distress. Our results support the idea that search engine algorithm designers should be conscientious of the role that structural factors play in expressions of distress and they should attempt to design search engine algorithms and interfaces to close gaps in care.

(JMIR Ment Health 2023;10:e43253) doi:10.2196/43253

KEYWORDS

mental health; search engine algorithms; digital mental health; health equity

Introduction

Background

Over the course of a given year, 1 in 5 US adults will experience mental illness [1]. This ratio is starker for young adults, with nearly 1 in 2 US adolescents experiencing mental illness within a given year [1]. Most people experiencing mental illness do not receive care, with only 46.2% of US adults experiencing mental illness receiving any form of care in 2020 [1]. Early signs suggest that the stress of coping with the COVID-19 pandemic [2] has caused an increase in the prevalence of serious mental distress [3] with longitudinal impacts that are not yet fully understood.

However, the sudden and widespread move toward remote work during the pandemic has also influenced the provision of care for mental illness—what Shore et al [4] call "the rapid virtualization of psychiatric care." Given the requirements of social distancing and pandemic-incited isolation, an increasing number of individuals have turned to technology-mediated tools and resources to find help when in distress, including online support communities [5], helplines [6], resources recommended by search engines [7], teletherapy [8], and telepsychiatry [4], among other modalities. Recommendation algorithms that analyze individual language around mental health underlie how these tools suggest resources to people in need [9], including ways that may be opaque to those engaging with the technology- or algorithmically mediated support system [10].

As technology-mediated tools and resources expand access to care, structural and societal divides in American society are particularly important to study to ensure that inequities are not exacerbated by how algorithmically directed interventions are designed or deployed. Disparities in household income [11], care resources [12], race and ethnicity [13,14], sexual orientation [15,16], and gender identity [17] have all been demonstrated to have a substantial influence on how people experience mental illness and whether care is accessible. Much has also been written about the digital divide in the United States, with nearly 3 in 10 rural Americans without a broadband connection at home [18] and roughly 3 in 10 Black adults and 4 in 10 Hispanic adults without a broadband connection at home [19]. The sudden move to remote health care services during the COVID-19 pandemic made digital divides more sharply consequential [20], with some even arguing that they functioned as a social determinant of health [21,22] given that telehealth use was a primary means of care amid overburdened hospitals.

A similar divide exists in the availability of mental health care providers in the United States. In the United States, approximately 113 million people [23] live in mental health professional shortage areas (MHPSAs). MHPSAs are designated by the US Health Resources and Services Administration (HRSA) as facilities, population groups, and geographic areas in which there are "too few...mental health providers and services" to meet expected needs [24]. In this study, for brevity, we describe MHPSAs as "shortage areas" and areas that are not MHPSAs as "nonshortage areas." Although it is well recognized that economic, social, and geographic disparities influence the accessibility and use of care in shortage areas, individual and community expressions of mental health in shortage areas are not as well understood or investigated. This lack of understanding is a significant gap in providing culturally competent care and can have a severe impact on the provision of care, particularly for those in geographically sparse or hard-to-reach areas [25,26].

One way in which people understand their experience of health and learn more about symptoms is via search engine queries and the subsequent resources recommended by the engines [27,28]. A Pew Research study conducted in 2013 [29] found that over 1 in 3 Americans search on the web for information about their health experiences, and nearly 8 in 10 of those who do search on the web for information and care begin their journey through a search engine. As traces of private engagements with a technology-mediated tool, search engine data can provide valuable insights into how individuals understand their own mental health and express it to others without the limitations of external social stigma [30].

Objective

In this study, by leveraging deidentified and aggregated data from Google (Google LLC) searches regarding symptoms from 2017 to 2021 [31], we analyzed how differences in access to mental health resources relate with how individuals search for mental health conditions and related symptoms. The following research questions (RQs) were asked:

1. What are the differences between shortage and nonshortage areas regarding how people search for mental health symptoms?
2. How does the rural-urban divide interact with how people in shortage and nonshortage areas search for mental health symptoms?
3. How have patterns in searching for mental health symptoms in shortage and nonshortage areas changed over time, particularly given the onset of the COVID-19 pandemic?
4. How might search behavior concerning suicidal ideation differ between shortage and nonshortage areas?

Through our analysis, we demonstrate significant differences in how individuals in shortage and nonshortage areas articulate their experiences when seeking more information and resources regarding mental illness and health. We found that individuals tend to search for mental health less in shortage areas than in nonshortage areas. These results are consistent when accounting for the fact that shortage areas are predominantly rural. We also found that individuals in shortage areas tend to use broader (and often somatic) representations of mental illness when searching for resources, whereas individuals in nonshortage areas tend to search for more specific diagnoses and conditions that use clinical mental health language [32]. We also found distinct patterns regarding how people search for mental health and illness over time, including a specific look at suicidal ideation—a health issue with distinct prevalence and burden in shortage areas [33]. Overall, our approach demonstrates the viability of using symptom search data on the web to better understand the differences in how people understand and express their mental health experiences in resource-constrained areas.

Methods

Data

We leveraged data from the Google COVID-19 Search Trends symptoms data set [31]. Released during the COVID-19 pandemic and regularly updated, these data are "a publicly available dataset that shows aggregated, anonymized trends in Google searches for symptoms (and some related topics)" [34]. These data include searches related to 422 different symptoms, conditions, and diseases and not exclusively to mental health conditions. On their white page [32], Google described the process of aggregating and anonymizing health queries for each region in the data set. Using differential privacy, Google collected relative-to-area measures to search for queries related to the data set's target terms. Statistical noise was added to obfuscate individual queries, especially in areas with low population where generalized statistics could be more invasive to specific individuals. As a show of face validity, these data have been used in past work to predict the rates of transmission and mortality of COVID-19 in the United States, both independently [35] and in conjunction with other data sources [36].

To scope our data set, we leveraged symptom search data at a biweekly level (the most granular available time span at the time of data collection) and at the county level for all counties in the United States. For our analysis, we analyzed data from 2017 to 2021.

We limited our analysis to all areas labeled as "Geographic Area" shortage and nonshortage areas by the US HRSA. This excluded smaller facilities that were population-based (such as "migrant farmworkers") or facility-based (such as correctional facilities or "Indian Health Facilities") [24]. In addition, to classify the relative rural or urban nature of different counties, we used the 3-tiered 2013 classification system delineating rural, micropolitan, and metropolitan areas used by the National Center for Health Statistics [37,38]. As a simplification, this classification system groups both moderately populated suburban areas and densely populated urban areas as being "metropolitan." To make this analysis more specific and highlight the differences between the most rural and most urban counties, we also used the 6-tiered classification system [37] to examine the differences between the most rural counties and large "central" metropolitan areas (with populations of ≥1 million people). As some counties in the United States are classified as "partial shortage areas" [24], for a clearer comparison, we analyzed the differences between areas that were entirely classified as shortage areas or entirely not being shortage areas.

As a result of the robust process used to ensure that the search result data cannot be traced back to individual users [34], the Google COVID-19 Search Trends symptoms data set did not include data for counties that are extremely sparsely populated. In total, this included 105 counties (of the 3143 counties and county equivalents in the United States [39]), including 103 shortage area counties, 2 partial shortage counties, and 1 nonshortage county. In total, this is 3.37% of all counties or county equivalents (such as parishes or boroughs) in the United States. The specific counties that did not have available data, likely owing to extremely low population levels, are shown in Figure 1, colored in gray.

Figure 1. Map of US mental health searching percentage of total symptom search. Counties that tend toward purple have lower levels of mental health searching, whereas counties that have higher levels of mental health searching tend toward yellow. The maximum and minimum observed percentages of total searching that can be attributed to mental health can be seen in Table 1.

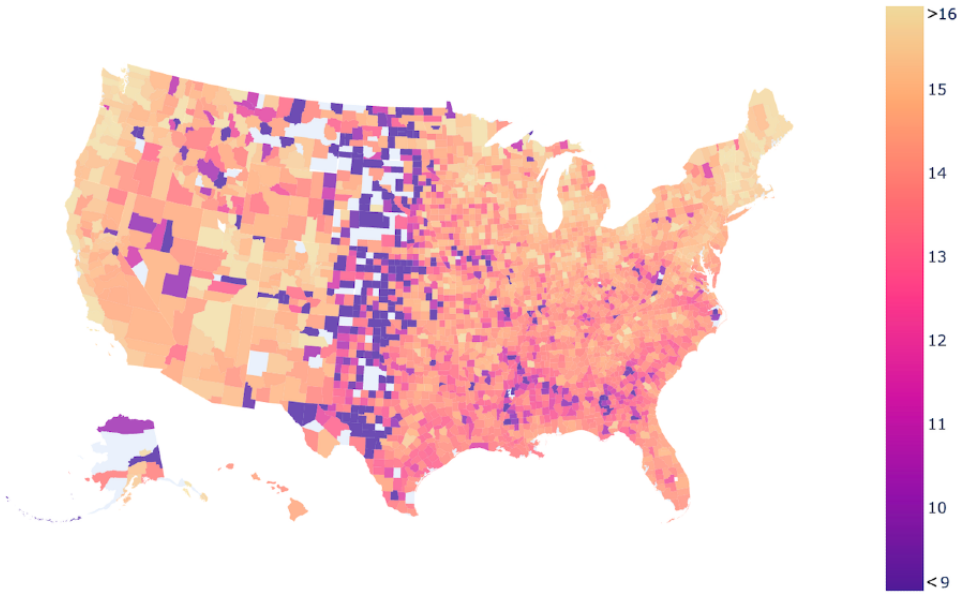


Table 1. Counties that search for mental health symptoms the most and the least.[a]

Rank	Counties with most mental health searching (%; area)	Counties with least mental health searching[b] (%; area)
1	Uintah County, Utah (24.06%; shortage area)	Taliaferro County, Georgia (3.70%; shortage area)
2	Hampshire County, Massachusetts (17.74%; nonshortage area)	Irion County, Texas (4.27%; shortage area)
3	Oglala Lakota County, South Dakota (17.60%; shortage area)	Loving County, Texas (4.47%; shortage area)
4	Boulder County, Colorado (17.27%; shortage area)	Elk County, Kansas (4.57%; shortage area)
5	Todd County, North Dakota (17.24%; shortage area)	Harper County, Oklahoma (4.74%; shortage area)
6	Whitman County, Washington (17.19%; shortage area)	Clark County, South Dakota (4.84%; shortage area)
7	Douglas County, Kansas (17.17%; nonshortage area)	Hitchcock County, Nebraska (4.99%; shortage area)
8	Williamsburg City County, Virginia (17.16%; nonshortage area)	Quitman County, Georgia (4.99%; shortage area)
9	Tompkins County, New York (17.12%; shortage area)	Echols County, Georgia (5.03%; shortage area)
10	Monroe County, Indiana (17.09%; shortage area)	Eureka, Nevada (5.05%; partial shortage area)

[a]Percentages of mental health–related searches (out of total health-related searches) are in parentheses.

[b]Note that all counties with least mental health–related searches are either shortage or partial shortage areas.

Identifying Mental Health Symptoms

As noted above, the Google COVID-19 Search Trends symptoms data set includes search topics that span both physical and mental health. To filter the particular topics that were primarily relevant to mental health, we began with a process of manual categorization by the 2 first coauthors (one of whom is a domain expert with nearly 6 years of experience in digital and global mental health). Each author selected search topics related to the conditions that were referenced in psychiatric manuals and resources. Comparing these independently generated lists resulted in a Cohen κ of 0.859, with an overall agreement of 97.6%. At this stage, we had identified 42 distinct topics pertaining specifically to mental health. We continued to refine and expand this list as described in the following paragraphs.

As discussed in previous research [40], broad and somatic symptoms can also be presentations of mental illness. As these were not captured in the initial pass of the 42 filtered symptoms above, we sought to expand our data set to include somatic symptoms that might be indicative of a mental illness. To do so, we identified symptoms that were most commonly searched on Google with the mental health diagnoses we had chosen. We leveraged Google Trends [41] to identify related search topics and queries associated with each of the 422 symptoms in our data set. We then probed these related topics and queries and created a list of symptoms and conditions that had connections to our manually generated list of 42 mental health symptoms and conditions. We included all symptoms that co-occurred with a manually selected mental health symptom or condition. As this process was likely to introduce noise in our data, we filtered out those associations that had a Google Trends relevance metric that was <2 (such as fecal incontinence or delayed onset muscle soreness). We decided to use 2 as our threshold after trying various threshold levels and optimizing the level that gave us the highest coverage with the least amount of noise in our list of symptoms. Using this strategy, we were able to discover relationships with symptoms that we had not selected manually but had important implications for mental health.

The affiliations output that we generated followed previous research in psychiatry. For example, there was a strong Google Trends signal between anxiety and depression, conforming to research in psychiatry that shows a high comorbidity between these 2 conditions [42,43]. There were also strong trends for other pairs that have been shown to have high comorbidities, such as insomnia and sleep apnea [44] or self-harm and physical scarring [45]. However, some of the strong affiliations were unexpected. For example, there was a strong association between confusion and urinary tract infections, likely attributable to the emergence of these symptoms together in older individuals potentially experiencing dementia [46].

Using these data on search affiliations from Google Trends, we were able to identify somatic symptom terminology that often co-occurred with mental health searches and were thus relevant to our analysis. To remove noise related to the COVID-19 pandemic, we manually removed any respiratory symptoms or any identified terms that were diagnosed as nonpsychiatric syndromes or broader diseases (as opposed to being solely symptoms). Our final list of 71 mental health–related search topics can be accessed in Multimedia Appendix 1.

Analytic Approach

To protect the privacy of their users [34], the Google COVID-19 Search Trends symptoms data set provides aggregate normalized weights representing how often a symptom was searched in comparison to other symptoms rather than reporting raw search frequencies. As noted in the data description [31], because of the different scaling factors, these weights were not consistent across regions or timescales. Thus, to standardize our comparisons, we divided each symptom's weight by the total summed weight of all symptoms for a given county and a given biweekly period to obtain an average percentage that an individual symptom had of all symptom searches. In our temporal analysis, we repeated this calculation but within a particular time frame, such as comparing the average percentage of a symptom between 2 given years, or in the case of our analysis, that of the influence of the pandemic on searching before and after the start of the COVID-19 pandemic. All

comparisons were performed using a statistical 2-tailed *t* test. To eliminate the potential for a significant result occurring owing to chance because of the large number of symptoms being used as comparison points, we calculated the false discovery rate, hereinafter called the *Q* value, at a significance level of .05.

Results

Shortage Areas and Nonshortage Areas

We began our analysis by broadly examining the differences in searching for mental health symptoms between shortage and nonshortage areas (*RQ1*). As shown in Figure 1, searching for mental health symptoms was broadly associated with population density, similar to the association between mental health shortage area and population density [47]. More sparsely populated areas had lower levels of mental health symptom searching, whereas more populated areas had higher levels of mental health symptom searching.

Associations could also be observed in relation to race. As seen in Figure 1, noticeably lower levels of mental health symptom search could be seen in the Black Belt region of the American South. This region, originally called the Black Belt because of the dark soil typical of the area, is home to a higher-than-average number of Black individuals [48] as a result of free descendants of enslaved individuals who were economically forced to continue living in the same area as sharecroppers or tenant farmers [49]. As discussed by Pendse et al [50], the illness constructs used to evaluate mental health symptoms were primarily trained in White, Western, and often female populations. Therefore, it may be the case that these areas have comparable levels of mental health searching but may use cultural framings of distress [51,52] that are not typically associated with dominant framings of mental illness.

In addition, Table 1 shows the counties with the highest and lowest levels of mental health symptom searches. As shown, 3 of the top 10 counties with the highest rates of mental health symptom–related searching were nonshortage areas, whereas none of the areas with the lowest rates of mental health symptom–related searching were nonshortage areas. Nationally, individuals in shortage areas generally search for mental health symptoms less often than those in nonshortage areas. As noted in Table 2, on average, 13.42% (SD 2.24%) of searches in shortage areas were for mental health symptoms, compared with 15.06% (SD 0.96%) of searches in nonshortage areas.

When we break down the individual symptoms that make up our broader "mental health symptom" construct, we see some distinct patterns. Previous studies have shown that people living in rural areas define health issues broadly [53]. Indeed, broader and more general symptom categories associated with mental health tended to be searched significantly more often in shortage areas, specifically *anxiety* (mean 2.03%, SD 0.44%; *Q*), *depression* (mean 1.15%, SD 0.27%; *Q*), *fatigue* (mean 1.21%, SD 0.28%; *Q*), and *headache* (mean 1.03%, SD 0.23%; *Q*). The more clinical versions of these terms such as *major depressive disorder* or *generalized anxiety disorder* are less distinct, with *major depressive disorder* being searched more often in shortage areas (mean 1.01%, SD 0.25%), whereas *generalized anxiety disorder* is searched more often in nonshortage areas (mean 0.25%, SD 0.03%). However, the fact that clinical language around anxiety was searched for more in nonshortage areas may be representative of a greater level of mental health literacy, awareness of clinical framings of distress in nonshortage areas, or a previously known self-diagnosis, potentially stemming from access to a greater number of mental health professionals.

These differences were more distinct when searching for topics related to more specific symptom categories and mental health disorders. Individuals in nonshortage areas were significantly more likely to search for specific symptom categories and mental health disorders than individuals in shortage areas. For example, *binge eating* (mean 0.02%, SD 0.02%; *Q*), *psychosis* (mean 0.37%, SD 0.06%; *Q*), and *attention-deficit/hyperactivity disorder* (*ADHD*; mean 0.77%, SD 0.10%; *Q*) were all searched significantly more in nonshortage areas. These are clinically recognized in the Diagnostic Manual for Mental Disorders [54] and commonly appear in clinical interview instruments used by therapists, psychiatrists, and social workers [55]. The same is true of all other specific diagnoses other than *alcoholism*, which is searched more in nonshortage areas, but the search frequency is not statistically significantly distinct from that in shortage areas.

Pertinent to *RQ1*, our results demonstrate that individuals in shortage areas were more likely than those in nonshortage areas to search for broader symptom categories, including those often associated with somatic symptom presentations of mental illness (such as headache or fatigue) [56,57]. These results point to a potentially strong relationship between the number of mental health practitioners in an area and how people come to understand and express their mental health in private and individual settings.

Table 2. Shortage versus nonshortage areas—*percentage of search*.[a,b]

	Shortage areas (%), mean (SD)	Nonshortage areas (%), mean (SD)	*Q* value
Overall average	13.42 (2.24)	15.06 (0.96)	1.26×10^{-25}
Broad symptoms			
Anxiety	2.03 (0.44)	1.8 (0.28)	6.36×10^{-14}
Depression	1.15 (0.27)	1.08 (0.11)	2.43×10^{-4}
Fatigue	1.21 (0.28)	1.05 (0.12)	5.64×10^{-15}
Headache	1.03 (0.23)	0.95 (0.07)	1.44×10^{-7}
Specific conditions			
Alcoholism	0.93 (0.29)	0.97 (0.14)	3.42×10^{-2}
Attention-deficit/hyperactivity disorder	0.73 (0.20)	0.77 (0.1)	8.80×10^{-4}
Binge eating	0.02 (0.02)	0.05 (0.02)	3.42×10^{-52}
Compulsive behavior	0.07 (0.06)	0.14 (0.03)	4.61×10^{-71}
Dysphoria	0.01 (0.02)	0.03 (0.02)	1.01×10^{-59}
Generalized anxiety disorder	0.18 (0.09)	0.25 (0.03)	7.76×10^{-27}
Hypochondriasis	0.008 (0.016)	0.017 (0.01)	4.57×10^{-18}
Hypomania	0.009 (0.02)	0.02 (0.01)	2.10×10^{-29}
Major depressive disorder	1.01 (0.25)	0.92 (0.09)	4.44×10^{-8}
Manic disorder	0.08 (0.06)	0.13 (0.03)	6.62×10^{-48}
Mood disorder	0.17 (0.09)	0.25 (0.04)	1.07×10^{-28}
Psychosis	0.30 (0.13)	0.37 (0.06)	3.80×10^{-13}
Suicidal ideation	0.05 (0.04)	0.10 (0.03)	2.57×10^{-56}

[a]Percentages indicate the percentage of all health searches, with symptoms searched significantly more often in nonshortage areas (significance level of $Q<.05$) italicized. Nonitalicized symptoms were searched significantly more often in the shortage areas.

[b]The significance level of all *Q* values were <.0003, except for those of *alcoholism* ($Q<.034$) and *attention-deficit/hyperactivity disorder* ($Q<.0009$).

The Rural-Urban-Metro Divide

To better understand whether the differences that we observed between shortage and nonshortage areas are influenced by rurality, we analyzed the differences between rural, urban, and metropolitan (areas with populations >1 million) shortage and nonshortage areas (*RQ2*). We began by examining the differences between rural shortage and rural nonshortage areas. We found that the searches for symptoms related to mental health are lower in rural shortage areas. We observed an average of 12.44% (SD 2.62%) mental health–related searches in rural shortage areas and of 14.71% (SD 2.44%) searches in rural nonshortage areas, consistent with what was observed in the broader shortage and nonshortage area comparison.

We found that broader symptoms were searched at similar rates between rural shortage and nonshortage areas. However, similar to our comparison between broad shortage and nonshortage areas, we found that though rural shortage areas search for *anxiety* more often, rural nonshortage areas search for the clinical term *generalized anxiety disorder* more often. Searches about *anxiety* ($Q=.022$) were observed significantly more frequently in rural shortage areas, accounting for an average of 2.24% (SD 0.49%) of searches related to mental health in rural shortage areas and of 1.87% (SD 0.29%) of searches in rural nonshortage areas. Searches for g*eneralized anxiety disorder* accounted for an average of 0.25% (SD 0.024%) of mental health–related searches in rural nonshortage areas and of 0.12% (SD 0.09%) of searches in rural shortage areas.

Next, consistent with our broad shortage and nonshortage area comparison, we find that more specific illness categories were searched significantly more often in rural nonshortage areas. *Compulsive behavior* ($Q=1.16 \times 10^{-15}$) and *psychosis* ($Q=.002$) were searched significantly more often in rural nonshortage areas (mean 0.121%, SD 0.049%; mean 0.383%, SD 0.12%, respectively) than in rural areas (mean 0.033%, SD 0.035%; mean 0.241%, SD 0.145%, respectively). Other specific conditions such as *manic disorder* or *mood disorder* that were searched significantly more often in rural nonshortage areas are shown in Table 3.

We then repeated this analysis for urban shortage and urban nonshortage areas to determine if the same patterns observed for rural shortage and nonshortage areas were preserved. Consistent with our past results, we noticed that individuals in urban nonshortage areas searched for mental health symptoms more often than individuals in urban shortage areas, with the percentage of all health searches being an average of 15.03% (SD 0.79%) in urban nonshortage areas and of 14.27% (SD 1.32%) in urban shortage areas.

We also observed that some broader and more somatic symptom categories, namely *fatigue* and *headache*, were higher in urban shortage areas than in urban nonshortage areas. However, broader psychiatric symptom categories were relatively similar, with no significant difference between urban shortage and nonshortage areas for searches related to *anxiety* (Q=.051) and *depression* (Q=.051), consistent with our rural analysis. Similar to the observations in our rural analysis, we found that specific conditions and disorders were searched significantly more often in urban shortage areas. As shown in Table 4, all specific conditions other than *major depressive disorder* were searched at higher rates in nonshortage areas. For example, *generalized anxiety disorder* (Q=5.69 × 10^{-7}) was searched on an average of 0.23% (SD 0.07%) occasions in urban shortage areas and of 0.24% (SD 0.03%) in urban nonshortage areas. Similarly, *compulsive behavior* (Q=2.17 × 10^{-20}) and *psychosis* (Q=.0014) were searched more often in urban nonshortage areas (mean 0.139%, SD 0.03%; mean 0.36%, SD 0.05%, respectively) than in urban shortage areas (mean 0.1%, SD 0.051%; mean 0.34%, SD 0.08%, respectively).

Analyzing our results in Tables 3 and 4 together, we see similar patterns when comparing shortage and nonshortage areas broadly. Individuals in shortage and nonshortage areas searched for broader symptom categories at roughly the same rates. However, individuals in nonshortage areas, rural or urban, relatively consistently searched for more specific conditions related to mental health. The number of mental health professionals that an area has might thus be an indication of how people in that area predominantly search for symptoms of mental illness, and this expresses how levels of mental health literacy relate to provider prevalence.

Subsequently, to investigate how rurality might influence how individuals in shortage areas express their symptoms, we examined differences in symptom searching in rural, urban, and metropolitan shortage areas. We found that rurality does influence how people frame their symptom searches. We discovered that the more rural a shortage area is, the lower their level of mental health searching is, with an average of 12.44% (SD 2.62%) of searches in rural shortage areas, 14.27% (SD 1.32%) of searches in urban shortage areas, and 15.11% (SD 0.70%) of searches in metropolitan shortage areas.

As shown in Table 5, when comparing rural and urban areas, we also found that searches for broader symptoms (such as *anxiety*, *depression*, *fatigue*, and *headache*) were significantly higher in rural shortage areas than in urban shortage areas. We also found that several specific diagnoses (such as *ADHD*, *generalized anxiety disorder*, and *psychosis* among others) are significantly more frequently searched in urban shortage areas, as shown in Table 5. This suggests that rurality is also a significant factor (along with the number of mental health professionals in an area) in influencing how people search for mental health resources on the web.

When comparing rural and metropolitan shortage areas, we observed similar patterns—searches were significantly higher for specific disorders and diagnoses in metropolitan shortage areas. However, we found that the overall degree of significance was smaller. For some of the broader symptoms, we found no statistically significant difference between the searches in rural shortage areas and metropolitan shortage areas (such as for *depression*, *major depressive disorder*, and *headache*). We did find that *anxiety* and *fatigue* were searched less frequently in metropolitan shortage areas than in rural shortage areas, consistent with our other results. Similarly, we did find that specific conditions were searched more often in metropolitan shortage areas than in rural shortage areas, as can be seen in Table 5.

Unique from other symptoms, we found no significant difference in the search rates of alcoholism among rural shortage areas, urban shortage areas, and metropolitan shortage areas (Q=.390; Q=.749). This confirms previous findings showing mixed results when attempting to examine differences in alcohol use disorders in urban and rural areas [58]. Taken together, these results enhance our understanding of the interaction between the rural and urban divide and shortage versus nonshortage discrepancies in mental health–related searching (*RQ2*).

Table 3. Rural shortage versus nonshortage areas—*percentage of search*.[a]

	Rural shortage areas (%), mean (SD)	Rural nonshortage areas (%), mean (SD)	*Q* value
Overall average	12.44 (2.62)	14.71 (2.44)	.007[b]
Broad symptoms			
Anxiety[c]	2.24 (0.49)	1.87 (0.29)	.022[d]
Depression	1.19 (0.35)	1.12 (0.26)	.550[e]
Fatigue	1.28 (0.35)	1.06 (0.25)	.057[e]
Headache	1.05 (0.30)	0.95 (0.15)	.343[e]
Specific conditions			
Alcoholism[f]	0.93 (0.37)	1.18 (0.20)	.041[d]
Attention-deficit/hyperactivity disorder	0.68 (0.25)	0.80 (0.24)	.192[e]
Binge eating	0.013 (0.03)	0.019 (0.011)	.549[e]
Compulsive behavior[f]	0.033 (0.035)	0.121 (0.049)	1.16×10^{-15}[g]
Dysphoria[f]	0.009 (0.02)	0.021 (0.02)	.042[d]
Generalized anxiety disorder[f]	0.119 (0.09)	0.25 (0.024)	1.16×10^{-6}[g]
Hypochondriasis	0.008 (0.021)	0.003 (0.002)	.540[e]
Hypomania	0.008 (0.020)	0.02 (0.03)	.330[e]
Major depressive disorder	1.05 (0.33)	0.95 (0.20)	.441[e]
Manic disorder[f]	0.040 (0.038)	0.13 (0.050)	2.13×10^{-14}[g]
Mood disorder[f]	0.117 (0.086)	0.255 (0.06)	1.48×10^{-7}[g]
Psychosis[f]	0.241 (0.145)	0.383 (0.12)	.002[b]
Suicidal ideation[f]	0.024 (0.029)	0.088 (0.053)	1.91×10^{-12}[g]

[a]Percentages indicate mental health searching out of all health searches.

[b]Significance level of $Q<.01$.

[c]Symptoms were searched significantly more in rural shortage areas.

[d]Significance level of $Q<.05$.

[e]Statistically insignificant, based on false discovery rate correction.

[f]Symptoms were searched significantly more in rural nonshortage areas.

[g]Significance level of $Q<.001$.

Table 4. Urban shortage versus nonshortage areas—*percentage of search*.[a]

	Urban shortage areas (%), mean (SD)	Urban nonshortage (%), mean (SD)	*Q* value
Overall average	14. (1.32)	15.03 (0.79)	5.71×10^{-13}[b]
Broad symptoms			
Anxiety	1.85 (0.32)	1.8 (0.3)	.051[c]
Depression	1.1 (0.17)	1.07 (0.09)	.051[c]
Fatigue[d]	1.13 (0.19)	1.05 (0.11)	1.54×10^{-7}[b]
Headache[d]	1.01 (0.15)	0.95 (0.07)	1.29×10^{-7}[b]
Specific conditions			
Alcoholism[e]	0.92 (0.19)	0.95 (0.12)	.033[f]
Attention-deficit/hyperactivity disorder[e]	0.76 (0.13)	0.77 (0.09)	.614[c]
Binge eating[e]	0.03 (0.02)	0.05 (0.02)	3.96×10^{-19}[b]
Compulsive behavior[e]	0.1 (0.051)	0.139 (0.03)	2.17×10^{-20}[b]
Dysphoria[e]	0.02 (0.018)	0.04 (0.02)	6.26×10^{-19}[b]
Generalized anxiety disorder[e]	0.23 (0.07)	0.24 (0.03)	5.69×10^{-7}[b]
Hypochondriasis[e]	0.01 (0.01)	0.02 (0.01)	9.97×10^{-20}[b]
Hypomania[e]	0.01 (0.01)	0.02 (0.01)	2.01×10^{-23}[b]
Major depressive disorder[d]	0.97 (0.16)	0.92 (0.08)	6.46×10^{-6}[b]
Manic disorder[e]	0.11 (0.05)	0.13 (0.03)	8.23×10^{-11}[b]
Mood disorder[e]	0.21 (0.07)	0.24 (0.04)	4.81×10^{-8}[b]
Psychosis[e]	0.34 (0.08)	0.36 (0.05)	.0014[g]
Suicidal ideation[e]	0.07 (0.04)	0.09 (0.03)	5.21×10^{-13}[b]

[a]Percentages indicate mental health searching out of all health searches.

[b]Significance level of $Q<.001$.

[c]Statistically insignificant, based on false discovery rate correction.

[d]Symptoms searched significantly more in urban shortage areas.

[e]Symptoms searched significantly more in urban nonshortage areas.

[f]Significance level of $Q<.05$.

[g]Significance level of $Q<.01$.

Table 5. Percentage of searches in rural versus urban versus metro shortage areas.[a]

	Rural shortage areas (%), mean (SD)	Urban shortage areas (%), mean (SD)	Metro shortage areas (%), mean (SD)	Rural versus urban, *Q* value	Rural versus metro, *Q* value
Overall average	12.44 (2.62)	14.27 (1.32)	15.11 (0.70)	2.58×10^{-68}[b]	.0003
Broad symptoms					
Anxiety[c,d]	2.24 (0.49)	1.85 (0.32)	.97 (0.14)	1.08×10^{-79}[b]	.0002[b]
Depression[d]	1.19 (0.35)	1.1 (0.17)	1.05 (0.09)	5.89×10^{-12}[b]	.16[e]
Fatigue[c,d]	1.28 (0.35)	1.13 (0.19)	0.98 (0.03)	1.54×10^{-26}[b]	.003[f]
Headache[d]	1.05 (0.30)	1.01 (0.15)	0.91 (0.02)	.00023[b]	.083[e]
Specific conditions					
Alcoholism	0.93 (0.37)	.92 (0.19)	0.97 (0.14)	.390[e]	.749[e]
Attention-deficit/hyperactivity disorder[g]	0.68 (0.25)	0.76 (0.13)	0.78 (0.07)	7.26×10^{-16}[b]	.193[e]
Binge eating[g,h]	0.013 (0.03)	0.03 (0.02)	0.06 (0.01)	6.86×10^{-61}[b]	1.18×10^{-10}[b]
Compulsive behavior[g,h]	0.033 (0.035)	0.1 (0.051)	0.19 (0.14)	1.93×10^{-206}[b]	8.82×10^{-42}[b]
Dysphoria[d,g,h]	0.009 (0.02)	0.02 (0.018)	0.05 (0.01)	1.11×10^{-48}[b]	8.83×10^{-13}[b]
Generalized anxiety disorder[g,h]	0.119 (0.09)	0.23 (0.07)	0.24 (0.007)	4.65×10^{-134}[b]	2.56×10^{-06}[b]
Hypochondriasis[g,h]	0.008 (0.021)	0.01 (0.01)	0.03 (0.002)	.0009[b]	.0017[f]
Hypomania[g,h]	0.008 (0.020)	0.01 (0.01)	0.03 (0.005)	3.29×10^{-08}[b]	7.36×10^{-05}[b]
Major depressive disorder[d]	1.05 (0.33)	0.97 (0.16)	0.88 (0.06)	2.10×10^{-09}[b]	.074[e]
Manic disorder[g,h]	0.040 (0.038)	0.11 (0.05)	0.14 (0.01)	1.37×10^{-208}[b]	2.07×10^{-20}[b]
Mood disorder[g,h]	0.117 (0.086)	0.21 (0.07)	0.24 (0.018)	1.36×10^{-132}[b]	5.87×10^{-07}[b]
Psychosis[g,h]	0.241 (0.145)	0.34 (0.08)	0.38 (0.03)	1.04×10^{-66}[b]	.0015[f]
Suicidal ideation[g,h]	0.024 (0.029)	0.07 (0.04)	0.11 (0.02)	9.11×10^{-172}[b]	8.78×10^{-22}[b]

[a]Percentages indicate the percentage of all health searches, with SDs in parentheses.

[b]Significance level of $Q<.001$.

[c]Symptoms searched more often in rural shortage areas than in metro shortage areas.

[d]Symptoms searched more often in urban shortage areas.

[e]Statistically insignificant, based on false discovery rate correction.

[f]Significance level of $Q<.01$.

[g]Symptoms searched significantly more often in rural shortage areas than in urban shortage areas.

[h]Symptoms searched more often in metropolitan shortage areas.

Temporal Analysis of Differences and the COVID-19 Pandemic

In this section, we examine the differences in search behavior regarding mental health in shortage and nonshortage areas over time (*RQ3*). As demonstrated in Figures 2 and 3, though trends tended to be similar among rural, urban, and metropolitan areas, the relative gaps (in searching) between areas tended to remain consistent over time in both shortage and nonshortage areas. The similarity in these dynamics suggests that symptom searching in rural, urban, and metropolitan areas is likely to be similarly affected by external factors or events, though to different magnitudes. In addition, as demonstrated in Figure 4, using the example of searches related to *suicidal ideation*, we also noticed that gaps between shortage and nonshortage areas also remained consistent over time.

One trend we want to draw particular attention to is the decrease in searching for mental health symptoms at the onset of the COVID-19 pandemic (March 11, 2020 [60]). As indicated by the vertical lines in Figures 2, 3, and 4, at the onset of the pandemic, individuals in the shortage and nonshortage areas began to search for mental health significantly less often. This likely demonstrates that mental health took a backseat as individuals searched for health information related to COVID. Over time, levels of mental health symptom searching stabilized to similar levels in both shortage and nonshortage areas, a finding that has also been observed in other contexts regarding

the expression of psychosocial concerns during the pandemic [61]. Researchers attribute this plateauing effect to habituation to a "new normal," given the protracted nature of the pandemic [61].

Although these overall trends tend to be similar, we did observe some differences in mental health symptom searching before and after the pandemic's onset. As shown in Table 6 for shortage areas and in Table 7 for nonshortage areas, searches for *headache*, *ADHD*, *compulsive behavior*, *dysphoria*, *hypochondriasis*, and *manic disorder* significantly increased after the start of the COVID-19 pandemic in both shortage and nonshortage areas. In addition, in both the shortage and nonshortage areas, searches for *depression*, *fatigue*, *alcoholism*, *major depressive disorder*, *mood disorder*, and *suicidal ideation* significantly declined. However, it has been observed that overall, alcohol use [59] and depressive symptoms [62,63] increased during the pandemic. A demonstrated lack of engagement with web-based resources may thus be an indication of a lack of engagement with offline resources.

Figure 2. Symptom search distributions for several mental health symptoms in shortage areas. Vertical line represents March 11, 2020, the day that the World Health Organization (WHO) declared COVID-19 a global pandemic [61]. CIs represented in transparent purple (the CIs tightly follow the mean). Note that in the broader terms on the top row, rural shortage areas search at a higher rate than other types of areas. For the more particular terms on the bottom row, rural shortage areas search at significantly lower rates.

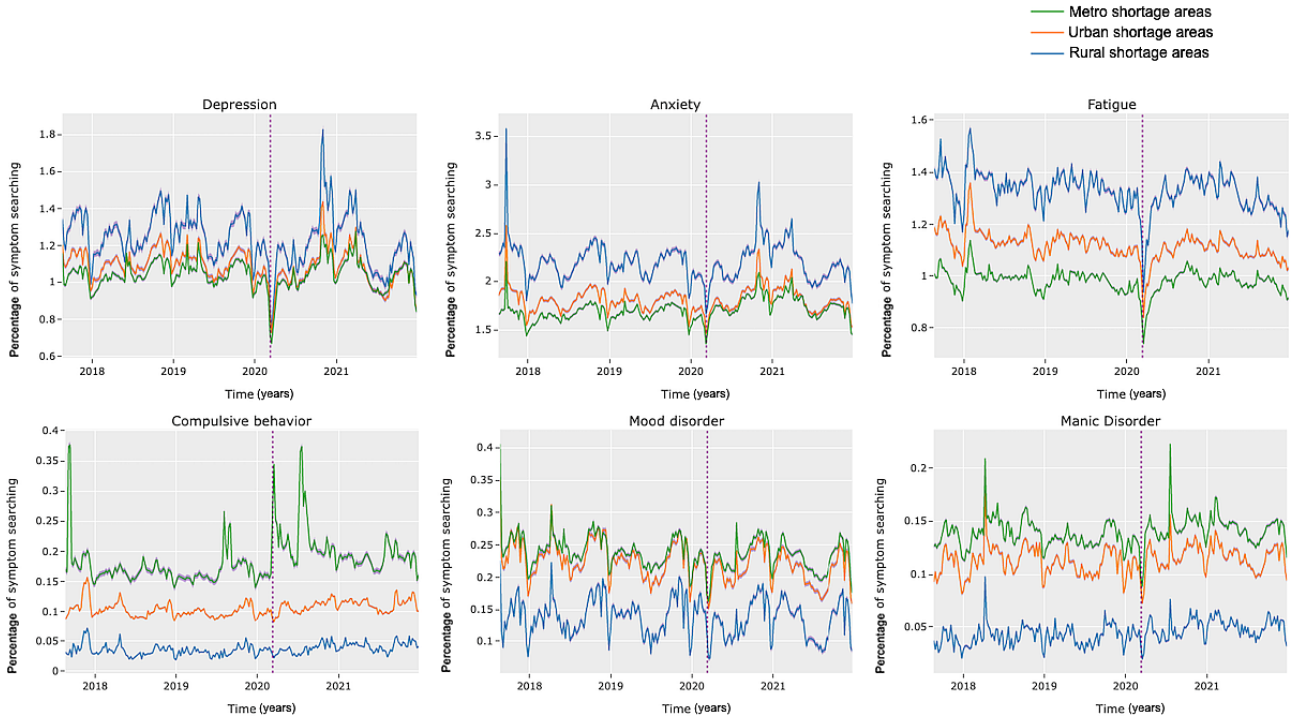


Figure 3. Symptom search distributions for several mental health symptoms in nonshortage areas. Vertical line represents March 11, 2020, the day that the World Health Organization (WHO) declared COVID-19 a global pandemic [61]. CIs represented in transparent purple (the CIs tightly follow the mean).

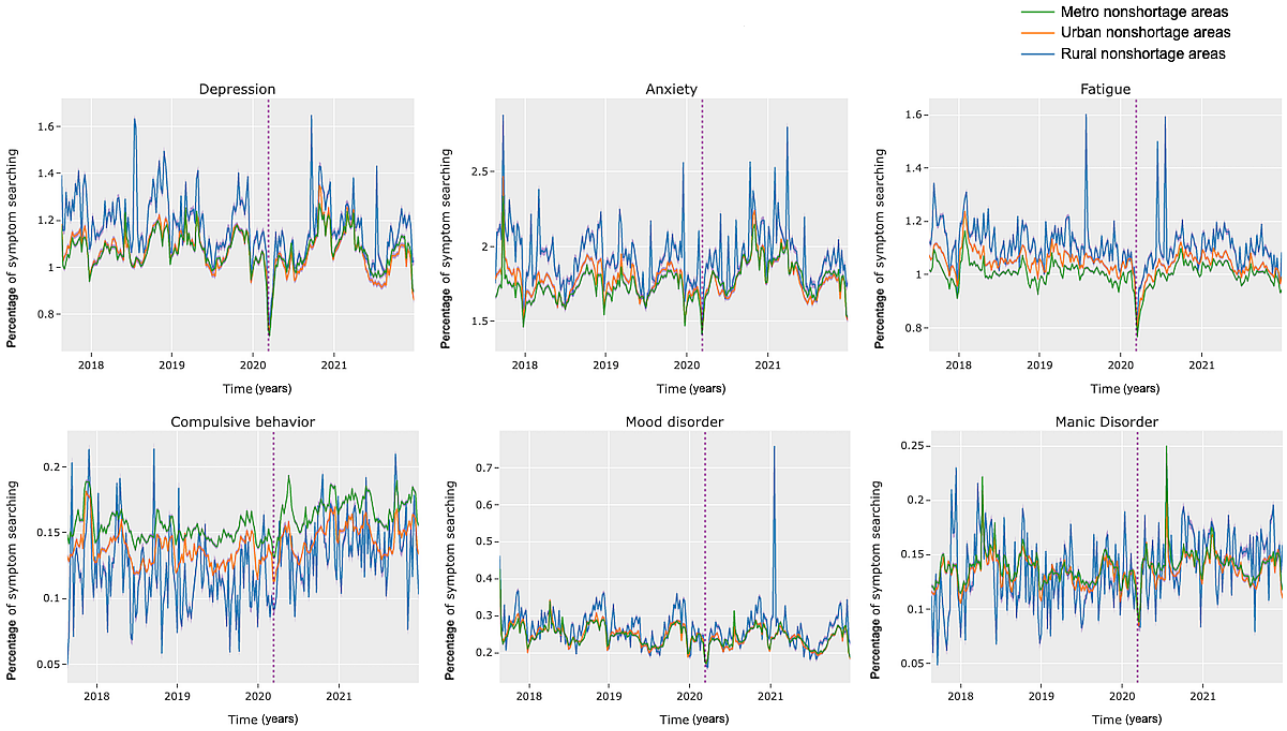


Figure 4. Symptom search distributions for suicidal ideation. Vertical line represents March 11, 2020, the day that the World Health Organization (WHO) declared COVID-19 a global pandemic [61]. Note that symptom searching in both shortage and nonshortage areas trends downward. Also note that shortage areas (with higher overall suicide rates) have lower levels of searching about suicidal ideation. Jan: January; Jul: July.

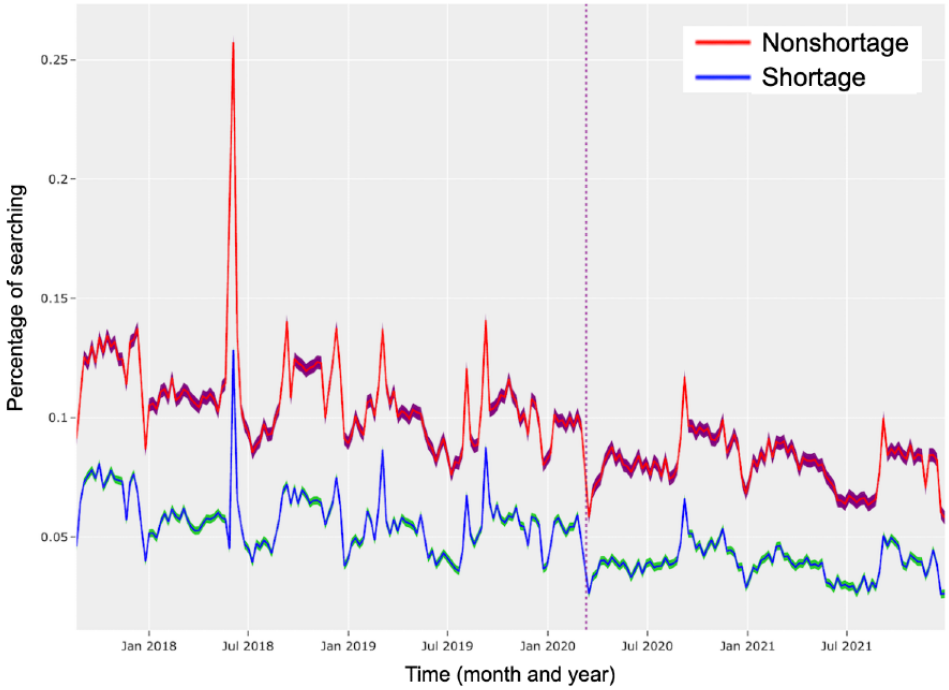


Table 6. *Percentage of search* in shortage areas during pre–COVID-19 pandemic versus post–COVID-19 pandemic.[a]

	Pre–COVID-19 pandemic (%), mean (SD)	Post–COVID-19 pandemic (%), mean (SD)	*Q* value
Overall average	13.65 (0.70)	13.89 (0.97)	.031[b]
Broad symptoms			
Anxiety	2.006 (0.153)	2.021 (0.181)	.505[c]
Depression[d]	1.186 (0.091)	1.136 (0.147)	.0023[e]
Fatigue[d]	1.242 (0.057)	1.203 (0.065)	3.95×10^{-6}[f]
Headache[g]	1.047 (0.047)	1.069 (0.055)	.0026[e]
Specific conditions			
Alcoholism[d]	0.957 (0.058)	0.928 (0.07)	.0017[e]
Attention-deficit/hyperactivity disorder[g]	0.729 (0.109)	0.788 (0.147)	.0010[e]
Binge eating	0.022 (0.005)	0.023 (0.005)	.303[c]
Compulsive behavior[g]	0.068 (0.011)	0.074 (0.009)	3.49×10^{-6}[f]
Dysphoria[g]	0.012 (0.003)	0.017 (0.004)	8.26×10^{-23}[f]
Generalized anxiety disorder	0.183 (0.019)	0.179 (0.023)	.354[c]
Hypochondriasis[g]	0.007 (0.002)	0.008 (0.002)	.008[e]
Hypomania	0.009 (0.003)	0.010 (0.002)	.048[b]
Major depressive disorder[d]	1.066 (0.088)	1.008 (0.098)	1.36×10^{-5}[f]
Manic disorder[g]	0.076 (0.012)	0.082 (0.011)	9.64×10^{-5}[f]
Mood disorder[d]	0.184 (0.029)	0.170 (0.029)	5.36×10^{-4}[f]
Psychosis[g]	0.310 (0.036)	0.322 (0.035)	.020[b]
Suicidal ideation[d]	0.057 (0.013)	0.038 (0.007)	3.48×10^{-27}[f]

[a]Percentages indicate percentage of all health searching.

[b]Significance level of $Q<.05$.

[c]Statistically insignificant, based on false discovery rate correction.

[d]Symptoms searched significantly more before March 15.

[e]Significance level of $Q<.01$.

[f]Significance level of $Q<.001$.

[g]Symptoms searched significantly more after March 15.

Table 7. *Percentage of search* in nonshortage areas during pre–COVID-19 pandemic versus post–COVID-19 pandemic.[a]

	Pre–COVID-19 pandemic (%), mean (SD)	Post–COVID-19 pandemic (%), mean (SD)	*Q* value
Overall average	14.985 (0.64)	15.234 (0.86)	.02[b]
Broad symptoms			
Anxiety[c]	1.772 (0.116)	1.808 (0.135)	.038[b]
Depression[d]	1.088 (0.106)	1.051 (0.106)	.0031[e]
Fatigue[d]	1.056 (0.050)	1.032 (0.050)	1.98×10^{-4}[f]
Headache[c]	0.936 (0.0004)	0.955 (0.0004)	6.38×10^{-4}[f]
Specific conditions			
Alcoholism[d]	0.977 (0.051)	0.943 (0.068)	5.03×10^{-5}[f]
Attention-deficit/hyperactivity disorder[c]	0.747 (0.084)	0.804 (0.124)	8.23×10^{-5}[f]
Binge eating[c]	0.049 (0.008)	0.052 (0.008)	.0042[e]
Compulsive behavior[c]	0.136 (0.012)	0.148 (0.012)	1.01×10^{-11}[f]
Dysphoria[c]	0.030 (0.005)	0.043 (0.008)	2.42×10^{-33}[f]
Generalized anxiety disorder[d]	0.251 (0.019)	0.241 (0.020)	.002[b]
Hypochondriasis[c]	0.017 (0.003)	0.020 (0.004)	2.18×10^{-9}[f]
Hypomania[c]	0.022 (0.005)	0.024 (0.003)	9.06×10^{-5}[f]
Major depressive disorder[d]	0.929 (0.069)	0.895 (0.079)	8.06×10^{-4}[f]
Manic disorder[c]	0.133 (0.014)	0.139 (0.015)	.002[e]
Mood disorder[d]	0.254 (0.031)	0.237 (0.026)	3.96×10^{-5}[f]
Psychosis	0.369 (0.031)	0.368 (0.028)	.760[g]
Suicidal ideation[d]	0.108 (0.021)	0.081 (0.010)	4.77×10^{-24}[f]

[a]Percentages indicate percentage of all health searching.

[b]Significance level of $Q<.05$.

[c]Symptoms searched significantly more after March 15.

[d]Symptoms searched significantly more before March 15.

[e]Significance level of $Q<.01$.

[f]Significance level of $Q<.001$.

[g]Statistically insignificant, based on false discovery rate correction.

Suicidal Ideation

In this section, we analyze the trends in suicidal ideation between shortage and nonshortage areas (*RQ4*). Previous research by Ku et al [33] found that the rates of death by suicide are consistently higher in shortage areas than those in nonshortage areas. In addition, Ku et al also found that the association between having a higher suicide rate and being in a shortage area has grown over time.

Overall, searches for suicidal ideation tended to be at an average of 0.05% (SD 0.04%) of all mental health–related searches in shortage areas and of 0.10% (SD 0.03%) of all the searches in nonshortage areas. This pattern was also observed in rural and urban shortage areas, with averages of 0.024% (SD 0.029%) of searches in rural shortage areas, 0.088% (SD 0.053%) of searches in rural nonshortage areas, 0.07% (SD 0.04%) of searches in urban shortage areas, and 0.09% (SD 0.03%) of searches in urban nonshortage areas. Rurality was also significantly related to how people search about suicide in shortage areas—individuals in rural shortage areas (mean 0.024%, SD 0.029%) tended to search for suicide at lower rates than individuals in urban shortage areas (mean 0.07%, SD 0.03%; $Q=9.11 \times 10^{-172}$) and even more so than those in metropolitan shortage areas (mean 0.11%, SD 0.03%; $Q=8.78 \times 10^{-22}$). Although the rates of suicide are consistently known to be higher in shortage [33] and rural areas [64], we observed that the rates of *searching* for suicidal ideation were actually lower in shortage areas ($Q<.0003$), and more so when a shortage area was rural ($Q<2 \times 10^{-12}$).

In addition, previous work has shown that the rates of suicide have increased over time (from 1999 to 2016), with the steepest rises in rural areas [65]. However, considering the rates of searching, we actually observed that the rates of searching for suicidal ideation slowly trending downward over time. We conjecture that this discrepancy might be representative of the fact that people who search for resources receive more support (via internet resources) and are thus less likely to die by suicide.

Discussion

Contextualizing Findings in Previous Research: Implications for Search Engines

We observed that search topics pertaining to specific clinical language regarding mental health were searched more often in nonshortage areas than in shortage areas, regardless of whether the shortage area was rural or urban. This finding allowed us to uncover the relationship between the number (or availability) of mental health professionals in an area and the language that people use to describe their experiences of distress (via search data). Having more specialized language to express mental health status can be key to obtaining more specific treatment options. However, individuals in shortage areas tend to use this specific language less frequently.

Search engines can help close this inequity in care by referring individuals to specialized resources regarding mental health, even when the symptom categories being searched for are broader. For example, the results in a search for "headache" might include emotional support materials and information about mental health conditions alongside broader information about the different causes of or remedies for a headache. Doing so might also moderate the potential for search engine results to escalate anxieties related to the diagnosis for a user, especially from a shortage area. White and Horvitz [66] described how the use of search engines to gain diagnostic understanding of illnesses by individuals with limited medical training may sometimes result in unfounded escalation of concerns regarding common symptomatology. The presence of emotional support and mental health resources alongside traditional health information could provide individuals with language and resources that are more in line with their experience of distress and serve to de-escalate concerns about symptoms that are likely to be nonlethal.

The need for an equitable design of search engines that attends to the diverse mental health needs of underserved populations is underscored by the fact that technology-mediated systems may be the most accessible form of care or resources available to communities in need [67]. Therefore, in addition to the search engines being intentional about the type of support resources to be directed to underserved mental health search users, we also emphasize the need to consider alternative ways of optimizing matching and personalization in search. In particular, while search engines do consider geolocation as one of many variables in tailoring search results [68], our findings show that given the nature of searches in shortage and nonshortage areas, personalization should consider social determinants of health and structural inequities in care experienced by these communities. For instance, search engines, which are equipped with the insights gained from this work, could strive to provide more educational pointers that enhance understanding of one's experience as a result of broad symptom searching in shortage areas. Similarly, as users in nonshortage areas tend to use more clinical terms to seek information on mental health, search engine algorithms may prioritize those results that provide advice complementary to formal treatment.

We note that search technologies encode certain values about what sort of content is "important" or "authoritative" [69], and previous work has discovered that a lack of consideration of equities can result in amplified biases against minoritized identities [70,71]. Although we do not discover the underlying intent or offline context of specific search behaviors, when catering to mental health–related search queries, these values would need to be punctuated with a deeper understanding of the types of searches in shortage and nonshortage areas. More specifically, we argue that to fulfill the potential of serving as an algorithmically mediated care resource, search engine design will need to decouple itself from the biases encoded in web-based advertising delivery. For instance, it has been demonstrated that search results for Black-identifying first names are associated with more advertisements for public record searches (eg, arrest records), in contrast to those for White-identifying first names [72]. Search engines would need to pay careful attention to ensure that queries for broader mental health symptoms, as observed more often in shortage areas, provide empirically grounded results and not advertisements for unverified treatments.

Contextualizing Findings in Previous Research: Implications for Public Health

Our approach demonstrates the potential of using search engine log data to better understand emerging symptom presentations and the use of care resources, expanding earlier investigations that found individuals using search engines to gather new information and resources for their health [30]. Therefore, search engine log data can be used to identify emerging symptom presentations by analyzing what searches co-occur with mental health symptoms. Moreover, from previous work, we already know that interactions, which are mediated by search engine ranking algorithms and their corresponding interfaces, can have tangible impacts on how individuals understand, describe, and present their symptoms when speaking to clinicians [73,74]. Thus, search log data, such as the one used here, could also be used to identify emerging shortage areas or nonshortage areas, even in areas with less reliable offline data. Areas in which individuals seem to tend toward searching broader symptom categories could be flagged as potential shortage areas, whereas areas with individuals who seem to tend toward searching more specific diagnoses could be flagged as potential nonshortage areas. Combined with traditional metrics from the HRSA, this information could be used in new programs to incentivize mental health professionals to work in less covered areas. Search engines could differentiate between resources based on the shortage status and corresponding mental health literacy rate that are inferred from a region's search queries. This could be particularly helpful in the case of suicidal ideation. We find that though rates of suicide are higher in shortage areas and on the increase, broadly, rates of *searching* for suicidal ideation are

lower in shortage areas and on the decline overall. Providing additional resources to support someone experiencing suicidal ideation, even when queries may not be specific to suicidal ideation, may lead individuals to discover resources to which they may not have otherwise been exposed (such as helplines, warmlines, or peer respite centers). This finding also underscores the need to augment public health informational campaigns that may be specifically targeted to underserved communities, as stigma and limited literacy around suicidal risk may be underlying factors driving lower search for information regarding suicide in shortage areas.

Ethical Implications and Future Work

On the basis of our results, we believe and argue that search engine algorithm designers have an ethical imperative to use US public health information (such as whether an area is a shortage area) and apparent search trends to offer mental health resources more readily. However, as Pendse et al [50] discussed, we also consider the fact that the form of resources recommended might influence how people come to think of their mental health state. Privileging biomedical approaches to health might be counterproductive or crowd out minoritized forms of care. An alternative approach (such as an issue-specific support group or online forum) might be more in line with how an individual understands their distress. We contend that search engine algorithm designers should ensure that resources offered to individuals who appear to be in distress in prioritized medical information cards (such as those from Google's Knowledge Panel [75]) are diverse, such as broader resources for stress, warmlines, peer support, and other forms of care that may not fit a traditional biomedical lens, as Pendse et al [50] recommend. Similarly, we are cognizant of the potential for analyses of search engine data to be used to expose users' personal information without consent, particularly the information of those living in geographically sparse or underserved areas. We appreciate the use of several privacy-preserving mechanisms in the Google COVID-19 Search Trends data set [34].

Our work highlights the importance of better understanding the differences between how individuals in shortage and nonshortage areas perceive their experience of distress and express it to others. However, the data we analyzed were anonymized metadata and did not include other axes of oppression [76,77] that have an influence on expressions of mental illness, such as race, gender identity, sexual orientation, class, caste, ethnicity, or nationality [50]. Future work must qualitatively investigate the specific idioms of distress [78] and explanatory models [79] that underlie these search queries. Our work is limited to resource-constrained areas in the United States, and future work could explore how resource constraints interact with broader mental health symptom–related searches in other countries.

Conclusions

Technology-mediated tools for mental health support (and the algorithms that underlie them) can often be the first point of contact with any form of mental health resources, particularly for those in underserved groups. The diverse ways in which individuals understand and express their mental illness thus have a direct influence on both what resources are recommended and what resources are proactively excluded. In this study, we investigated mental health symptom searching among one such underserved group or individuals in the US MHPSAs. We demonstrated strong differences between how individuals in areas with fewer mental health professionals search about their mental health, with clear implications for algorithm design and health equity. We leveraged this analysis to discuss how search engine algorithm designers might be conscientious of the role that structural factors play in expressions of distress and how they can design search engine algorithms and interfaces to close gaps in care.

Acknowledgments

The authors thank Evgeniy Gabrilovich for introduction to the Google symptom search data set and for initial conversations about the data. SP was supported partly through a Microsoft gift to MDC and NK; SP, BR, and MDC were partly supported through a grant from the National Institutes of Health (R01MH117172) and a grant from National Science Foundation Information and Intelligent Systems (2027689), both to MDC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of mental health–related search topics.
[TXT File , 1 KB - mental_v10i1e43253_app1.txt]

References

1. Mental Health Information - Statistics - Mental Illness. National Institute of Mental Health. 2022. URL: https://www.nimh.nih.gov/health/statistics/mental-illness [accessed 2022-04-20]
2. Coronavirus disease 2019 (COVID-19) : situation report, 72. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/331685 [accessed 2022-04-20]

3. Breslau J, Finucane ML, Locker AR, Baird MD, Roth EA, Collins RL. A longitudinal study of psychological distress in the United States before and during the COVID-19 pandemic. Prev Med 2021 Feb;143:106362. [doi: 10.1016/j.ypmed.2020.106362] [Medline: 33388325]
4. Shore JH, Schneck CD, Mishkind MC. Telepsychiatry and the coronavirus disease 2019 pandemic-current and future outcomes of the rapid virtualization of psychiatric care. JAMA Psychiatry 2020 Dec 01;77(12):1211-1212. [doi: 10.1001/jamapsychiatry.2020.1643] [Medline: 32391861]
5. Kaveladze B, Chang K, Siev J, Schueller SM. Impact of the COVID-19 pandemic on online obsessive-compulsive disorder support community members: survey study. JMIR Ment Health 2021 Feb 17;8(2):e26715 [FREE Full text] [doi: 10.2196/26715] [Medline: 33595449]
6. Brülhart M, Klotzbücher V, Lalive R, Reich SK. Mental health concerns during the COVID-19 pandemic as revealed by helpline calls. Nature 2021 Dec;600(7887):121-126. [doi: 10.1038/s41586-021-04099-6] [Medline: 34789873]
7. Lekkas D, Gyorda JA, Price GD, Wortzman Z, Jacobson NC. Using the COVID-19 pandemic to assess the influence of news affect on online mental health-related search behavior across the united states: integrated sentiment analysis and the circumplex model of affect. J Med Internet Res 2022 Jan 27;24(1):e32731 [FREE Full text] [doi: 10.2196/32731] [Medline: 34932494]
8. Maier CA, Riger D, Morgan-Sowada H. "It's splendid once you grow into it:" client experiences of relational teletherapy in the era of COVID-19. J Marital Fam Ther 2021 Apr;47(2):304-319 [FREE Full text] [doi: 10.1111/jmft.12508] [Medline: 33721348]
9. Cheng Q, Yom-Tov E. Do search engine helpline notices aid in preventing suicide? Analysis of archival data. J Med Internet Res 2019 Mar 26;21(3):e12235 [FREE Full text] [doi: 10.2196/12235] [Medline: 30912753]
10. Pendse SR, Sharma A, Vashistha A, De Choudhury M, Kumar N. “Can I not be suicidal on a Sunday?”: understanding technology-mediated pathways to mental health support. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 May Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan p. 1-16. [doi: 10.1145/3411764.3445410]
11. Hodgkinson S, Godoy L, Beers LS, Lewin A. Improving mental health access for low-income children and families in the primary care setting. Pediatrics 2017 Jan;139(1):e20151175 [FREE Full text] [doi: 10.1542/peds.2015-1175] [Medline: 27965378]
12. Cummings JR, Allen L, Clennon J, Ji X, Druss BG. Geographic access to specialty mental health care across high- and low-income US communities. JAMA Psychiatry 2017 May 01;74(5):476-484 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0303] [Medline: 28384733]
13. Cook BL, Trinh NH, Li Z, Hou SS, Progovac AM. Trends in racial-ethnic disparities in access to mental health care, 2004-2012. Psychiatr Serv 2017 Jan 01;68(1):9-16 [FREE Full text] [doi: 10.1176/appi.ps.201500453] [Medline: 27476805]
14. Ojeda VD, Bergstresser SM. Gender, race-ethnicity, and psychosocial barriers to mental health care: an examination of perceptions and attitudes among adults reporting unmet need. J Health Soc Behav 2008 Sep;49(3):317-334. [doi: 10.1177/002214650804900306] [Medline: 18771066]
15. Willging CE, Salvador M, Kano M. Brief reports: unequal treatment: mental health care for sexual and gender minority groups in a rural state. Psychiatr Serv 2006 Jun;57(6):867-870. [doi: 10.1176/ps.2006.57.6.867] [Medline: 16754766]
16. Hoy-Ellis CP, Ator M, Kerr C, Milford J. Innovative approaches address aging and mental health needs in LGBTQ communities. Generations 2016;40(2):56-62 [FREE Full text] [Medline: 28366982]
17. Snow A, Cerel J, Loeffler DN, Flaherty C. Barriers to mental health care for transgender and gender-nonconforming adults: a systematic literature review. Health Soc Work 2019 Aug 02;44(3):149-155. [doi: 10.1093/hsw/hlz016] [Medline: 31359065]
18. Vogels EA. Some digital divides persist between rural, urban and suburban America. Pew Research Center. 2021 Aug 19. URL: https://www.pewresearch.org/fact-tank/2021/08/19/some-digital-divides-persist-between-rural-urban-and-suburban-america/ [accessed 2022-04-20]
19. Atske S, Perrin A. Home broadband adoption, computer ownership vary by race, ethnicity in the U.S. Pew Research Center. 2021 Jul 16. URL: https://www.pewresearch.org/fact-tank/2021/07/16/home-broadband-adoption-computer-ownership-vary-by-race-ethnicity-in-the-u-s/ [accessed 2022-04-20]
20. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]
21. Clare CA. Telehealth and the digital divide as a social determinant of health during the COVID-19 pandemic. Netw Model Anal Health Inform Bioinform 2021;10(1):26 [FREE Full text] [doi: 10.1007/s13721-021-00300-y] [Medline: 33842187]
22. Zhang D, Shi L, Han X, Li Y, Jalajel NA, Patel S, et al. Disparities in telehealth utilization during the COVID-19 pandemic: findings from a nationally representative survey in the United States. J Telemed Telecare (forthcoming) 2021 Oct 11:1357633X211051677. [doi: 10.1177/1357633X211051677] [Medline: 34633882]
23. Mental Health Care Health Professional Shortage Areas (HPSAs). KFF. 2019. URL: https://www.kff.org/other/state-indicator/mental-health-care-health-professional-shortage-areas-hpsas/ [accessed 2022-04-20]
24. Shortage Designation Scoring Criteria. Health Resources and Services Administration. 2020. URL: https://bhw.hrsa.gov/shortage-designation/hpsa-criteria [accessed 2022-04-20]

25. Baquet CR, Commiskey P, Daniel Mullins C, Mishra SI. Recruitment and participation in clinical trials: socio-demographic, rural/urban, and health care access predictors. Cancer Detect Prev 2006;30(1):24-33 [FREE Full text] [doi: 10.1016/j.cdp.2005.12.001] [Medline: 16495020]
26. López CM, Qanungo S, Jenkins C, Acierno R. Technology as a means to address disparities in mental health research: a guide to "Tele-Tailoring" your research methods. Prof Psychol Res Pr 2018 Feb;49(1):57-64 [FREE Full text] [doi: 10.1037/pro0000176] [Medline: 30034085]
27. Yom-Tov E, Lampos V, Inns T, Cox IJ, Edelstein M. Providing early indication of regional anomalies in COVID-19 case counts in England using search engine queries. Sci Rep 2022 Feb 11;12(1):2373 [FREE Full text] [doi: 10.1038/s41598-022-06340-2] [Medline: 35149764]
28. Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY, Snot Force Alliance. Correlations of online search engine trends with coronavirus disease (COVID-19) incidence: infodemiology study. JMIR Public Health Surveill 2020 May 21;6(2):e19702 [FREE Full text] [doi: 10.2196/19702] [Medline: 32401211]
29. Fox S, Duggan M. Health Online 2013. Pew Research Center. 2013 Jan 15. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2022-04-20]
30. De Choudhury M, Morris MR, White RW. Seeking and sharing health information online: comparing search engines and social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Apr Presented at: CHI '14; April 26-May 1, 2014; Toronto, Canada p. 1365-1376. [doi: 10.1145/2556288.2557214]
31. COVID-19 Search Trends symptoms dataset. Google. 2022. URL: https://pair-code.github.io/covid19_symptom_dataset/?country=US [accessed 2022-04-20]
32. Pendse SR, Niederhoffer K, Sharma A. Cross-cultural differences in the use of online mental health support forums. Proc ACM Hum Comput Interact 2019 Nov 07;3(CSCW):1-29. [doi: 10.1145/3359169]
33. Ku BS, Li J, Lally C, Compton MT, Druss BG. Associations between mental health shortage areas and county-level suicide rates among adults aged 25 and older in the USA, 2010 to 2018. Gen Hosp Psychiatry 2021;70:44-50 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.02.001] [Medline: 33714795]
34. Bavadekar S, Dai A, Davis J, Desfontaines D, Eckstein I, Everett K, et al. Google COVID-19 search trends symptoms dataset: anonymization process description (version 1.0). arXiv. Preprint posted online on September 2, 2020 [FREE Full text]
35. Alruily M, Ezz M, Mostafa A, Yanes N, Abbas M, El-Manzalawy Y. Prediction of covid-19 transmission in the United States using google search trends. Comput Mater Contin 2022;71(1):1751-1768. [doi: 10.32604/cmc.2022.020714]
36. McDonald DJ, Bien J, Green A, Hu AJ, DeFries N, Hyun S, et al. Can auxiliary indicators improve COVID-19 forecasting and hotspot prediction? Proc Natl Acad Sci U S A 2021 Dec 21;118(51):e2111453118 [FREE Full text] [doi: 10.1073/pnas.2111453118] [Medline: 34903655]
37. Ingram DD, Franco SJ. 2013 NCHS urban-rural classification scheme for counties. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. 2014 Apr. URL: https://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf [accessed 2022-04-20]
38. Guy Jr JG, Haegerich TM, Evans ME, Losby JL, Young R, Jones CM. Vital signs: pharmacy-based naloxone dispensing - United States, 2012-2018. MMWR Morb Mortal Wkly Rep 2019 Aug 09;68(31):679-686 [FREE Full text] [doi: 10.15585/mmwr.mm6831e1] [Medline: 31393863]
39. 2020 Census Tallies. United States Census Bureau. 2020. URL: https://www.census.gov/geographies/reference-files/time-series/geo/tallies.html [accessed 2022-04-20]
40. Pendse SR, Karusala N, Siddarth D, Gonsalves P, Mehrotra S, Naslund JA, et al. Mental health in the global south: challenges and opportunities in HCI for development. In: Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies. 2019 Jul Presented at: COMPASS '19; July 3-5, 2019; Accra, Ghana p. 22-36. [doi: 10.1145/3314344.3332483]
41. Google Trends. Google. URL: https://trends.google.com/trends/ [accessed 2022-04-20]
42. Gorman JM. Comorbid depression and anxiety spectrum disorders. Depress Anxiety 1996;4(4):160-168. [doi: 10.1002/(SICI)1520-6394(1996)4:4<160::AID-DA2>3.0.CO;2-J] [Medline: 9166648]
43. Cummings CM, Caporino NE, Kendall PC. Comorbidity of anxiety and depression in children and adolescents: 20 years after. Psychol Bull 2014 May;140(3):816-845 [FREE Full text] [doi: 10.1037/a0034733] [Medline: 24219155]
44. Benetó A, Gomez-Siurana E, Rubio-Sanchez P. Comorbidity between sleep apnea and insomnia. Sleep Med Rev 2009 Aug;13(4):287-293. [doi: 10.1016/j.smrv.2008.09.006] [Medline: 19246219]
45. Ho W, Jones CD, Anderson W. Deliberate self-harm scars: review of the current literature. JPRAS Open 2018 Jun;16:109-116 [FREE Full text] [doi: 10.1016/j.jpra.2018.03.001] [Medline: 32158821]
46. Mayne S, Bowden A, Sundvall PD, Gunnarsson R. The scientific evidence for a potential link between confusion and urinary tract infection in the elderly is still confusing - a systematic literature review. BMC Geriatr 2019 Feb 04;19(1):32 [FREE Full text] [doi: 10.1186/s12877-019-1049-7] [Medline: 30717706]
47. Gamm L, Tai-Seale M, Stone S. White paper: meeting the mental health needs of people living in rural areas. Department of Health Policy and Management, Texas A&M University System Health Science Center. 2002 Jun. URL: https://www.

researchgate.net/publication/239531705_White_Paper_Meeting_the_Mental_Health_Needs_of_People_Living_in_Rural_Areas [accessed 2022-04-20]
48. Wimberley RC, Morris LV. The Reference Book on Regional Well-Being: U.S. Regions, the Black Belt, Appalachia. Southern Rural Development Center. 1996 Nov. URL: http://www.srdc.msstate.edu/publications/archive/203.pdf [accessed 2022-04-20]
49. Domosh M. Genealogies of race, gender, and place. Ann Am Assoc Geogr 2017 Mar 17;107(3):765-778. [doi: 10.1080/24694452.2017.1282269]
50. Pendse SR, Nkemelu D, Bidwell NJ, Jadhav S, Pathare S, De Choudhury M, et al. From treatment to healing: envisioning a decolonial digital mental health. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Apr Presented at: CHI '22; April 29-May 05, 2022; New Orleans, LA, USA p. 1-23. [doi: 10.1145/3491102.3501982]
51. Campbell RD. Revisiting African American idioms of distress: are we speaking the same mental health language? Health Soc Work 2020 Jan 07;45(1):55-58. [doi: 10.1093/hsw/hlz038] [Medline: 31989166]
52. Snowden LR. African American folk idiom and mental health services use. Cultur Divers Ethnic Minor Psychol 1999 Nov;5(4):364-370. [doi: 10.1037/1099-9809.5.4.364]
53. Wathen CN, Harris RM. "I try to take care of it myself." How rural women search for health information. Qual Health Res 2007 May;17(5):639-651. [doi: 10.1177/1049732307301236] [Medline: 17478646]
54. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th edition. Washington, DC, USA: American Psychiatric Association; 2013.
55. Spitzer RL, Williams JB, Gibbon M, First MB. The structured clinical interview for DSM-III-R (SCID). I: history, rationale, and description. Arch Gen Psychiatry 1992 Aug;49(8):624-629. [doi: 10.1001/archpsyc.1992.01820080032005] [Medline: 1637252]
56. Desai G, Chaturvedi SK. Idioms of distress. J Neurosci Rural Pract 2017 Aug;8(Suppl 1):S94-S97 [FREE Full text] [doi: 10.4103/jnrp.jnrp_235_17] [Medline: 28936079]
57. Katon W, Berg A, Robins A, Risse S. Depression: pattern of medical utilization and somatization in primary care. In: McHugh S, Vallis TM, editors. Illness Behavior: A Multidisciplinary Model. New York, NY, USA: Springer; 1986:355-364.
58. Dixon MA, Chartier KG. Alcohol use patterns among urban and rural residents: demographic and social influences. Alcohol Res 2016;38(1):69-77 [FREE Full text] [Medline: 27159813]
59. Grossman ER, Benjamin-Neelon SE, Sonnenschein S. Alcohol consumption during the COVID-19 pandemic: a cross-sectional survey of US adults. Int J Environ Res Public Health 2020 Dec 09;17(24):9189 [FREE Full text] [doi: 10.3390/ijerph17249189] [Medline: 33316978]
60. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Nov 3. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2022-04-20]
61. Saha K, Torous J, Caine E, De Choudhury M. Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. J Med Internet Res 2020 Nov 24;22(11):e22600 [FREE Full text] [doi: 10.2196/22600] [Medline: 33156805]
62. Hawes MT, Szenczy AK, Klein DN, Hajcak G, Nelson BD. Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic. Psychol Med 2022;52(14):3222-3230 [FREE Full text] [doi: 10.1017/S0033291720005358] [Medline: 33436120]
63. Racine N, McArthur BA, Cooke JE, Eirich R, Zhu J, Madigan S. Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis. JAMA Pediatr 2021 Nov 01;175(11):1142-1150 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.2482] [Medline: 34369987]
64. Americans in rural areas more likely to die by suicide. Centers for Disease Control and Prevention. 2017 Oct 5. URL: https://www.cdc.gov/media/releases/2017/p1005-rural-suicide-rates.html [accessed 2022-04-20]
65. Steelesmith DL, Fontanella CA, Campo JV, Bridge JA, Warren KL, Root ED. Contextual factors associated with county-level suicide rates in the United States, 1999 to 2016. JAMA Netw Open 2019 Sep 04;2(9):e1910936 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.10936] [Medline: 31490540]
66. White RW, Horvitz E. Cyberchondria: studies of the escalation of medical concerns in web search. ACM Trans Inf Syst 2009 Nov;27(4):1-37. [doi: 10.1145/1629096.1629101]
67. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of digital mental health for marginalized and underserved populations. Curr Treat Options Psych 2019 Jul 5;6(3):243-255. [doi: 10.1007/s40501-019-00181-z]
68. Kliman-Silver C, Hannak A, Lazer D, Wilson C, Mislove A. Location, location, location: the impact of geolocation on web search personalization. In: Proceedings of the 2015 Internet Measurement Conference. 2015 Oct Presented at: IMC '15; October 28-30, 2015; Tokyo, Japan p. 121-127. [doi: 10.1145/2815675.2815714]
69. Diaz A. Through the Google goggles: sociopolitical bias in search engine design. In: Spink A, Zimmer M, editors. Web Search: Multidisciplinary Perspectives. Berlin, Germany: Springer; 2008:11-34.
70. Kay M, Matuszek C, Munson SA. Unequal representation and gender stereotypes in image search results for occupations. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Apr Presented at: CHI '15; April 18-23, 2015; Seoul, South Korea p. 3819-3828. [doi: 10.1145/2702123.2702520]

71. Noble SU. Algorithms of Oppression. New York, NY, USA: New York University Press; Feb 2018.
72. Sweeney L. Discrimination in online ad delivery. Commun ACM 2013 May;56(5):44-54. [doi: 10.1145/2447976.2447990]
73. Iverson SA, Howard KB, Penney BK. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008 Dec;108(12):699-711. [Medline: 19075034]
74. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
75. Scull A. Dr. Google will see you now: Google's health information previews and implications for consumer health. Med Ref Serv Q 2020;39(2):165-173. [doi: 10.1080/02763869.2020.1726151] [Medline: 32329674]
76. Costanza-Chock S. Design Justice: Community-Led Practices to Build the Worlds We Need. Cambridge, MA, USA: MIT Press; 2020.
77. Collins PH. Black feminist thought in the matrix of domination. In: Collins PH, editor. Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment. Boston, MA, USA: Unwin Hyman; 1990:221-238.
78. Nichter M. Idioms of distress: alternatives in the expression of psychosocial distress: a case study from South India. Cult Med Psychiatry 1981 Dec;5(4):379-408. [doi: 10.1007/BF00054782] [Medline: 7326955]
79. Kleinman A. Patients and Healers in the Context of Culture: An Exploration of the Borderland between Anthropology, Medicine, and Psychiatry. Berkeley, CA, USA: University of California Press; 1980.

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
HRSA: Health Resources and Services Administration
MHPSA: Mental Health Professional Shortage Area
RQ: research question

Edited by J Torous; submitted 06.10.22; peer-reviewed by A Strzelecki, A Lustig; comments to author 05.11.22; revised version received 16.11.22; accepted 20.11.22; published 30.01.23.

Please cite as:
Rochford B, Pendse S, Kumar N, De Choudhury M
Leveraging Symptom Search Data to Understand Disparities in US Mental Health Care: Demographic Analysis of Search Engine Trace Data
JMIR Ment Health 2023;10:e43253
URL: https://mental.jmir.org/2023/1/e43253
doi:10.2196/43253
PMID:36716082

Original Paper

Digital Technology Use and Mental Health Consultations: Survey of the Views and Experiences of Clinicians and Young People

Raphael Rifkin-Zybutz[1], BA, BM BCh; Nicholas Turner[2], BSc, MSc; Jane Derges[1,2], MSc, PhD; Helen Bould[1,3,4], BA, MA, BM BCh; Felicity Sedgewick[5], BA, MA, PhD; Rachael Gooberman-Hill[6], MA, PhD; Myles-Jay Linton[2,5], BSc, PhD, CPsychol; Paul Moran[1,7,8], MBBS, MSc, MD; Lucy Biddle[2,7], BA, MA, PhD

[1]Centre for Academic Mental Health, Bristol University Medical School, Bristol, United Kingdom

[2]Population Health Sciences, Bristol University Medical School, Bristol, United Kingdom

[3]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom

[4]Gloucestershire Health and Care National Health Service Foundation Trust, Gloucester, United Kingdom

[5]School of Education, University of Bristol, Bristol, United Kingdom

[6]Bristol University Medical School, University of Bristol, Bristol, United Kingdom

[7]The National Institute for Health Research Applied Research Collaboration West, University Hospitals Bristol and Weston National Health Service Foundation Trust, Bristol, United Kingdom

[8]Biomedical Research Centre, University Hospitals Bristol and Weston National Health Service Foundation Trust, Bristol, United Kingdom

Corresponding Author:
Lucy Biddle, BA, MA, PhD
Population Health Sciences
Bristol University Medical School
Canynge Hall, 39 Whatley Road
Bristol, BS8 2PS
United Kingdom
Phone: 44 (0)117 928 9000
Email: lucy.biddle@bristol.ac.uk

Abstract

Background: Digital technologies play an increasingly important role in the lives of young people and have important effects on their mental health.

Objective: We aimed to explore 3 key areas of the intersection between digital technology and mental health: the views and experiences of young people and clinicians about digital technology and mental health; implementation and barriers to the UK national guidance recommendation—that the discussion of digital technology use should form a core part of mental health assessment; and how digital technology might be used to support existing consultations.

Methods: Two cross-sectional web-based surveys were conducted in 2020 between June and December, with mental health clinicians (n=99) and young people (n=320). Descriptive statistics were used to summarize the proportions. Multilinear regression was used to explore how the answers varied by gender, sexuality, and age. Thematic analysis was used to explore the contents of the extended free-text answers. Anxiety was measured using the Generalized Anxiety Disorder Questionnaire-7 (GAD-7).

Results: Digital technology use was ubiquitous among young people, with positive and negative aspects acknowledged by both clinicians and young people. Negative experiences were common (131/284, 46.1%) and were associated with increased anxiety levels among young people (GAD-7 3.29; 95% CI 1.97-4.61; *P*<.001). Although the discussion of digital technology use was regarded as important by clinicians and acceptable by young people, less than half of clinicians (42/85, 49.4%) routinely asked about the use of digital technology and over a third of young people (48/121, 39.6%) who had received mental health care had never been asked about their digital technology use. The conversations were often experienced as unhelpful. Helpful conversations were characterized by greater depth and exploration of how an individual's digital technology use related to mental health. Despite most clinicians (59/83, 71.1%) wanting training, very few (21/86, 24.4%) reported receiving training. Clinicians were open to viewing mental health data from apps or social media to help with consultations. Although young people were generally, in theory, comfortable sharing such data with health professionals, when presented with a binary choice, most reported not wanting to share social media (84/117, 71.8%) or app data (67/118, 56.8%) during consultations.

Conclusions: Digital technology use was common, and negative experiences were frequent and associated with anxiety. Over a third of young people were not asked about their digital technology use during mental health consultations, and potentially valuable information about relevant negative experiences on the web was not being captured during consultations. Clinicians would benefit from having access to training to support these discussions with young people. Although young people recognized that app data could be helpful to clinicians, they appeared hesitant to share their own data. This finding suggests that data sharing has barriers that need to be further explored.

(JMIR Ment Health 2023;10:e44064) doi:10.2196/44064

KEYWORDS

internet; adolescent; child; mental health; anxiety; patient-physician relationship

Introduction

Background

Digital technologies play an increasingly important role in the lives of young people. Survey data from the European Commission show that across Europe, 95% of young people used the internet daily in 2021, compared with 82% in 2012 [1]. Meanwhile, 87% of those aged between 12 and 15 years use social media sites or apps, and over 90% own their own smartphones [2]. Negative experiences on the web have become a common occurrence, with >50% of children reporting having had a negative experience on the web [2]. Therefore, there has been increasing concern about possible negative impacts on mental health, including increased levels of depression and anxiety, loneliness, and social isolation, in addition to exposure to cyberbullying [3-7].

It is also increasingly asserted that asking about digital technology use and its associated risks is an important part of a comprehensive assessment of young people during mental health consultations [8]. Royal College of Psychiatrists (the United Kingdom) guidance, for instance, stipulates that "questions around technology use should become a core part of biopsychosocial assessments and formulations" and highlights the need for research in this area [9]. Currently, little is known about whether this recommendation is being implemented or whether there are barriers preventing such discussions from taking place.

However, there are also positive aspects to the increasing use of digital technology. In principle, digital technology provides resources to support and educate those seeking help [10,11], a function that played a particularly important role during the COVID-19 pandemic [12]. Furthermore, there has been an increasing push to harness digital technology to support mental health, for example, the proliferation of mood-monitoring apps [13]; web-based support services during crises [14]; and novel social media platforms designed to cater for mental health, such as those providing a forum for young people to express their feelings and obtain peer support [15].

Apps provide a digital record of an individual's health and behavior that, in principle, can be used to help inform routine mental health assessments. Preliminary surveys have shown that among adults in the United States, psychiatric outpatients would generally be willing to share their social media posts with their current therapists [16]. Surveys of clinicians have shown that in practice, they had often viewed patient's electronic media such as texts and occasionally their social media posts as part of their care and had generally found this helpful [17]. However, none of these studies examined patients aged <18 years.

Objective

Therefore, we aimed to address these gaps in knowledge by conducting a web-based survey of young people and clinicians. We aimed to investigate the following three main areas of interest: (1) clinicians' and young people's attitudes toward and experiences of the digital world in relation to mental health, (2) the discussion of digital technology use by young people and clinicians during mental health consultations, and (3) the use of digital technology to aid consultations (including the potential to use novel data streams to inform the consultation).

Methods

We conducted 2 cross-sectional web-based surveys: 1 with mental health clinicians and 1 with young people.

Ethics Approval

Research ethics approval was provided by the Health Sciences Research Ethics Committee at the University of Bristol (references: 103102; 104403).

Recruitment

We recruited young people aged between 14 and 24 years from the general population. Any young person was eligible to take part, irrespective of their mental health status or experience of service use. Upon clicking the survey link, potential participants were taken to a study information sheet and were required to provide e-consent before starting to complete the survey. Parental assent was obtained for participants aged <16 years.

Eligible clinicians were health or social care workers working in the statutory or nonstatutory sector, who spent at least a proportion of their working time supporting the mental health needs of young people aged ≤24 years. Both surveys were conducted in the United Kingdom, but international participants were eligible to participate.

Convenience sampling was used, in which participants were recruited through public advertisements circulated via professional social media accounts, tagging relevant organizations such as mental health charities, youth groups, and professional bodies. The advertisements were "re-tweeted," creating a web-based snowballing. We also approached local schools, clinical networks, university services, and national

charities directly through email. Furthermore, an advertisement was placed on the UK-based young person's mental health app ("Tellmi").

Data Collection

Each survey consisted of 1 questionnaire, with complementary versions for clinicians and young people. These were hosted on SurveyMonkey for 4 months in 2020 (clinician survey: June to September; young person survey: August to December).

The clinician and young people questionnaires followed the same format and covered identical topic areas. The majority of questions were fixed choice or Likert-scale questions were used, with a smaller number of open-ended questions to provide additional qualitative insights. The full versions of both questionnaires can be found in Multimedia Appendices 1 and 2. The main topic areas included views on the effect of digital technology use on mental health, experiences of discussing digital technology use during mental health consultations, and experiences of using digital technology to help facilitate mental health consultations. In addition, young people were asked to complete the Generalized Anxiety Disorder Questionnaire 7 (GAD-7) to assess current anxiety symptoms. The GAD-7 has been validated for use in this age range [18]. Clinicians were asked about their job roles, experiences, age of young people they supported, country of practice, and length of practice.

A draft copy of the young people questionnaire was shared with a young person's advisory group panel and a further group of neurodiverse young people who offered advice on acceptability, clarity, and ease of completion. Final amendments were made to address feedback before the finalized survey was administered. The clinician survey questionnaire was piloted with a small number of clinical colleagues to refine the final draft.

Upon completion, participants of both surveys could opt to enter into a prize draw. Signposting information for mental health services was also provided at the end of the young people questionnaire for participants who wanted further support.

Analysis

Quantitative Data

Quantitative data were analyzed using Stata (version 17; StataCorp) [19]. The clinician and young person surveys were analyzed separately. Survey questions varied in design, although most were statements that were linked to the choice of the following 6 options: "Strongly Disagree," "Disagree," "Neutral," "Agree," "Strongly Agree," and "Unsure." For the purposes of analysis, these were combined into positive (agree and strongly agree), neutral (neutral and unsure), or negative (disagree and strongly disagree) categories. All other questions were statements that offered categorical responses (such as "This has Happened," "I would like this to happen," and "I would not like this to happen"), which were analyzed as categorical variables. Descriptive statistics were used to summarize the proportions.

Some questions were restricted to subgroups of the survey; for example, the questions asking young people about their experiences when receiving help with their mental health were only asked to those who reported receiving such help. Fisher exact test was used to assess differential completion rates across gender and sexuality, and logistic regression was used to assess differential completion rates across age groups.

For the young people questionnaire, exploratory analyses were performed to determine whether answers to key questions varied by demographic background. Ordinal logistic regression was used to examine whether responses varied by age, gender, and reported sexuality. The proportional odds assumption was assessed using the Brant test. Because of the low number of individuals reporting gender other than male or female, analyses checking whether responses varied across gender examined only the differences between those reporting male or female. Age was recoded into groups of people aged <16, 16 to 18, and ≥18 years. Our reference category was the most common response in each group, which was female, age <16 years, and heterosexual. Given the low numbers for most ethnic minority groups, we could not meaningfully examine how experiences or views differed across ethnicities. Demographic data were not collected in the clinician questionnaires.

Linear regression was used to assess the association between total GAD-7 scores and previous negative web-based experiences. For the purposes of this analysis, we recoded answers to the question, "I have had bad experiences online that have affected my mental health" into a binary variable of yes ("strongly agree" or "agree") or no or unsure ("Neutral," "Disagree," "Strongly Disagree," or "Unsure"). A simple unadjusted regression was used followed by a multiple linear regression, which was adjusted for possible demographic confounders including age, gender, sexuality, ethnicity, and a history of being treated for a mental health condition.

Qualitative Data

Thematic analysis was used to explore the contents of the extended free-text answers. An open inductive approach was used in which codes were derived from the data and were used to label the text according to the meaning expressed. Coding was undertaken collaboratively by LB and JD. All individual codes were displayed on a code list and then organized into thematic categories by LB. The content of each theme was then examined by retrieving all relevant labeled data.

Results

Participant Demographics

A total of 320 young people completed the survey. The sample consisted of participants who are mainly White British (271/320, 84.7%), of female sex (196/320, 61.3%), and heterosexual (214/320, 67.1%). More than one-third of the participants (121/320, 437.8%) had received help for mental health difficulties. Smartphone ownership (295/320, 92.2%) and the use of social media (289/320, 90.3%) were almost ubiquitous.

Furthermore, 99 clinicians completed the survey. Most of them had lengthy clinical experience (median of 12 years), were based in the United Kingdom (88/99, 88.9%), and worked in the statutory sector (64/99, 64.7%). They were recruited from a wide variety of roles, of which psychologists (28/99, 28.3%), psychiatrists (23/99, 23.2%), and other (11/99, 11.1%) were the most common. The "other" category was composed of a diverse

range of associated professionals including nursing assistants, well-being workers, advisers, and counselors.

Full demographic information for both surveys can be found in Tables 1 and 2.

Table 1. Young people questionnaire demographics (n=320).

Characteristics	Participants, n (%)
Age (years)[a]	
14-15	132 (41.3)
16-17	82 (25.6)
18-25	106 (33.1)
Missing	0 (0)
Ethnicity	
White	271 (84.7)
Black, African, Caribbean, or Black British	5 (1.6)
Asian or Asian British	16 (5)
Mixed or Multiple ethnic groups	21 (6.6)
Other ethnic group	7 (2.2)
Missing	0 (0)
Sexuality	
Heterosexual	214 (67.1)
LGB+[b]	97 (30.4)
Prefer not to say	8 (2.5)
Missing	1 (0.3)
Gender	
Female	196 (61.3)
Male	110 (34.4)
Any other response (including nonbinary)	12 (3.8)
Prefer not to say	2 (0.6)
Missing	0 (0)
Current occupation	
Any education	264 (82.5)
Any employment	121 (37.8)
NEET[c]	19 (6)
Missing	0 (0)
Ever received help for a mental health problem	
Yes	121 (37.8)
No	148 (46.4)
Missing	51 (16.6)
Device ownership	
Smartphone	295 (92.2)
Laptop or desktop	275 (85.9)
Tablet	112 (35)
Missing	20 (6.2)
Use of social media	
Yes	289 (90.3)
No	10 (3.1)
Missing	21 (6.6)

[a]Median (IQR): 16.9 (15.1-19.8) years.

[b]LGB+: lesbian, gay, and bisexual.

[c]NEET: Not In education, employment, or training.

Table 2. Clinician demographics (n=99).

Characteristics	Value
Sector, n (%)	
Statutory	64 (64.7)
Multiple	4 (4)
Charitable	15 (15.2)
Private	10 (10.1)
Education	5 (5.1)
Other	1 (1)
Job role, n (%)	
Psychiatrist	23 (23.2)
Nurse	8 (8.1)
Psychologist	28 (28.3)
Social worker	3 (3)
GP[a]	3 (3)
Student well-being support	10 (10.1)
Manager	4 (4)
Practitioner	9 (9.1)
Other	11 (11.1)
Age range of young people supported (years), n (%)	
<12	54 (54.5)
12-15	61 (61.6)
16-17	72 (72.7)
18-24	46 (46)
Country of practice, n (%)	
The United Kingdom	88 (88.9)
Europe (excluding the United Kingdom)	5 (5.1)
Rest of world	5 (5.1)
Missing	1 (1)
Length of practice (years), median (IQR)	12 (5.9-20)

[a]GP: general practitioner.

Views and Experiences of Digital Technology Use in Relation to Mental Well-Being

We explored young people's use of digital technology in relation to mental health and young people's and practitioners' views about the impact of digital technology on mental health. Table 3 presents a summary of the most important data quoted in this section. The full data for all the questions in this section are displayed in Table S1 in Multimedia Appendix 3.

Just over half of the young person sample (157/297, 52.9%) reported having used digital technology or web-based resources to support their mental well-being. Of the 157 young people, 128 (81.5%) answered an open-ended question regarding the resources they had used to do this. These responses were coded into the following 5 categories: formal treatment (17/128, 13.3%), talking to others (32/128, 25%), apps (34/128, 26.6%), psychoeducation (40/128, 31.2%), and distraction (50/128, 39.1%).

Most clinicians agreed that digital technology use could have harmful (91/95, 96%) and helpful (78/95, 82%) effects on young people's mental health. However, although many young people agreed that social media could be harmful (122/284, 43%) and helpful (102/284, 36%) to mental health, similar proportions

remained neutral or unsure about whether social media was harmful (101/284, 36%) or helpful (105/284, 37%).

Exploratory analyses of demographics suggested that compared with female young people, male young people were less likely to agree that social media was harmful to their mental health (odds ratio [OR] 0.31, 95% CI 0.19-0.52; *P*<.001) and more likely to agree that social media had been helpful for their mental health (OR 2.19, 95% CI 1.34-3.58; *P*=.001). Age was associated with views about the harms of social media (χ^2_2=16.7, *P*<.001) and whether it could be helpful (χ^2_2=9.5, *P*=.009), with older individuals having more skeptical views of social media. Compared with young people aged <16 years, both those aged 16 to 18 years (OR 2.36, 95% CI 1.32-4.25; *P*=.004) and >18 years (OR 3.67, 95% CI 1.31-4.25; *P*<.001) were more likely to agree that social media was harmful to their mental health. Compared with young people aged <16 years, those aged >18 years were less likely to agree that social media had been helpful for their mental health (OR 0.49, 95% CI 0.29-0.84; *P*=.009). There was no evidence that young people who identified as lesbian, gay, and bisexual (LGB+) had different views on the harms of social media (*P*=.31; Table S2 in Multimedia Appendix 3), but they were more likely to agree that social media had been beneficial for their mental health (OR 2.24, 95% CI 1.33-3.75; *P*=.002).

A large proportion of young people reported having bad web-based experiences, which affected their mental health (131/283, 46.3%). Male young people were less likely to report having these experiences than female young people (OR 0.56, 95% CI 0.34-0.91; *P*=.02). There was no evidence that the likelihood of reporting these experiences varied by age (*P*=.17) or sexuality (*P*=.12; Table S2 in Multimedia Appendix 3).

Within the linear regression analysis, reporting of bad experiences on the web was significantly associated with higher scores on the GAD-7 at the time of the survey (3.26, 95% CI 1.93-4.58; *P*<.001). This association attenuated but remained after adjusting for age, gender, sexuality, and a history of being treated for a mental health condition (1.93, 95% CI 0.72-3.14; *P*=.002).

Table 3. General attitudes and experiences of digital technology.

Questions	Participant sample	Participants, n/N (%)			Missing participants, n/N (%)
		Agree	Unsure or neutral	Disagree	
Social media helpful or harmful?					
Social media has been helpful to my mental health	YP[a]	102/284 (35.9)	105/284 (37)	77/284 (27.1)	36/320 (11.3)
Young people's digital technology use can be helpful to their mental health	C[b]	78/95 (82.1)	15/95 (15.8)	2/95 (2.1)	4/99 (4)
Social media has been harmful to my mental health	YP	122/284 (43)	101/284 (35.6)	61/284 (21.5)	36/320 (11.3)
Young people's digital technology use can be harmful to their mental health	C	91/95 (95.8)	4/95 (4.2)	0/95 (0)	4/99 (4)
Bad experiences on the web					
I have had bad experiences online that have affected my mental health	YP	131/283 (46.3)	59/283 (20.9)	93/283 (32.9)	37/320 (11.6)

[a]YP: young person (n=320).

[b]C: clinician (n=99).

Discussing Digital Technology Use During Mental Health Consultations

We explored whether practitioners ask young people about their digital technology use and its possible impacts on their well-being when an individual presents with mental health difficulties. A summary of the most important data quoted in this section can be found in Tables 4 and 5, and the full data for all the questions in this section are displayed in Table S3 in Multimedia Appendix 3.

Most clinicians (61/83, 74%) agreed that exploring digital technology use should form an essential part of mental health risk assessment. This was also generally acceptable to young people in the survey, most of whom (177/265, 66.8%) agreed, "I would find it okay to discuss my digital technology use with a health professional." There was no evidence of differences in views between genders (*P*=.70) or across different sexualities (*P*>.27; Table S2 in Multimedia Appendix 3). There was a trend toward views differing by age (χ^2_2=5.3, *P*=.07), but these results were not statistically significant. Compared with young people aged <16 years, those aged >18 years (OR 2.05, 95% CI 1.09-3.88) were more likely to find it acceptable to discuss their digital technology use.

Although most clinicians (63/84, 75%) reported feeling confident talking to young people about their digital technology use, a large proportion of young people (111/264, 42.1%) felt that health professionals did not understand the nature of their behavior on the web. Despite a high level of confidence, most clinicians (59/83, 71%) also reported wanting training on how to discuss digital technology use in a health care setting. Notably, very few health care professionals reported having

received any such training or guidance (21/86, 24%) or access to a protocol (6/86, 7%) to help guide discussions in this area.

Despite recognizing the importance of the discussion, only about half of the clinicians (48/86, 56%) reported that asking about digital technology use was integrated into their assessments, and just under half (42/85, 49%) reported asking about digital technology use routinely. Most others (34/85, 40%) only occasionally discussed the topic. This did not appear to differ across job roles (Fisher exact test, *P*=.70). There was concordance with this finding in our young person survey, in which 48 (39.6%) of the 121 young people who reported that they had received mental health help had never been asked about their digital technology use by any professional. Although gender was not associated with the likelihood of being asked about digital technology use (*P*=.23; Table S2 in Multimedia Appendix 3), age was associated (χ^2_2=8.4, *P*=.02), with those aged >18 years being much less likely to have been asked compared with those aged <16 years (OR 0.22, 95% CI 0.07-0.67; *P*=.008). There was also some evidence that young people who identified as LGB+ were more likely to be asked (OR 2.32, 95% CI 1.00-5.39; *P*=.05).

Table 4. Discussing digital technology use during mental health consultations: role and knowledge of clinicians.

Questions	Participant sample	Participants, n/N (%)			Missing participants, n/N (%)
		Agree	Unsure or neutral	Disagree	
The role of health professionals					
I would find it okay to discuss digital tech with health professionals	YP[a]	177/265 (66.8)	59/265 (22.3)	29/265 (10.9)	55/320 (17.2)
Exploring Digital Technology should form an essential part of mental health consultation	C[b]	61/83 (73.5)	18/83 (21.7)	4/83 (4.8)	16/99 (16.2)
Health professionals: understanding the world on the web					
Most health professionals do not understand the way young people use the online world	YP	111/264 (42.1)	115/264 (43.6)	38/264 (14.4)	56/320 (17.5)
Overall, I have a good understanding of how young people use digital technology	C	73/95 (76.8)	15/95 (15.8)	7/95 (7.4)	4/99 (4)
I would like training in how to talk to young people about their digital technology use	C	59/83 (71.1)	21/83 (25.3)	3/83 (3.6)	16/99 (16.2)

[a]YP: young person.

[b]C: clinician.

Table 5. Discussing digital technology use during mental health consultations: implementation.

Questions	Participants, n/N (%)
Clinician	
How often do you discuss digital technology use with young people	
Routinely	42/85 (49.4)
Occasionally	34/85 (40)
Rarely	2/85 (2.4)
Never	7/85 (8.2)
Missing	14/99 (14.1)
Have you received any training or guidance for talking to young people about their digital technology use?	
Yes	21/86 (24.4)
No	65/86 (75.6)
Missing	13/99 (13.1)
Young person	
Have you ever been asked about your digital technology use?	
Yes	73/121 (60.3)
No	48/121 (39.7)
Missing	51/172 (29.7)

Of the 121 individuals who reported receiving help with their mental health, 67 (55.4%) reported having a bad experience on the web that affected their mental health. Of them, over a third (25/67, 37%) had never been asked about their digital technology use and so had not had their relevant negative experiences on the web explored during their mental health contacts.

When asked about potential barriers to discussing digital technology use, clinicians were most likely to identify none (35/84, 42%), lack of time (29/84, 35%), and lack of knowledge (24/84, 29%). It is also notable that around a fifth (18/84, 21%) reported worrying that young people would be unwilling to have a conversation about their digital behavior.

Survey respondents were asked to describe the topics covered during conversations on the use of digital technology. The 3 most reported topics among young people were the use of social media (52/73, 71%), available mental health apps (49/73, 67%), and website signposts to help resources (43/73, 59%). Clinicians reported a very similar result, with the most discussed topics being social media (71/77, 92%), mental health apps (66/77, 86%), website signposting (60/77, 78%), and discussions about negative experiences on the web (60/77, 78%). Overall, clinicians reported asking more often about topics than young people recalled being asked.

Discussing Digital Technology Use During Mental Health Consultations—Qualitative Findings

Young people in the survey who had been asked by their clinician about their digital technology use in the context of a mental health consultation were asked whether this conversation had been helpful and to explain their answers in free texts. Of the 58 young people who provided a response, 21 (36%) reported the conversation as helpful, 34 (58%) said it was not helpful, and 3 (5%) were unsure.

Conversations about digital technology use were considered helpful when they allowed young people to talk through a negative experience on the web or to gain insights into how web-based content or behaviors could have a negative effect on their mental health. Some participants indicated that such conversations sometimes meant that they learned how to behave more safely on the web, and 1 stated that they had never had the opportunity to discuss their experiences on the web elsewhere:

> *It was helpful to break down exactly what I look at on social media and how it impacts my day-to-day life as well as my outlook on society*
>
> *It was helpful because I'd rather talk to someone about it if I'm struggling with stigma or bullying on social media than attempting to harm myself/end my life*

Another participant thought that the conversation about their digital behavior had facilitated greater insights and shared an understanding about their difficulties:

> *It helped me identify what was harmful to me at the time…I'm glad that I did talk about it as the professionals working within my care were good at responding to that and understood a lot more about me and why I felt the way I did, partly because of the content I was viewing at the time.*

However, simply receiving signposts to mental health apps or help sites was generally not considered a useful outcome and could be perceived as dismissive:

> *[Clinician] did not personally help me and immediately told me to use an app or visit websites to look for help by myself*
>
> *It was not particularly helpful as my phone does not have space to download mental health apps. We also did not talk enough about the impact of social media and negative experiences online on mental health.*

The most commonly reported reason for finding conversations unhelpful was that the clinician conveyed a "negative view" of young people's digital technology use or appeared not to understand it, which could result in the young person feeling judged or blamed:

> *It was not helpful because I felt like they were judging me for my use of social media and [saw] social media as something that is always bad.*

Digital technology seemed frowned upon:

> *'Oh, this is what's causing your anxiety. You're too glued to your computer…' It wasn't helpful because usage seems frowned upon and is treated insensitively*

Two participants found conversation about their use of technology to be intrusive, whereas 4 considered such discussions irrelevant to them (and therefore unhelpful) because they regarded their digital technology use as unproblematic.

Can Digital Technology Be Used to Assist the Mental Health Consultation?

Our third research area explored whether digital technology could be used as a tool to enhance mental health consultations with young people. Full results of this section are presented in Table S4 in Multimedia Appendix 3.

Most of the clinicians (83/91, 91%) surveyed had used digital technology to support patients. When integrating digital technology into mental health care, clinicians in this survey mostly reported "Prescribing a self-care or mood-monitoring app" (52/87, 60%), meeting with young people via web before face-to-face consultations (31/86, 36%), and offering therapy to be completely on the web (26/87, 30%). Interestingly, although most young people had either experienced or would like to experience being recommended a web-based resource (105/119, 88.2%) or an app (104/118, 88.1%), most did not want web-based appointments either before the first face-to-face meeting (66/119, 55.5%) or generally (71/118, 60.2%).

The survey also explored how novel data streams might be used to help inform mental health consultations, for example, by inviting young people to share mental health apps or social media information with their clinicians. Most clinicians (80/87, 92%) had not asked young people to share their app data with them, but most (63/87, 72%) would consider doing so. This was true for history taking, triaging, risk assessment, and monitoring.

Most clinicians (54/83, 65%) thought that having access to app data would be useful to them, with the rest (25/83, 30%) either neutral or unsure.

Most young people (170/264, 64.4%) reported that they would feel comfortable about sharing their mental health app data with a health professional. Young people who had received mental health care had generally not been asked to share data either from social media (107/117, 91.4%) or a mental health app such as mood-monitoring app (106/118, 89.8%). Despite reporting feeling comfortable with sharing data in principle, most young people did not want this to form a part of the mental health consultation, with most of young people (84/117, 71.8%) going on to state that they would not like to share data posted on social media, and a smaller majority (67/118, 56.8%) not wanting to share data from a mental health–specific app. Compared with young people aged <16 years, those aged >18 years were much more likely to want to be asked to share data from mental health apps (OR 10.51, 95% CI 2.48-44.48; *P*=.001), and there was some evidence that they were more likely to want to share posts on social media (OR 3.28, 95% CI 0.82-13.09; *P*=.09). Male participants were more likely to want to be asked to share posts made on social media (OR 3.53, 95% CI 1.06-11.80; *P*=.04), but there was no evidence of a change in attitude toward sharing app data (*P*=.49; Table S2 in Multimedia Appendix 3). There was no evidence of differing views by sexuality (*P*=.78; Table S2 in Multimedia Appendix 3).

Discussion

Principal Findings

Our study encompassed views and experiences about digital technology use and its relevance to mental health consultations among both young people and clinicians involved in delivering mental health services to children, adolescents, and young people. To our knowledge, this is the first study to explore the frequency of discussions about digital technology during mental health consultations from the perspectives of both professionals and young people.

Views and Experiences of Digital Technology Use in Relation to Mental Well-being

Our results align with prior research demonstrating that most young people now own devices that give them instant access to the web [2]. In concordance with previous research, the participants in our study often reported that their experiences on the web affected their mental health, and there was evidence that female participants were more likely to report having had such experiences [20,21]. Therefore, it follows that female young people were more likely to view social media as harmful and less likely to appraise it as helpful. This aligns with previous work that has also suggested that women might be more susceptible to the negative effects of social media [22]. The mixed impacts of digital technology use were also recognized by clinicians.

LGB+ youth were more likely to feel that social media could be helpful and less likely to think of social media as harmful. This could be explained by previous findings that web space is especially important for LGB+ individuals to find like-minded individuals and seek support from friends on the web [23]. However, the same research also identified that LGB+ individuals are more frequently abused on the web. Our sample size may have been too small to demonstrate the same effect [23].

Furthermore, we showed that even after adjusting for gender, sexuality, ethnicity, and mental health history, the reporting of these experiences was associated with higher levels of anxiety, suggesting possible important and enduring effects. This aligns with prior work showing that negative experiences on the web such as cyberbullying are linked to poor mental health outcomes, although this may be true for a wider range of experiences [24]. However, it is important to note that the cross-sectional nature of the analysis means that we are not able to establish the direction of causation, and these findings could also indicate that people with higher anxiety are more likely to have negative experiences on the web or to recall and report negative experiences on the web. Longitudinal studies are needed to further investigate this.

Discussing Digital Technology Use During Mental Health Consultations

Despite clinicians generally agreeing that enquiry about a young person's use of digital technology should form an essential part of routine mental health consultations [9], we also found that clinicians asked about digital technology use in an ad hoc manner rather than routinely. This leads to a considerable proportion of young people not having this area explored in mental health contexts. Indeed, around 40% of those who have had adverse experiences on the web that they recognized as affecting their mental health did not have these explored during mental health consultations, suggesting that this is an important area of deficit. The main barriers to discussion appeared to be time, knowledge, and lack of expertise, with most clinicians lacking specific training or protocols to help guide them.

Although our qualitative data indicate good potential for conversations to be beneficial, this lack of guidance may explain why over half of the young people in our survey who had been asked about digital technology use by their clinicians described the conversation as unhelpful. Some young people have also reported experiencing negative attitudes from practitioners regarding social media. Such experiences may underlie the belief expressed by >40% of young people that clinicians do not understand how young people engage with the web, despite most clinicians believing otherwise. This schism of opinion points to a gap in understanding, which we explored in greater depth in subsequent qualitative research and described elsewhere [25]. Similarly, the belief that young people would be unwilling to discuss their behavior on the web was also cited by some clinicians as a barrier to asking; nevertheless, two-thirds of the young people in our survey agreed that they would find it acceptable for a clinician to raise this topic, and indeed those who had been asked wished to explore this in some depth.

To address the need for guidance, a recent study by our group [26] used the Delphi format to investigate good practice indicators for discussions about digital technology use between young people and clinical professionals. This study highlighted the need for discussions to move beyond information gathering

to take the form of a deeper conversation with a focus on encouraging individuals to reflect on the meanings and impact of their behavior on the web. This builds on the findings reported here, in which young people clearly indicated that discussion is unhelpful if it is merely focused on signposting to help resources. Given the expressed desire for training and guidance and that time is identified as a barrier to having these discussions, research such as this that seeks to guide and suggest areas of focus may be of help to professionals working with young people.

Can Digital Technology Be Used to Assist the Mental Health Consultation?

In terms of using digital technology to augment mental health conversations, most clinicians reported that digital technology has become routinely involved in the delivery of services, including meeting young people on the web as well as recommending digital interventions such as apps or helpful websites. This may be partially because of changing practice since the COVID-19 pandemic, which caused a pivot to increased use of digital technology to deliver remote therapy [27]. However, although previous studies have shown a general acceptability of web-based meetings [28], our results suggest that most young people would not want web-based meetings to replace face-to-face consultations. This agrees with previous research suggesting that mental health professionals see technology as primarily a preventative or psychoeducational tool rather than a replacement for face-to-face therapy [29]. However, other studies have suggested that there is a minority of individuals who would prefer web-based to face-to-face therapy [28], which aligns with our findings that around a fifth of young people would prefer to meet on the web rather than face to face, and around a third favor meeting on the web first before meeting in person. Research capturing young people's experiences of remote provision during COVID-19 is limited to date but may shed further light going forward [30]. It seems likely that there is a mix of needs within the population that may require flexible service provision.

Clinicians reported enthusiasm for the prospect of potentially integrating data from young people's mood-monitoring apps or social media to enable better assessments, and young people overall felt comfortable sharing data with clinicians. This fits with other preliminary research in this area [31]. However, young people remained skeptical of the idea of this happening during mental health consultations, with the majority reporting that they would not want this when presented with a binary choice. In addition, there was much greater reluctance to share social media data than mental health app data. This possibly represents the fact that social media information is less curated and viewed as a personal sphere separate from the mental health consultation. In addition, the aforementioned concern expressed by many young people that clinicians do not understand how young people use the web is likely to be of considerable relevance here. Sharing personal data would demand a strong sense of trust, but low confidence in clinician understanding is likely to erode this while also heightening the fear of judgment. Methods to facilitate data sharing, including strategies to remove such barriers, warrant further investigation.

Strengths and Limitations

This research constituted an in-depth survey of a group of both young people and clinicians. This enabled a comparison between the 2 perspectives. Furthermore, the use of both quantitative and qualitative approaches enabled a more detailed contextualization of some of the quantitative findings. The surveys addressed a recommendation specified by the Royal College of Psychiatrists (the United Kingdom) and is to our knowledge the first piece of research to investigate the uptake of this advice. In addition, the collection of both gender and sexuality information for the young people enabled us to explore how views differed across these demographics.

Some limitations of this study are that the sample was self-selected and is likely to be biased toward practitioners and young people with a particular interest in digital technology and the web. On this basis, it is possible that the practitioners included in the study were those most likely to ask young people about their digital technology use and were most willing to consider integrating digital technologies in the consultations. The actual prevalence may in fact be lower. It is also possible that the discussion of digital technology has increased since our data were collected, although evidence from our ongoing work does not support this [26]. Further limitations are the lack of ethnic diversity in the sample of young people; most individuals responding to the survey were White British, although the study did include a large number of individuals with a high proportion of those who identified as LGB+ and with prior experience of mental health issues. The low diversity meant that we were unable to compare the results across different ethnicities. As with all cross-sectional studies, we cannot draw any conclusions on causation for links found, for example, between negative experiences on the web and anxiety.

It is also important to note that the survey was conducted in 2020 in the context of the COVID-19 pandemic, during which professionals' attitudes and approaches to digital technology had to shift rapidly to accommodate services as face-to-face provision was not possible. Although we asked individuals to report their prepandemic practice, it is likely that some of the attitudes reflected changes in practice that had occurred because of the pandemic. Most practitioners were required to use digital technology to assist consultations during this time, which could have inflated the proportion of reporting meeting young people through the web before face-to-face consultation and providing web-based therapy. Concerns about overuse of or reliance on social media among young people may also have been heightened [32,33]. The pandemic is likely to have changed the way that services operate and accelerated the digitization of some mental health services, which may only be partially captured by these data.

Conclusions

This is the first study to have simultaneously gathered information from clinicians and young people about digital technology use and how it is explored during mental health consultations. Over a third of young people were not asked about their digital technology use during mental health consultations, and therefore the clinicians missed valuable

information including relevant negative experiences on the web impacting mental well-being.

Clinicians are currently operating in this space with little specific training, guidance, or protocols. Almost a third cited a lack of knowledge as a barrier to discussion, and 71% (59/83) reported that they would like training in how to talk to young people about their digital technology use. Therefore, important next steps include supporting clinicians by developing and disseminating effective and efficient methods for managing these important conversations with young people. Particularly notable is that despite the demand for training, over three-quarters of the clinicians also thought they had "a good understanding" of how young people use digital technology. This confidence may be misplaced because it was not shared by young people. This highlights a potential communication and understanding gap and, in turn, a need for knowledge exchange covering fundamental issues regarding the meaning of the web within young people's lives and in relation to their mental health. Successfully addressing this gap is essential to the viability of open communication and the potential for clinicians to capitalize on digital technology use as a means for enhancing mental health care.

Currently, when discussions about digital technology use happen, young people often regard them as unhelpful because they can feel dismissed, judged, or are merely signposted to web-based help sites or apps. Conversations were helpful when the young person had the opportunity to explore the impact of their digital technology use and when such use was used as a lens to better understand mental health problems more generally. Finding spaces where clinicians can have more time to discuss digital technology in greater depth is likely important given that these conversations are likely to take longer, but lack of time is identified as a barrier to these discussions by clinicians.

Acknowledgments

LB and PM were partly funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration West. RR-Z (Academic Clinical Fellow, ACF-2021-17-018) was funded by the Health Education England/NIHR for this research project. The views expressed in this publication are those of the authors and not necessarily those of the NIHR, National Health Service, or the UK Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full survey for young people.
[PDF File (Adobe PDF File), 182 KB - mental_v10i1e44064_app1.pdf]

Multimedia Appendix 2
Full survey for clinicians.
[PDF File (Adobe PDF File), 121 KB - mental_v10i1e44064_app2.pdf]

Multimedia Appendix 3
Full details of results and regression analyses.
[DOCX File , 65 KB - mental_v10i1e44064_app3.docx]

References

1. Being young in Europe today - digital world. Eurostat. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Being_young_in_Europe_today_-_digital_world [accessed 2022-06-01]
2. Children and parents: media use and attitudes report 2022. Ofcom. 2022. URL: https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2022 [accessed 2023-03-31]
3. Karim F, Oyewande AA, Abdalla LF, Chaudhry Ehsanullah R, Khan S. Social media use and its connection to mental health: a systematic review. Cureus 2020 Jun 15;12(6):e8627 [FREE Full text] [doi: 10.7759/cureus.8627] [Medline: 32685296]
4. Lee Y, Jeon YJ, Kang S, Shin JI, Jung Y, Jung SJ. Social media use and mental health during the COVID-19 pandemic in young adults: a meta-analysis of 14 cross-sectional studies. BMC Public Health 2022 May 17;22(1):995 [FREE Full text] [doi: 10.1186/s12889-022-13409-0] [Medline: 35581597]
5. Dyson MP, Hartling L, Shulhan J, Chisholm A, Milne A, Sundar P, et al. A systematic review of social media use to discuss and view deliberate self-harm acts. PLoS One 2016;11(5):e0155813 [FREE Full text] [doi: 10.1371/journal.pone.0155813] [Medline: 27191728]

6. Biddle L, Derges J, Goldsmith C, Donovan JL, Gunnell D. Using the internet for suicide-related purposes: contrasting findings from young people in the community and self-harm patients admitted to hospital. PLoS One 2018;13(5):e0197712 [FREE Full text] [doi: 10.1371/journal.pone.0197712] [Medline: 29795637]
7. Holland G, Tiggemann M. A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. Body Image 2016 Jun;17:100-110. [doi: 10.1016/j.bodyim.2016.02.008] [Medline: 26995158]
8. Abi-Jaoude E, Naylor KT, Pignatiello A. Smartphones, social media use and youth mental health. CMAJ 2020 Feb 10;192(6):E136-E141 [FREE Full text] [doi: 10.1503/cmaj.190434] [Medline: 32041697]
9. Technology use and the mental health of children and young people. Royal College of Psychiatrists. 2020 Jan. URL: https://www.rcpsych.ac.uk/improving-care/campaigning-for-better-mental-health-policy/college-reports/2020-college-reports/Technology-use-and-the-mental-health-of-children-and-young-people-cr225 [accessed 2022-06-01]
10. Clarke AM, Chambers D, Barry MM. Bridging the digital disconnect: exploring the views of professionals on using technology to promote young people's mental health. Sch Psychol Int 2017 Aug;38(4):380-397 [FREE Full text] [doi: 10.1177/0143034317700937] [Medline: 29081558]
11. Lupton D. Young people's use of digital health technologies in the global north: narrative review. J Med Internet Res 2021 Jan 11;23(1):e18286 [FREE Full text] [doi: 10.2196/18286] [Medline: 33427684]
12. Pretorius C, Coyle D. Young people's use of digital tools to support their mental health during COVID-19 restrictions. Front Digit Health 2021;3:763876 [FREE Full text] [doi: 10.3389/fdgth.2021.763876] [Medline: 34927133]
13. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. J Med Internet Res 2017 Jul 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
14. Cohen R, Rifkin-Zybutz R, Moran P, Biddle L. Web-based support services to help prevent suicide in young people and students: a mixed-methods, user-informed review of characteristics and effective elements. Health Soc Care Community 2022 Nov;30(6):2404-2413. [doi: 10.1111/hsc.13819] [Medline: 35514202]
15. Ravaccia GG, Johnson SL, Morgan N, Lereya ST, Edbrooke-Childs J. Experiences of using the digital support tool MeeToo: mixed methods study. JMIR Pediatr Parent 2022 Oct 20;5(4):e37424 [FREE Full text] [doi: 10.2196/37424] [Medline: 36264619]
16. Rieger A, Gaines A, Barnett I, Baldassano CF, Connolly Gibbons MB, Crits-Christoph P. Psychiatry outpatients' willingness to share social media posts and smartphone data for research and clinical purposes: survey study. JMIR Form Res 2019 Aug 29;3(3):e14329 [FREE Full text] [doi: 10.2196/14329] [Medline: 31493326]
17. Hobbs KW, Monette PJ, Owoyemi P, Beard C, Rauch SL, Ressler KJ, et al. Incorporating information from electronic and social media into psychiatric and psychotherapeutic patient care: survey among clinicians. J Med Internet Res 2019 Jul 12;21(7):e13218 [FREE Full text] [doi: 10.2196/13218] [Medline: 31301127]
18. Mossman SA, Luft MJ, Schroeder HK, Varney ST, Fleck DE, Barzman DH, et al. The Generalized Anxiety Disorder 7-item scale in adolescents with generalized anxiety disorder: signal detection and validation. Ann Clin Psychiatry 2017 Nov;29(4):227-34A [FREE Full text] [Medline: 29069107]
19. StataCorp. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC; 2021.
20. Hubbard L. Online hate crime report 2020 : challenging online homophobia, biphobia and transphobia. Galop. URL: https://www.report-it.org.uk/files/online-crime-2020_0.pdf [accessed 2022-06-01]
21. Magis-Weinberg L, Gys CL, Berger EL, Domoff SE, Dahl RE. Positive and negative online experiences and loneliness in Peruvian adolescents during the COVID-19 lockdown. J Res Adolesc 2021 Sep;31(3):717-733 [FREE Full text] [doi: 10.1111/jora.12666] [Medline: 34448303]
22. Thorisdottir IE, Sigurvinsdottir R, Kristjansson AL, Allegrante JP, Lilly CL, Sigfusdottir ID. Longitudinal association between social media use and psychological distress among adolescents. Prev Med 2020 Dec;141:106270 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106270] [Medline: 33031868]
23. Ybarra ML, Mitchell KJ, Palmer NA, Reisner SL. Online social support as a buffer against online and offline peer and sexual victimization among U.S. LGBT and non-LGBT youth. Child Abuse Negl 2015 Jan;39:123-136 [FREE Full text] [doi: 10.1016/j.chiabu.2014.08.006] [Medline: 25192961]
24. Bottino SM, Bottino CM, Regina CG, Correia AV, Ribeiro WS. Cyberbullying and adolescent mental health: systematic review. Cad Saude Publica 2015 Mar;31(3):463-475 [FREE Full text] [doi: 10.1590/0102-311x00036114] [Medline: 25859714]
25. Derges J, Bould H, Gooberman-Hill R, Moran P, Linton M, Rifkin-Zybutz R, et al. Talking about social media in mental health consultations: the experiences of mental health practitioners and young people. JMIR Form Res (Preprint under consideration) 2023 (forthcoming). [doi: 10.2196/preprints.43115]
26. Biddle L, Rifkin-Zybutz R, Derges J, Turner N, Bould H, Sedgewick F, et al. Developing good practice indicators to assist mental health practitioners to converse with young people about their online activities and impact on mental health: a two-panel mixed-methods Delphi study. BMC Psychiatry 2022 Jul 19;22(1):485 [FREE Full text] [doi: 10.1186/s12888-022-04093-w] [Medline: 35854237]

27. Taylor CB, Fitzsimmons-Craft EE, Graham AK. Digital technology can revolutionize mental health services delivery: the COVID-19 crisis as a catalyst for change. Int J Eat Disord 2020 Jul;53(7):1155-1157 [FREE Full text] [doi: 10.1002/eat.23300] [Medline: 32449523]
28. Sweeney GM, Donovan CL, March S, Forbes Y. Logging into therapy: adolescent perceptions of online therapies for mental health problems. Internet Interv 2019 Mar;15:93-99 [FREE Full text] [doi: 10.1016/j.invent.2016.12.001] [Medline: 30792959]
29. Cliffe B, Croker A, Denne M, Stallard P. Clinicians' use of and attitudes towards technology to provide and support interventions in child and adolescent mental health services. Child Adolesc Ment Health 2020 May;25(2):95-101. [doi: 10.1111/camh.12362] [Medline: 32307835]
30. Nicholas J, Bell IH, Thompson A, Valentine L, Simsir P, Sheppard H, et al. Implementation lessons from the transition to telehealth during COVID-19: a survey of clinicians and young people from youth mental health services. Psychiatry Res 2021 May;299:113848 [FREE Full text] [doi: 10.1016/j.psychres.2021.113848] [Medline: 33725578]
31. Southwick L, Suh R, Kranzler E, Bradley M, Merchant RM. Perspectives of patients and therapists on social media and digital data use in mental health therapy: thematic analysis. JMIR Form Res 2022 Jul 07;6(7):e32103 [FREE Full text] [doi: 10.2196/32103] [Medline: 35797103]
32. Cellini N, Canale N, Mioni G, Costa S. Changes in sleep pattern, sense of time and digital media use during COVID-19 lockdown in Italy. J Sleep Res 2020 Aug;29(4):e13074 [FREE Full text] [doi: 10.1111/jsr.13074] [Medline: 32410272]
33. Vall-Roqué H, Andrés A, Saldaña C. The impact of COVID-19 lockdown on social network sites use, body image disturbances and self-esteem among adolescent and young women. Prog Neuropsychopharmacol Biol Psychiatry 2021 Aug 30;110:110293 [FREE Full text] [doi: 10.1016/j.pnpbp.2021.110293] [Medline: 33662532]

Abbreviations

GAD-7: Generalized Anxiety Disorder Questionnaire-7
LGB+: lesbian, gay, and bisexual
OR: odds ratio

Edited by J Torous; submitted 04.11.22; peer-reviewed by N Esfandiari, B Sindelar; comments to author 05.02.23; revised version received 19.02.23; accepted 20.02.23; published 17.04.23.

Please cite as:
Rifkin-Zybutz R, Turner N, Derges J, Bould H, Sedgewick F, Gooberman-Hill R, Linton MJ, Moran P, Biddle L
Digital Technology Use and Mental Health Consultations: Survey of the Views and Experiences of Clinicians and Young People
JMIR Ment Health 2023;10:e44064
URL: https://mental.jmir.org/2023/1/e44064
doi:10.2196/44064
PMID:37067869

Original Paper

Validation of the Attitudes Towards Psychological Online Interventions Questionnaire Among Black Americans: Cross-cultural Confirmatory Factor Analysis

Donovan Michael Ellis[1], MA; Page Lyn Anderson[1], PhD

Department of Psychology, Georgia State University, Atlanta, GA, United States

Corresponding Author:
Page Lyn Anderson, PhD
Department of Psychology
Georgia State University
Urban Life Bldg, 11th Floor
140 Decatur Street
Atlanta, GA, 30303
United States
Phone: 1 404 413 6258
Email: panderson@gsu.edu

Abstract

Background: Acceptability of digital mental health interventions is a significant predictor of treatment-seeking behavior and engagement. However, acceptability has been conceptualized and operationalized in various ways, which decreases measurement precision and leads to heterogeneous conclusions about *acceptability*. Standardized self-report measures of acceptability have been developed, which have the potential to ameliorate these problems, but none have demonstrated evidence for validation among Black communities, which limits our understanding of attitudes toward these interventions among racially minoritized groups with well-documented barriers to mental health treatment.

Objective: This study aims to examine the psychometric validity and reliability of one of the first and most widely used measures of acceptability, the Attitudes Towards Psychological Online Interventions Questionnaire, among a Black American sample.

Methods: Participants (N=254) were recruited from a large southeastern university and the surrounding metropolitan area and completed the self-report measure via a web-based survey. A confirmatory factor analysis using mean and variance adjusted weighted least squares estimation was conducted to examine the validity of the underlying hierarchical 4-factor structure proposed by the original authors of the scale. An alternative, hierarchical 2-factor structure model and bifactor model were examined for comparative fit.

Results: The findings indicated that the bifactor model demonstrated a superior fit (comparative fit index=0.96, Tucker-Lewis index=0.94, standardized root mean squared residual=0.03, and root mean square error of approximation=0.09) compared with both 2- and 4-factor hierarchical structure models.

Conclusions: The findings suggest that, within a Black American sample, there may be greater utility in interpreting the Attitudes Towards Psychological Online Interventions Questionnaire subscales as attitudinal constructs that are distinct from the global *acceptability* factor. The theoretical and practical implications for culturally responsive measurements were explored.

(JMIR Ment Health 2023;10:e43929) doi:10.2196/43929

KEYWORDS

acceptability; Black American; iCBT; internet-based cognitive behavioral therapy; digital treatment; confirmatory factor analysis; bifactor model

Introduction

Background

Black communities face persistent barriers to mental health treatment, including cost, accessibility, and stigma [1-3]. Internet-based psychological interventions that implement evidence-based techniques, including psychoeducation, behavioral activation, mindfulness strategies, and symptom tracking [4], may prove useful for improving equitable access to mental health treatment as they are often more cost-effective [5,6], private [7], and readily accessible [8]. Digital interventions that are empirically driven and incorporate elements of cognitive behavioral therapy are typically referred to as internet-based cognitive behavioral therapy (iCBT) [9]. People benefit from iCBT when paired with therapist support or used alone, although the magnitude of the effect is often higher for programs with therapist assistance [10,11] (for more conservative findings on the comparative benefit of therapist support with iCBT, see the study by Bernstein et al [12]). Although iCBT programs are effective for a variety of anxiety, mood, and substance use disorders [13,14], studies have consistently reported their underutilization by the public [15,16].

Acceptability of iCBT

Studies examining this research-to-practice gap have revealed a complex picture of user acceptance of digital mental health interventions. Although therapist-supported iCBT is generally rated as more acceptable than self-guided programs [17,18], the overall willingness to use iCBT is low. In one study, 16% of non–treatment-seeking adults reported a willingness to consider using a digital mental health intervention to address a mental health concern [19], and another study reported that only 12% of participants were "definitely interested" in internet-based treatment [20]. Overall, people reported that they significantly preferred face-to-face therapy over iCBT and other digital mental health interventions [20,21].

A problem in this budding literature is that the construct of acceptability has been defined in a variety of ways, which may contribute to heterogeneous results regarding consumer attitudes toward iCBT [22]. Retrospective study outcomes, such as treatment satisfaction, engagement, usability, and feasibility, are often used interchangeably with acceptability [23]. Other researchers propose more prospective metrics, conceptualizing acceptability as "cognitively based, positive attitudes towards such interventions" that aim to predict treatment seeking [24]. Acceptability has sometimes been operationalized with measures of similar constructs, such as outcome expectancy—the expectation that one will benefit from treatment [25]. In some studies, acceptability was operationalized using single Likert scale items measuring willingness to use an intervention [20,26,27], and in other studies, researchers developed their own measure of acceptability [19,28]. The lack of precision in conceptualization and measurement may explain why conclusions about the acceptability of iCBT vary widely across studies.

A total of 6 self-report measures of consumer acceptability of digital mental health interventions now exist, with evidence of their psychometric properties and factor structure [24,29-33]. However, reflecting existing heterogeneity in the literature, these measures operationalize acceptability in various ways. The Attitudes Towards Psychological Online Interventions (APOI) questionnaire conceptualizes acceptability as a set of positive and negative appraisals and is designed to be used with various forms of digital mental health interventions [24]. The e-Therapy Attitudes and Process Questionnaire [29] includes items specifically related to users' anticipated engagement with and short-term adherence to digital interventions. The Online Psychoeducational Intervention–Brief Attitudes Scale [32] is an abbreviated measure of attitudes (5 items) that makes the conceptual distinction that attitudes toward web-based psychoeducational interventions should incorporate elements of both psychotherapy and learning methods. In addition, 3 measures have been developed to assess working alliances in different digital contexts, akin to the therapeutic alliance fostered in face-to-face therapy [34]. The Working Alliance Inventory for guided internet interventions [30] measures the perception of an emotional attachment or collaborative bond with a digital mental health intervention, and the Working Alliance Inventory applied to virtual and augmented reality [33] measures participant comfort and trust in a virtual reality environment. Similarly, the Virtual Therapist Alliance Scale [31] measures perceptions of the therapeutic alliance with digital therapist avatars common to automated virtual reality exposure therapies. Table 1 shows the characteristics of the acceptability measures.

Table 1. Measures of acceptability toward digital mental health interventions.

Study	Title	Abbreviation	Intervention modality
Clough et al [29], 2019	e-Therapy Attitudes and Process Questionnaire	eTAP	All
Gómez Penedo et al [30], 2020	Working Alliance Inventory for Guided Internet Interventions	WAI-I	Guided interventions
Miloff et al [31], 2020	Virtual Therapist Alliance Scale	VTAS	Augmented and virtual reality
Miragall et al [33], 2015	Working Alliance Inventory Applied to Virtual and Augmented Reality	WAI-VAR	Augmented and virtual reality
Schröder et al [24], 2015	Attitudes Towards Psychological Online Interventions Questionnaire	APOI	All
Teles et al [32], 2021	Online Psychoeducational Intervention—Brief Attitudes Scale	OPI-BAS	Psychoeducation

Racially Minoritized Communities Are Underrepresented in Acceptability Research

Further complicating matters are the dearth of acceptability research that is inclusive of ethnically or racially minoritized communities. In 1 meta-analysis, 62 of 64 randomized controlled trials examining the efficacy and acceptability of iCBT did not include (or did not report) racial minorities in their studies [13]. All but one [33] of the existing measures of consumer attitudes toward digital mental health interventions have collected data from White majority (and predominantly European language) samples [24,29-32], including the first and most highly cited measure of acceptability toward digital mental health interventions, the APOI questionnaire [24]. The APOI was developed with German-speaking participants who reported mild to moderate depression (N=1013) and were recruited from outpatient clinics, web-based health forums, and health insurance referrals.

No research to date has evaluated the reliability or validity of the APOI scale among racially or ethnically minoritized communities, including Black Americans. This is highly problematic because even though Black communities may disproportionately benefit from the advantages afforded by iCBT and related digital mental health interventions, it is unknown whether the APOI demonstrates good psychometric properties in this population.

This Study

This study addresses this problem by assessing the psychometric properties of the APOI questionnaire in a sample of Black Americans. Using confirmatory factor analyses, this study examined whether the APOI demonstrates reliability and construct validity within a Black population. In this study, 2 measurement models were examined using 16 ordered categorical (ordinal) response items retained in the exploratory factor analysis of the APOI. The first model presents a 2-factor, hierarchical measurement model (positive and negative subfactors) distinct from the 4-factor hierarchical model proposed by Schröder et al [24]. Given considerations for equivalent models [35,36] modification indexes will be reviewed to examine new and replicative factor structures to illuminate the underlying construct of *acceptability*.

Methods

Recruitment

Participants were self-identified Black or African American adults (N=254 participants). The participants ranged in age from 18 to 85 (mean 27.11, SD 13.40) years and were predominantly women (172.7/254, 68%), single (167.6/254, 66%), and highly educated (at least 70% had some college education; see Table 2 for more demographic and clinical characteristics of the sample). Participants were recruited from 2 primary sources: students recruited from the participant pool of a southeastern university in an urban setting who received course credit for their participation and community participants who were solicited in public places throughout the metropolitan area (eg, parks) and had the opportunity to enter a raffle for a US $25 Amazon gift card.

Table 2. Demographics and clinical characteristics of participants.

Variables	Values
Age (years; n=254), mean (SD)	27.11 (13.40)
Sex (n=254), n (%)	
Male	82 (32.3)
Female	172 (67.7)
Sexual identity (n=252), n (%)	
Heterosexual	210 (83.3)
Lesbian, gay, and bisexual	36 (14.3)
Self-identify	6 (2.4)
Current education status (n=253), n (%)	
High school	1 (0.4)
Some college or currently in college	173 (68.1)
Graduate or professional degree	5 (2.0)
Nondegree student or other	3 (1.2)
Nonstudent[a]	71 (28.0)
Relationship status (n=252), n (%)	
Single	166 (65.9)
Serious dating or committed relationship	55 (21.8)
Married or civil union	16 (6.4)
Separated, divorced, or widowed	15 (6.0)
Symptom severity, mean (SD)	
DASS[b]—total (n=243)	29.58 (20.84)
DASS—depression (n=250)	8.99 (8.49)
DASS—anxiety (n=249)	8.35 (7.10)
DASS—stress (n=250)	11.96 (7.88)

[a]Reflects current noneducational status but does not indicate the highest level of education completed (ie, may include college graduates).

[b]DASS: Depression Anxiety Stress Scale.

Procedure

Participants completed a survey developed via the Qualtrics web-based platform as part of an experimental study assessing the impact of treatment rationale on the acceptability of iCBT. Participants were randomly assigned via Qualtrics (1:1 allocation) to read either a treatment rationale or definition of iCBT (see the study by Ellis and Anderson [37] for full details). The APOI questionnaire was administered as a primary measure of acceptability. The Depression, Anxiety, and Stress Scale-21 items (DASS-21) was used to characterize the sample, as experiences of depression and anxiety have been linked to mental health treatment–seeking attitudes [38] and to provide comparative evidence to Schröder et al [24] who recruited participants with mild to moderate depression.

All the data were collected on the web and will be made available upon request.

Measures

The APOI questionnaire [24] is a measure of attitudes toward digital mental health interventions that, for the purposes of this project, was modified to reference therapist-assisted iCBT. The development of the APOI included both exploratory and confirmatory factor analyses to identify clustering of latent constructs, resulting in 16 items comprising four subscales measuring attitudes toward psychological web-based interventions, which are as follows: (1) skepticism and perception of risk (SKE), which measures negative attitudes concerning the efficacy and security of a psychological web-based intervention; (2) confidence in effectiveness (CON), which measures positive attitudes concerning the utility and credibility of a psychological web-based intervention; (3) technologization threat (TET), which measures negative attitudes toward the lack of personal contact and the remote nature of the intervention; and (4) anonymity benefits (ABE), which measures positive attitudes related to increased privacy. Participants rate their agreement with each item (eg, “I have the feeling that iCBT can help me.”) on a 5-point Likert scale (1=totally agree

to 5=totally disagree). Positively valenced items were reverse coded. The total scores ranged from 16 to 80, with higher scores indicating more positive attitudes toward iCBT. The APOI demonstrated strong overall internal consistency (Cronbach α=.77) and showed evidence of construct validity in a sample of 1013 participants [24].

The DASS-21 [39] is a measure of mental illness comprising 3 subscales: depression, anxiety, and stress. Participants rated each item on a 4-point Likert scale (0=never to 3=always). Sum scores were computed by adding the scores across items and multiplying by 2. Scores on the total DASS-21 scale ranged from 0 to 126, with higher scores indicating more distress or impairment. Scores for each subscale were determined by summing the scores for the relevant 7 items and multiplying by 2 (range 0-42). The DASS-21 demonstrates strong convergent validity with both the Beck Anxiety Inventory (r=0.81) and Beck Depression Inventory (r=0.74), indicating a satisfactory ability to discriminate between anxiety and depressive symptoms [40]. The DASS-21 was normed on a nonclinical sample (N=717), and subsequent research has supported the validity and reliability of the DASS-21 across racial groups, including Black Americans (subscales: Cronbach α=.81−.88 [41]).

Statistical Analysis

The variables used for the factor analysis are listed in Table 3. See Tables 4 and 5 for the interitem correlation matrix and descriptive statistics.

Confirmatory factor analyses were performed using Mplus (version 8.4; Muthén & Muthén) with a sample of Black American adults (N=254) to examine the cross-cultural equivalence of the factor structure derived from the final set of 16 items indicated in the study by Schröder et al [24]. The weighted least squares means and variance adjusted (WLSMV) estimation method was used to analyze the covariance matrix structure of ordinal items. Several indices were used to evaluate the model fit: the discrepancy chi-square statistic (df≤5), standardized root mean squared residual (SRMR; SRMR≤0.08), root mean square error of approximation (RMSEA; RMSEA≤0.08), comparative fit index (CFI; CFI≥0.90), and Tucker-Lewis index (TLI; TLI≥0.90), which are commonly recommended at the indicated thresholds [42-44]. Latent variables were scaled by fixing the latent variances to 1, which allowed all indicator factor loadings to be estimated. Finally, reliability analyses of the APOI were conducted by calculating the internal consistency (Cronbach α) and corrected item-total correlations (discrimination) to facilitate comparisons with reliability metrics reported in the original publication.

In model 1, we examined a 2-factor, hierarchical confirmatory measurement model (2 first-order factors loading on 1 second-order global factor). We posited that the set of attitudes endorsed on the APOI would indicate a “positive attitudes towards internet-based treatments” latent factor as well as a “negative attitudes towards internet-based treatments” latent factor. Indicators drawn from the confidence in effectiveness (CON) and anonymity benefits (ABE) subscales comprise positive attitudes toward iCBT and were tested to examine statistically significant loading onto the “positive” latent factor. Indicators derived from the skepticism and perception of risk (SKE) and technologization threat (TET) subscales of the APOI comprise negative attitudes and were tested for statistically significant loading onto the “negative” latent factor. Both “positive” and “negative” first-order factors loaded onto the second-order global factor (termed *Acceptability* for the purposes of this study; Figure 1).

In model 2, we attempted a replication of the 4-factor, hierarchical confirmatory measurement model (4 first-order factors loading on 1 second-order global factor) proposed in the study by Schröder et al [24]. Indicators drawn from the 4 subscales were modeled per the provided confirmatory factor analysis specifications [24]. All 4 first-order factors (CON, ABE, SKE, and TET) were loaded onto the second-order global factor acceptability (Figure 2).

If neither hypothesized model 1 nor model 2 demonstrates adequate model fit, the modification fit indexes provided by the WLSMV estimation will be reviewed, and the comparative fit of a third alternative model (model 3) will be examined.

Table 3. Attitudes Towards Psychological Online Interventions Questionnaire: subscale and item descriptions[a].

Measure name and scale or item label	Description
Confidence in effectiveness subscale[b]	Measures positive attitudes concerning the efficacy and credibility of therapist-assisted iCBT[c]
CON1	A therapist-assisted iCBT program can help me to recognize the issues that I have to challenge.
CON2	I have the feeling that a therapist-assisted iCBT can help me.
CON3	A therapist-assisted iCBT program can inspire me to better approach my problems.
CON4	I believe that the concept of therapist-assisted iCBT programs makes sense.
Anonymity benefits subscale[b]	Measures positive attitudes related to the privacy and confidentiality of using a therapist-assisted iCBT
ABE1	A therapist-assisted iCBT program is more confidential and discreet than visiting a therapist.
ABE2	By using a therapist-assisted iCBT program, I can reveal my feelings more easily than with a therapist.
ABE3	I would be more likely to tell my friends that I use a therapist-assisted iCBT program than that I visit a therapist.
ABE4	By using a therapist-assisted iCBT program, I do not have to fear that someone will find out that I have psychological problems.
Skepticism and perception of risk subscale[d]	Measures negative attitudes concerning the efficacy and security of a therapist-assisted iCBT
SKE1	Using therapist-assisted iCBT programs, I do not expect long-term effectiveness.
SKE2	Using therapist-assisted iCBT programs, I do not receive professional support.
SKE3	It is difficult to implement the suggestions of a therapist-assisted iCBT effectively in everyday life.
SKE4	Therapist-assisted iCBT programs could increase isolation and loneliness.
Technologization threat subscale[d]	Measures negative attitudes related to the independent and remote nature of therapist-assisted iCBT
TET1	In crisis situations, a therapist can help me better than a therapist-assisted iCBT program.
TET2	I learn skills to better manage my everyday life from a therapist rather than from a therapist-assisted iCBT program.
TET3	I am more likely to stay motivated with a therapist than when using a therapist-assisted iCBT program.
TET4	I do not understand therapeutic concepts as well with a therapist-assisted iCBT.

[a]Response scale (1=totally disagree to 5=totally agree).

[b]Higher scores represent greater acceptability.

[c]iCBT: internet-based cognitive behavioral therapy.

[d]Higher scores indicate lower acceptability.

Table 4. Bivariate correlations between the 16 Attitudes Towards Psychological Online Interventions items.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
CON[a]1	1	—[b]	—	—	—	—	—	—	—	—	—	—	—	—	—	—
CON2	0.74	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—
CON3	0.76	0.79	1	—	—	—	—	—	—	—	—	—	—	—	—	—
CON4	0.71	0.65	0.75	1	—	—	—	—	—	—	—	—	—	—	—	—
ABE[c]1	0.38	0.46	0.47	0.41	1	—	—	—	—	—	—	—	—	—	—	—
ABE2	0.37	0.42	0.43	0.44	0.72	1	—	—	—	—	—	—	—	—	—	—
ABE3	0.20	0.34	0.26	0.25	0.53	0.56	1	—	—	—	—	—	—	—	—	—
ABE4	0.38	0.41	0.40	0.45	0.61	0.58	0.66	1	—	—	—	—	—	—	—	—
SKE[d]1	−0.05	−0.10	−0.07	0.01	−0.27	−0.31	−0.15	−0.17	1	—	—	—	—	—	—	—
SKE2	−0.01	−0.10	−0.02	0.02	−0.12	−0.30	−0.19	−0.18	0.63	1	—	—	—	—	—	—
SKE3	−0.15	−0.21	−0.15	0.03	−0.19	−0.26	−0.22	−0.15	0.71	0.72	1	—	—	—	—	—
SKE4	−0.09	−0.18	−0.07	0.04	−0.22	−0.28	−0.28	−0.25	0.63	0.69	0.75	1	—	—	—	—
TET[e]1	−0.44	−0.42	−0.50	0.58	−0.42	−0.41	−0.28	−0.33	0.24	0.21	0.24	0.22	1	—	—	—
TET2	−0.36	−0.39	−0.42	0.33	−0.43	−0.45	−0.39	−0.43	0.41	0.34	0.41	0.45	0.63	1	—	—
TET3	−0.39	−0.34	−0.41	0.36	−0.47	−0.38	−0.34	−0.41	0.38	0.25	0.30	0.38	0.66	0.72	1	—
TET4	−0.22	−0.22	−0.29	0.18	−0.45	−0.50	−0.33	−0.40	0.54	0.41	0.48	0.51	0.39	0.68	0.62	1

[a]CON: confidence in effectiveness.

[b]Not applicable.

[c]ABE: anonymity benefits.

[d]SKE: skepticism and perception of risk.

[e]TET: technologization threat.

Table 5. Descriptive statistics of the 16 Attitudes Towards Psychological Online Interventions items.

	CON[a]1	CON2	CON3	CON4	ABE[b]1	ABE2	ABE3	ABE4	SKE[c]1	SKE2	SKE3	SKE4	TET[d]1	TET2	TET3	TET4
Values, mean (SD)	3.6 (1.0)	3.4 (1.0)	3.6 (1.0)	3.7 (1.0)	3.3 (1.0)	3.2 (0.09)	3.0 (1.0)	3.2 (1.1)	3.1 (1.2)	3.3 (1.1)	3.1 (1.1)	3.2 (1.1)	2.5 (1.0)	2.7 (1.0)	2.6 (1.0)	2.9 (1.1)
Skew	−0.41	−0.15	−0.51	−0.50	−0.03	0.04	0.01	−0.08	−0.09	−0.19	−0.07	−0.13	0.26	0.03	0.18	0.11
Kurt	0.07	0.24	0.34	0.16	−0.02	0.09	−0.12	−0.14	−0.50	−0.34	−0.18	−0.33	0.30	0.16	0.07	−0.06

[a]CON: confidence in effectiveness.

[b]ABE: anonymity benefits.

[c]SKE: skepticism and perception of risk.

[d]TET: technologization threat.

Figure 1. Higher-order, 2-factor model depicting hierarchical relationship among indicators of 2 latent factors: positive and negative attitudes toward treatment loading on a global acceptability factor. ABE: anonymity benefits; CON: confidence in effectiveness; SKE: skepticism and perception of risk; TET: technologization threat. Note: threshold structure not shown.

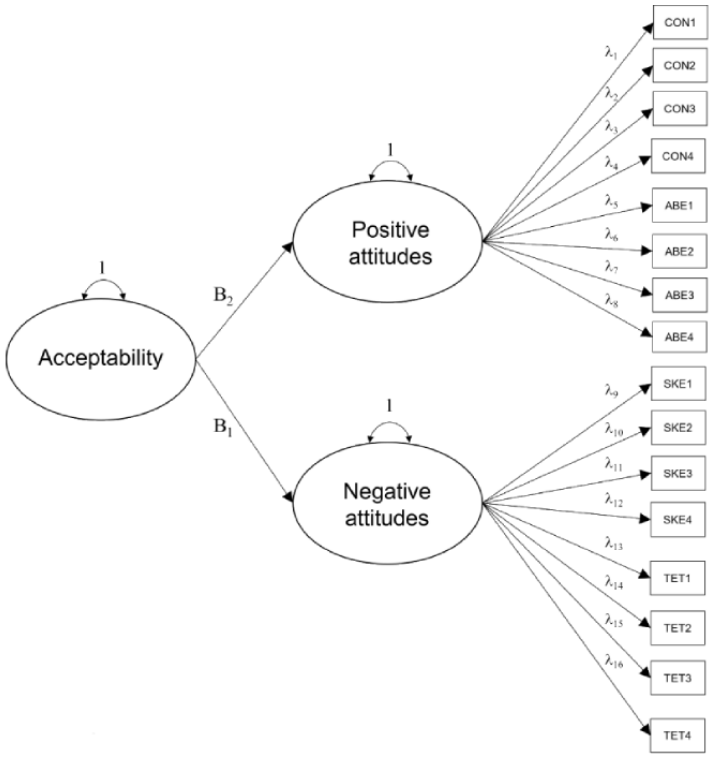
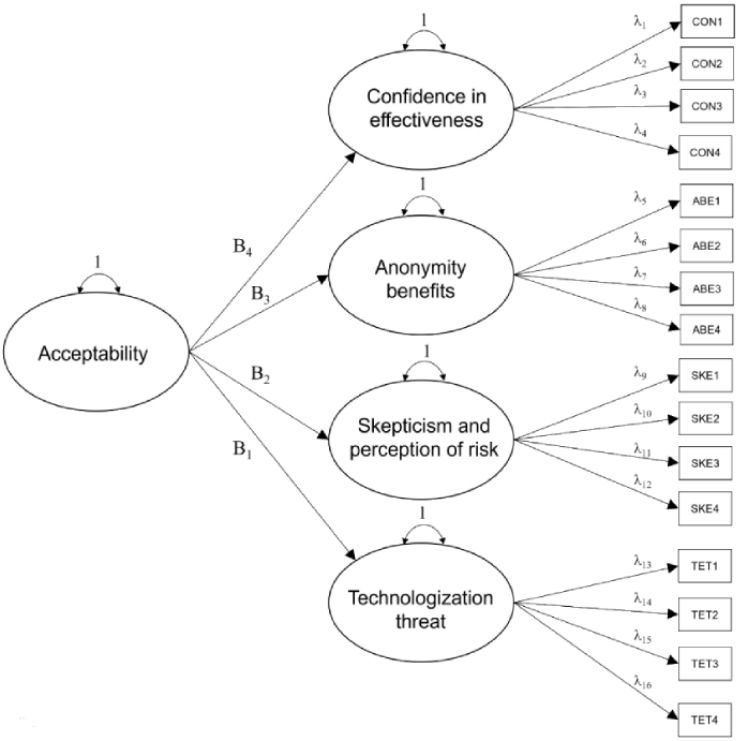


Figure 2. Higher-order, 4-factor model depicting hierarchical relationship among indicators of 4 latent factors: confidence, anonymity benefits, skepticism, and technologization threat loading on a global acceptability factor. ABE: anonymity benefits; CON: confidence in effectiveness; SKE: skepticism and perception of risk; TET: technologization threat. Note: threshold structure not shown.



Ethics Approval

This study was conducted in compliance with The Georgia State University institutional review board protocol #H18341 and preregistered with the Open Science Framework [45].

Results

Sample Characteristics

A total of 268 participants were enrolled in the study and completed the survey. Of these, 14 participants were excluded

because they did not complete the APOI questionnaire, thus yielding a sample of 254 participants. Participant ratings suggested mild symptoms of anxiety (mean 8.35, SD 7.10) and stress (mean 11.96, SD 7.88) and normal levels of depressive symptoms (mean 9.00, SD 8.49) according to standard thresholds of the DASS-21 [39].

Construct Validity

The 2 proposed models explored the construct of acceptability as a hierarchical, 2-factor model comprising "positive attitudes" and "negative attitudes" toward therapist-assisted iCBT, or as a hierarchical, 4-factor model comprising 4 distinct domains of attitudes toward therapist-assisted iCBT (confidence in effectiveness, anonymity benefits, skepticism and perception of risk, and technologization threat). See Table 6 for a full description of the model's fit indices.

Neither model had a perfect absolute model fit according to the chi-square test (model 1: χ^2_{103}=1579., P<.001; model 2: χ^2_{101}=595.3, P<.001). There was variation in the absolute values of correlation residuals, as residuals frequently exceeded 0.10 in model 1 (mean 0.14, SD 0.01), contrary to recommendations for ordered categorical variables [36]. Correlation residuals were largely below 0.10 in model 2 (mean 0.07, SD 0.01). Model 1 indicated poor fit according to CFI (0.65), TLI (0.59), SRMR (0.12), and RMSEA (0.24, 90% CI 0.23-0.25). Model 2 demonstrated better fit estimates with CFI (0.88), TLI (0.86), SRMR (0.08), and marginally improved RMSEA (0.14, 90% CI 0.13-0.15). As neither model 1 nor model 2 demonstrated adequate fit indices, an alternative bifactor model 3 (shown in Table 6) was examined because it retains theoretical similarity to the structure proposed by Schröder et al [24], and hierarchical models (ie, model 2) have more parameter constraints and are nested within less constrained bifactor models (ie, model 3) [46-48]. In model 3, the 4 factors (CON, ABE, SKE, and TET) were specified as orthogonal (instead of hierarchical) to the global factor of acceptability (Figure 3). Chi-square tests did not indicate an absolute model fit: χ^2_{82}=248.7, P<.001, although the chi-square:df ratio was 3.03, which is within the recommended range between 2 and 5 [44]. Furthermore, model 3 indicated better estimates with CFI=0.96, TLI=0.94, SRMR=0.03, and RMSEA=0.09, 90% CI 0.08-0.10. Overall, model 3 demonstrated adequate to good fit according to accepted thresholds [42-44] and the absolute values of correlation residuals did not exceed 0.10 (mean 0.03, SD 0.002). Other equivalent models were investigated (informed by statistically significant modification indices and theoretical rationale), but none demonstrated both structural fit and conceptual interpretability or parsimony (see Multimedia Appendix 1 for all tested confirmatory factor analysis models).

As models 1, 2, and 3 were nested, comparisons were conducted to verify the statistically improved model fit by examining the change in the chi-square statistic. As the scaled chi-square value for WLSMV cannot be used for traditional chi-square difference testing, the DIFFTEST option in Mplus (version 8.4) was used [49]. As shown in Table 6, comparisons indicated a significant chi-square change, $\Delta\chi^2_2$=327.7, P<.001, suggesting that model 2 was significantly better than model 1. Similarly, there was a significant chi-square change, $\Delta\chi^2_{19}$=231.9, P<.001, suggesting that model 3 was significantly better than model 2. Model 3 was the best fitting model and is described in more detail below (see Table 7 for full factor loadings and Figure 4 for the model with parameter estimates).

When examining the standardized factor loadings of the bifactor model, the absolute value of loadings for the categorical indicators ranged from 0.52 to 0.87 on their original 4 factors. Consistent with the findings of Schröder et al [24], all indicators significantly loaded onto their respective latent factors (CON, ABE, SKE, and TET), supporting the theory that these 4 domains are valid indicators of attitudes toward internet-delivered treatment. Furthermore, the 2 positively valenced latent factors (CON and ABE) significantly covaried as similar yet distinct factors (ψ=0.54; P<.001) as did the 2 negatively valenced latent factors (SKE, TET; ψ=0.70; P<.001).

The relationship between the 16 ordinal indicators and the global acceptability factor was more complex, as the absolute value of the loadings ranged from 0.004 to 0.70. Although the factor loadings for both CON and ABE indicators were positively correlated with the global acceptability factor, only CON indicators demonstrated adequate strength (0.35-0.70), whereas loadings for ABE items ranged from 0.02 to 0.28, suggesting a relatively weak relationship with the global factor. One item of the ABE subscale (ABE3) "I would be more likely to tell my friends that I use a therapist-assisted iCBT program than that I visit a therapist" did not load significantly on the global factor (λ=0.016; P=.83). Furthermore, there was significant heterogeneity in the factor loadings for both the SKE and TET indicators on the global factor. Despite its conceptualization as "negative attitudes," factor loadings of indicators of SKE ranged from 0.15 to 0.20 and were *positively* correlated with the global acceptability factor. Conversely, factor loadings of indicators of TET ranged from 0.39 to 0.64 and were negatively correlated with the global acceptability factor. One item of the TET subscale (TET4) "I do not understand therapeutic concepts as well with a therapist-assisted iCBT as I do with a live therapist" did not load significantly on the global factor (λ=0.004; P=.95).

Overall, the results from the bifactor model structure of the APOI provide evidence that the 4 factors proposed by Schröder et al [24] exhibit an orthogonal relationship with the global factor of acceptability. As expected, positively valenced factors were positively related to one another, negatively valenced factors were positively related to one another, and each item was a significant indicator of the 4 distinct subscales when controlling for the common variance shared by the global factor. The bifactor model shows that most (but not all) of the 16 APOI items are significant indicators of the global factor, although all SKE items were related in the opposite direction.

Table 6. Goodness-of-fit indexes of models tested in confirmatory factor analysis.

Model name	Chi-square (*df*)	*P* value	CFI[a]	TLI[b]	SRMR[c]	RMSEA[d] (95% CI)	Comparison		
							ΔChi-square (*df*)	*P* value	Note
2 factor	1579.8 (103)	<.001	0.65	0.59	0.12	0.24 (0.23-0.25)	__[e]	—	—
4 factor[f]	595.3 (101)	<.001	0.88	0.86	0.08	0.14 (0.13-0.15)	984.45 (2)	<.001	Versus model 1
Bifactor[f]	248.7 (82)	<.001	0.96	0.94	0.03	0.09 (0.08-0.10)	346.57 (19)	<.001	Versus model 2

[a]CFI: comparative fit index.

[b]TLI: Tucker-Lewis index.

[c]SRMR: standardized root mean squared residual.

[d]RMSEA: root mean square error of approximation.

[e]Not available.

[f]DIFFTEST command used for weighted least squares means and variance adjusted estimators to test differences in model fit.

Figure 3. Bifactor model depicting orthogonal relationship among indicators of 4 latent factors: confidence, anonymity benefits, skepticism, and technologization threat loading alongside a global acceptability factor. ABE: anonymity benefits; CON: confidence in effectiveness; SKE: skepticism and perception of risk; TET: technologization threat. Note: threshold structure not shown.

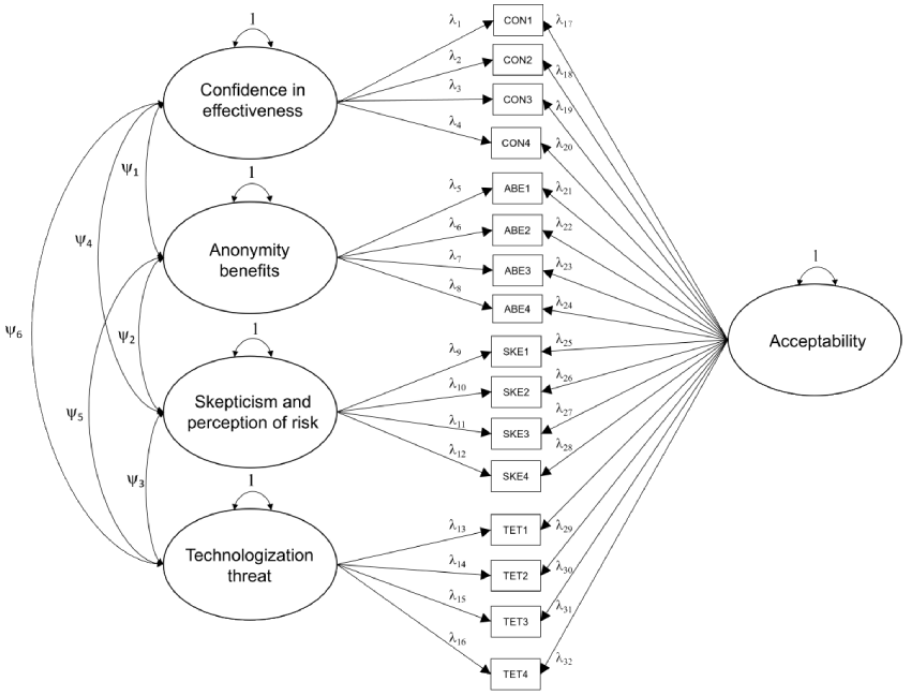


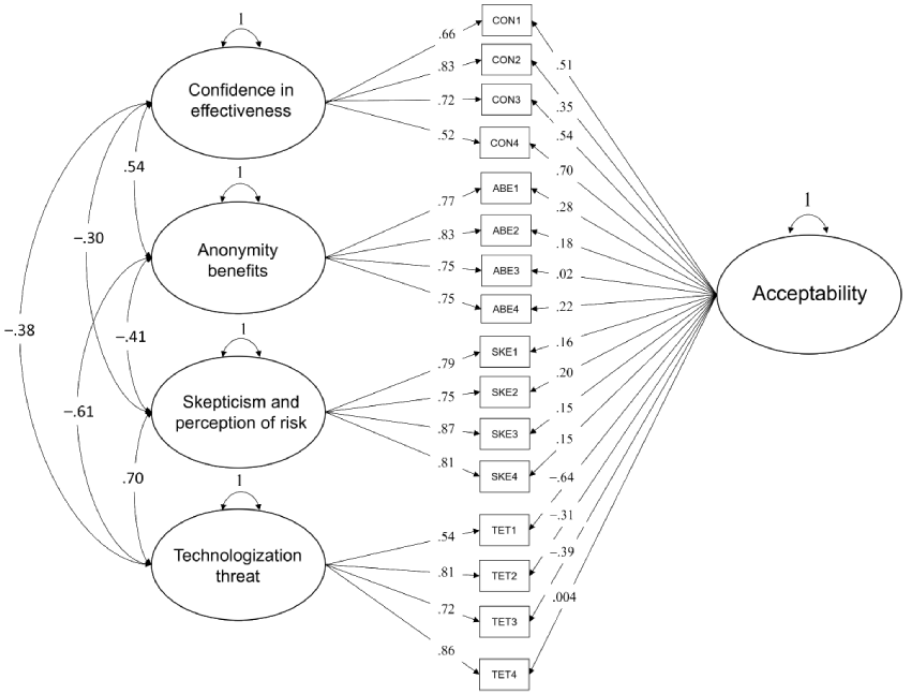
Table 7. Model 3 (bifactor) standardized factor loadings with SEs.

Relation or variable	Estimate (SE)	*P* value
Loadings		
Confidence in effectiveness (CON) BY		
CON1	0.66 (0.06)	<.001
CON2	0.83 (0.04)	<.001
CON3	0.72 (0.06)	<.001
CON4	0.52 (0.07)	<.001
Anonymity benefits (ABE) BY		
ABE1	0.77 (0.03)	<.001
ABE2	0.83 (0.03)	<.001
ABE3	0.75 (0.03)	<.001
ABE4	0.75 (0.03)	<.001
Skepticism and perception of risk (SKE) BY		
SKE1	0.79 (0.02)	<.001
SKE2	0.75 (0.03)	<.001
SKE3	0.87 (0.02)	<.001
SKE4	0.81 (0.02)	<.001
Technologization threat (TET) BY		
TET1	0.54 (0.06)	<.001
TET2	0.81 (0.03)	<.001
TET3	0.72 (0.04)	<.001
TET4	0.86 (0.03)	<.001
Acceptability BY		
CON1	0.51 (0.07)	<.001
CON2	0.35 (0.08)	<.001
CON3	0.54 (0.08)	<.001
CON4	0.70 (0.07)	<.001
ABE1	0.28 (0.07)	<.001
ABE2	0.18 (0.08)	.01
ABE3	0.02 (0.08)	.83
ABE4	0.22 (0.07)	.001
SKE1	0.16 (0.06)	.01
SKE2	0.20 (0.06)	.001
SKE3	0.15 (0.06)	.02
SKE4	0.15 (0.06)	.008
TET1	−0.64 (0.05)	<.001
TET2	−0.31 (0.07)	<.001
TET3	−0.39 (0.07)	<.001
TET4	<.01 (0.08)	.95
Factor covariances		
Confidence in effectiveness WITH		
Anonymity benefits	0.54 (0.06)	<.001
Skepticism and perception of risks	−0.30 (0.05)	<.001

Relation or variable	Estimate (SE)	*P* value
Technologization threat	−0.38 (0.06)	<.001
Acceptability	0.00 (—[a])	—
Anonymity benefits WITH		
Skepticism and perception of risks	−0.41 (0.06)	<.001
Technologization threat	−0.61 (0.05)	<.001
Acceptability	0.00 (—)	—
Skepticism and perception of risk WITH		
Technologization threat	0.70 (0.05)	<.001
Acceptability	0.00 (—)	—
Technologization threat WITH		
Acceptability	0.00 (—)	—

[a]Not available.

Figure 4. Bifactor model depicting orthogonal relationship among indicators of 4 latent factors: confidence, anonymity benefits, skepticism, and technologization threat loading alongside a global acceptability factor. Standardized parameter estimates shown. ABE: anonymity benefits; CON: confidence in effectiveness; SKE: skepticism and perception of risk; TET: technologization threat. Note: threshold structure not shown.



Reliability

The APOI demonstrated excellent internal consistency for the total scale (Cronbach α=.89) and retained good-to-excellent reliability across subscales (Cronbach α=.84 for ABE, .85 for TET, .87 for SKE, and .90 for CON). Across subscales, the corrected item-total correlations ranged from 0.59 to 0.83, with a mean adjusted correlation of 0.71 indicating good item discrimination within subscales. The corrected item - total correlations for the APOI total scale ranged from 0.45 to 0.68, with a mean adjusted correlation of 0.55, indicating good item discrimination within the total scale.

Discussion

Principal Findings

This study evaluated the psychometric properties of the APOI questionnaire [24], which is the most robust and widely used measure of *acceptability* for digital mental health interventions within a sample of Black Americans. The APOI demonstrated good-to-excellent internal consistency in the current sample, both as a total score and across subscales (Cronbach α=.84−.90), which is stronger than the internal consistency reported in the original publication (Cronbach α=.62−.77).

However, the original hierarchical, 4-factor model proposed by Schröder et al [24] exhibited relatively poor goodness-of-fit indices. Instead, the APOI showed the strongest evidence for

construct validity of a bifactor model in which each of the indicators loaded on a global factor of acceptability and the global factor of acceptability was orthogonally related to the 4 subscales. Although this unexpected finding is inconsistent with the hierarchical model proposed by Schröder et al [24], it is consistent with the literature showing that bifactor models fit better than their equivalent higher-order model in more than 90% of comparisons for mental abilities test batteries [50] and can be particularly valuable in evaluating the plausibility of subscales [51,52]. The strong, positive correlations between positively valenced subscales (confidence in effectiveness and anonymity benefits) and negatively valenced subscales (skepticism and perception of risk and technologization threat), and the negative correlations across oppositely valenced subscales are compelling evidence that the subscales have meaningful discriminant validity and can be interpreted in their own right.

The heterogeneity of findings regarding model fit may be explained by the nature of the coefficients of the factor loadings and overall structure. Modeling both positive and negatively valenced factors onto a unitary, higher-order construct (ie, acceptability) can prove difficult, especially when variance exists among indicators of lower-order constructs. The factor loadings between the 16 indicators and global acceptability factor varied substantially. Several indicators loading on the ABE, SKE, and TET subscales exhibited relatively weak or null relations with acceptability or were in the opposite direction than expected. Items loaded on the ABE subscale, in particular, may indicate both facilitators and barriers to engagement with digital interventions, given the user's conflicting perceptions of digital privacy and confidentiality [8]. Items that loaded on the SKE subscale were positively correlated with acceptability which is contrary to the conceptualization of this subscale as a construct reflecting negative attitudes, although this is interpreted with caution, given their weak correlations.

Scholars have called for better conceptualizations of acceptability [15,23], which have the potential to produce even more parsimonious measures by exploring new factors or consolidating indicators to reduce conceptual overlap. In particular, there is a growing need for evidence of the dimensions of acceptability that are demonstrably correlated with uptake, engagement, and adherence to digital mental health interventions. As discussed in prior research, this apparent discrepancy in consumer attitudes and behaviors may, in fact, be a consequence of the heterogeneous nature and definition of acceptability toward digital mental health interventions [22,24]. A considerable amount of research uses a single item to assess acceptability and results from this study, and others [29,30,32], demonstrate that single-items measures are inadequate for the operationalization of this heterogeneous construct.

Furthermore, these data suggest that within a Black American population, there is greater utility in interpreting the APOI subscales as attitudinal constructs distinct from a global acceptability factor. However, given that the higher-order model is nested within the bifactor model [46-48], these models are not necessarily at odds with one another. Ultimately, these results provide support for the underlying validity of the 4 factors proposed by the APOI but eschew traditional practices of prioritizing the calculation of a single acceptability score at the expense of adequately measuring each relevant dimension of acceptability and reporting them in tandem with the global score for contextualization.

Strengths and Limitations

This is the first study to investigate the psychometric properties of the APOI questionnaire among a racially minoritized population. This study is the first to provide evidence for the cross-cultural equivalence of APOI among Black Americans. This is a notable contribution to the literature, as the vast majority of randomized controlled trials examining the efficacy and acceptability of iCBT do not include (or do not report) racial minorities in their studies [13], and existing measures of consumer attitudes toward digital mental health interventions [24,29-33] have predominantly been developed and examined for validation within White majority (and predominantly European) samples. Furthermore, by modifying the target treatment from "psychological online interventions" to "therapist-assisted iCBT," this study provides preliminary evidence for the utility of the APOI for diverse digital interventions with varying degrees of specificity. Overall, the results suggest that the APOI is a robust measure.

Despite the strengths of this study, there are some limitations that warrant attention. The study sample consisted of participants with minimal symptoms of depression, anxiety, or stress. This was distinct from the participants who reported moderate levels of depression in the study by Schröder et al [24]. Future research needs to evaluate these measures among those with greater depression severity or other diagnoses. The participants in this study were predominantly young adult females. These demographic groups are more likely to use digital mental health interventions, and the relative impact of their positive and negative attitudes towards digital mental health intervention is likely to differ across diverse populations [8]. Relatedly, measurement invariance was not formally assessed across different subgroups within the sample (eg, male vs female), because of significant imbalances in sample size, which minimized the power to detect potential differences between these groups. Finally, the convergent validity of the APOI with other measures of acceptability within a Black American sample could not be determined because no other relevant measures of acceptability existed at the time of data collection for this study.

Future Directions

Future research should modify the APOI to apply it to other digital mental health interventions (eg, virtual reality exposure therapies and massively open web-based interventions) and translate the measure into additional languages (eg, Spanish) to further examine cross-intervention and cross-cultural equivalency. Although the APOI demonstrated good internal consistency reliability within the present sample, test-retest reliability was not examined. Indeed, with the exception of the study by Clough et al [29], there is a notable lack of investigation of the test-retest reliability of acceptability measures, which deserves further evaluation. Moreover, it would be compelling to investigate the criterion validity of the APOI to examine whether positive attitudes toward digital mental health interventions predict the willingness to use or actual use

of digital mental health interventions among racially and ethnically minoritized participants. Consistent with the Theory of Planned Behavior [53], which emphasizes the relationship among beliefs, attitudes, and behavioral intentions, positive attitudes toward acceptability would be expected to be the strongest predictor of behavioral intention, which in turn is the immediate determinant of actual treatment-seeking behavior. Investigations of the relationship between attitudes toward iCBT and the effectiveness of such interventions should be conducted, as those with more positive attitudes might derive greater clinical benefits. Finally, although studies examining the convergent validity of the APOI with related measures of acceptability toward digital mental health interventions have been recently conducted [29,30], these studies did not expressly recruit participants from racially and ethnically minoritized communities, and their results are predominantly based on White or European samples. This is concerning, as racially and ethnically minoritized communities may be positioned to benefit the most from the treatment accessibility advantages afforded by digital mental health interventions [54]. Understanding these communities' attitudes toward these treatments is paramount.

Conclusions

The APOI questionnaire is a valid and reliable measure of attitudes toward therapist-assisted iCBT among Black Americans. However, some of the indicators were only weakly associated with the global factor of acceptability, and a bifactor model demonstrated better goodness-of-fit than the hierarchical, 4-factor structure proposed by the original authors. This provides strong evidence that the APOI demonstrates multidimensionality and that there is greater utility in interpreting APOI subscales as attitudinal constructs distinct from a global acceptability factor. Indeed, attitudes of acceptability comprise both positive and negative attitudes toward the uptake of digital mental health interventions and must be evaluated in tandem to effectively understand the nuanced attitudes consumers may hold toward these interventions. This is the first study to examine the psychometric properties of any measure of consumer attitudes toward digital mental health interventions among Black participants. Demonstrating the reliability, validity, and cultural equivalency of existing measures of attitudes toward these interventions is needed to improve our understanding of the drivers of and barriers to using digital treatments among minoritized communities. For the full potential of digital mental health interventions to improve equitable access to treatment to be realized, more adequate representation of minoritized communities in research on these interventions must be achieved.

Acknowledgments

The authors would like to thank Lee Branum-Martin, PhD, for consultation on structural equation modeling and confirmatory factor analyses. The data in this study are a secondary analysis by Ellis and Anderson [37]. This paper presents original secondary analyses of a previously published experimental survey study.

Authors' Contributions

DME devised the project, main conceptual ideas, and protocol outline and conducted all the statistical analyses; designed the figures and tables; and wrote the manuscript. Both DME and PLA contributed to the final version of this manuscript. PLA supervised the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Materials depict fit indices for all examined confirmatory factor analyses. Mplus (version 8.4) syntax is provided for all analyses.
[DOCX File , 32 KB - mental_v10i1e43929_app1.docx]

References

1. Alegría M, Canino G, Ríos R, Vera M, Calderón J, Rusch D, et al. Inequalities in use of specialty mental health services among Latinos, African Americans, and non-Latino whites. Psychiatr Serv 2002 Dec;53(12):1547-1555. [doi: 10.1176/appi.ps.53.12.1547] [Medline: 12461214]
2. Ayalon L, Alvidrez J. The experience of Black consumers in the mental health system--identifying barriers to and facilitators of mental health treatment using the consumers' perspective. Issues Ment Health Nurs 2007 Dec 09;28(12):1323-1340. [doi: 10.1080/01612840701651454] [Medline: 18058337]
3. Gaston GB, Earl TR, Nisanci A, Glomb B. Perception of mental health services among Black Americans. Social Work in Mental Health 2016 Feb 16;14(6):676-695. [doi: 10.1080/15332985.2015.1137257]
4. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]

5. Gerhards SA, de Graaf LE, Jacobs LE, Severens JL, Huibers MJ, Arntz A, et al. Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2010 Apr;196(4):310-318. [doi: 10.1192/bjp.bp.109.065748] [Medline: 20357309]
6. Hedman E, Andersson E, Ljótsson B, Andersson G, Rück C, Lindefors N. Cost-effectiveness of internet-based cognitive behavior therapy vs. cognitive behavioral group therapy for social anxiety disorder: results from a randomized controlled trial. Behav Res Ther 2011 Nov;49(11):729-736. [doi: 10.1016/j.brat.2011.07.009] [Medline: 21851929]
7. Carolan S, de Visser RO. Employees' perspectives on the facilitators and barriers to engaging with digital mental health interventions in the workplace: qualitative study. JMIR Ment Health 2018 Jan 19;5(1):e8 [FREE Full text] [doi: 10.2196/mental.9146] [Medline: 29351900]
8. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
9. Himle JA, Weaver A, Zhang A, Xiang X. Digital mental health interventions for depression. Cognit Behavioral Pract 2022 Feb;29(1):50-59. [doi: 10.1016/j.cbpra.2020.12.009]
10. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul;12(7):861-9; quiz 870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
11. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct 09;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
12. Bernstein EE, Weingarden H, Wolfe EC, Hall MD, Snorrason I, Wilhelm S. Human support in app-based cognitive behavioral therapies for emotional disorders: scoping review. J Med Internet Res 2022 Apr 08;24(4):e33307 [FREE Full text] [doi: 10.2196/33307] [Medline: 35394434]
13. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
14. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Human Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
15. Apolinário-Hagen J, Kemper J, Stürmer C. Public acceptability of e-mental health treatment services for psychological problems: a scoping review. JMIR Ment Health 2017 Apr 03;4(2):e10 [FREE Full text] [doi: 10.2196/mental.6186] [Medline: 28373153]
16. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
17. Casey LM, Joy A, Clough BA. The impact of information on attitudes toward e-mental health services. Cyberpsychol Behav Soc Netw 2013 Aug;16(8):593-598. [doi: 10.1089/cyber.2012.0515] [Medline: 23679567]
18. Mitchell N, Gordon PK. Attitudes towards computerized CBT for depression amongst a student population. Behav Cognit Psychother 2007 May 14;35(4):421-430. [doi: 10.1017/s1352465807003700]
19. Travers MF, Benton SA. The acceptability of therapist-assisted, internet-delivered treatment for college students. J College Student Psychother 2014 Jan 14;28(1):35-46. [doi: 10.1080/87568225.2014.854676]
20. Mohr DC, Siddique J, Ho J, Duffecy J, Jin L, Fokuo JK. Interest in behavioral and psychological treatments delivered face-to-face, by telephone, and by internet. Ann Behav Med 2010 Aug;40(1):89-98 [FREE Full text] [doi: 10.1007/s12160-010-9203-7] [Medline: 20652466]
21. Choi I, Sharpe L, Li S, Hunt C. Acceptability of psychological treatment to Chinese- and Caucasian-Australians: internet treatment reduces barriers but face-to-face care is preferred. Soc Psychiatry Psychiatr Epidemiol 2015 Jan;50(1):77-87. [doi: 10.1007/s00127-014-0921-1] [Medline: 24993290]
22. Molloy A, Ellis DM, Su L, Anderson PL. Improving acceptability and uptake behavior for internet-based cognitive-behavioral therapy. Front Digit Health 2021;3:653686 [FREE Full text] [doi: 10.3389/fdgth.2021.653686] [Medline: 34713125]
23. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
24. Schröder J, Sautier L, Kriston L, Berger T, Meyer B, Späth C, et al. Development of a questionnaire measuring attitudes towards Psychological Online Interventions-the APOI. J Affect Disord 2015 Nov 15;187:136-141. [doi: 10.1016/j.jad.2015.08.044] [Medline: 26331687]
25. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4] [Medline: 11132119]
26. Handley T, Perkins D, Kay-Lambkin F, Lewin T, Kelly B. Familiarity with and intentions to use internet-delivered mental health treatments among older rural adults. Aging Ment Health 2015;19(11):989-996. [doi: 10.1080/13607863.2014.981744] [Medline: 25420968]

27. Wootton BM, Titov N, Dear BF, Spence J, Kemp A. The acceptability of internet-based treatment and characteristics of an adult sample with obsessive compulsive disorder: an internet survey. PLoS One 2011;6(6):e20548 [FREE Full text] [doi: 10.1371/journal.pone.0020548] [Medline: 21673987]
28. Apolinário-Hagen J, Vehreschild V, Alkoudmani RM. Current views and perspectives on e-mental health: an exploratory survey study for understanding public attitudes toward internet-based psychotherapy in Germany. JMIR Ment Health 2017 Feb 23;4(1):e8 [FREE Full text] [doi: 10.2196/mental.6375] [Medline: 28232298]
29. Clough B, Eigeland J, Madden I, Rowland D, Casey L. Development of the eTAP: a brief measure of attitudes and process in e-interventions for mental health. Internet Interv 2019 Dec;18:100256 [FREE Full text] [doi: 10.1016/j.invent.2019.100256] [Medline: 31890610]
30. Gómez Penedo JM, Berger T, Grosse Holtforth M, Krieger T, Schröder J, Hohagen F, et al. The Working Alliance Inventory for guided Internet interventions (WAI-I). J Clin Psychol 2020 Jun;76(6):973-986. [doi: 10.1002/jclp.22823] [Medline: 31240727]
31. Miloff A, Carlbring P, Hamilton W, Andersson G, Reuterskiöld L, Lindner P. Measuring alliance toward embodied virtual therapists in the era of automated treatments with the Virtual Therapist Alliance Scale (VTAS): development and psychometric evaluation. J Med Internet Res 2020 Mar 24;22(3):e16660 [FREE Full text] [doi: 10.2196/16660] [Medline: 32207690]
32. Teles S, Ferreira A, Paúl C. Assessing attitudes towards online psychoeducational interventions: psychometric properties of a Brief Attitudes Scale. Health Soc Care Community 2021 Sep;29(5):e1-10. [doi: 10.1111/hsc.13227] [Medline: 33170537]
33. Miragall M, Baños RM, Cebolla A, Botella C. Working alliance inventory applied to virtual and augmented reality (WAI-VAR): psychometrics and therapeutic outcomes. Front Psychol 2015 Oct 08;6:1531 [FREE Full text] [doi: 10.3389/fpsyg.2015.01531] [Medline: 26500589]
34. Horvath AO, Bedi RP. The alliance. In: Psychotherapy Relationships That Work: Therapist Contributions and Responsiveness to Patients. Oxford, United Kingdom: Oxford University Press; 2002.
35. McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. Psychol Method 2002 Mar;7(1):64-82 [FREE Full text] [doi: 10.1037/1082-989x.7.1.64] [Medline: 11928891]
36. Kline R. Principles and Practice of Structural Equation Modeling. New York City, NY: Guilford Publications; 2015.
37. Ellis DM, Anderson PL. Improving the acceptability of internet-based cognitive-behavioral therapy among Black Americans. Technol Mind Behav 2021 Oct 18;2(3). [doi: 10.1037/tmb0000044]
38. Magaard JL, Seeralan T, Schulz H, Brütt AL. Factors associated with help-seeking behaviour among individuals with major depression: a systematic review. PLoS One 2017;12(5):e0176730 [FREE Full text] [doi: 10.1371/journal.pone.0176730] [Medline: 28493904]
39. Lovibond S, Lovibond P, Psychology Foundation of Australia. Manual for the Depression Anxiety Stress Scales. Sydney, N.S.W: Psychology Foundation of Australia; 1995.
40. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
41. Norton PJ. Depression Anxiety and Stress Scales (DASS-21): psychometric analysis across four racial groups. Anxiety Stress Coping 2007 Sep;20(3):253-265. [doi: 10.1080/10615800701309279] [Medline: 17999228]
42. Fan X, Thompson B, Wang L. Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. Structural Equat Model Multidisciplinary J 1999 Jan;6(1):56-83. [doi: 10.1080/10705519909540119]
43. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equat Model Multidisciplinary J 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
44. Marsh HW, Hocevar D. Application of confirmatory factor analysis to the study of self-concept: first- and higher order factor models and their invariance across groups. Psychol Bull 1985 May;97(3):562-582. [doi: 10.1037/0033-2909.97.3.562]
45. Ellis D, Anderson P. Cross-cultural validation of the attitudes towards psychological online interventions questionnaire among Black Americans. OSF Registries. 2022. URL: https://osf.io/y3r2p [accessed 2022-02-16]
46. Markon KE. Bifactor and hierarchical models: specification, inference, and interpretation. Annu Rev Clin Psychol 2019 May 07;15(1):51-69. [doi: 10.1146/annurev-clinpsy-050718-095522] [Medline: 30649927]
47. Rijmen F. Formal relations and an empirical comparison among the bi-factor, the Testlet, and a second-order multidimensional IRT model. J Educ Measurement 2010;47(3):361-372. [doi: 10.1111/j.1745-3984.2010.00118.x]
48. Yung Y, Thissen D, McLeod LD. On the relationship between the higher-order factor model and the hierarchical factor model. Psychometrika 1999 Jun;64(2):113-128. [doi: 10.1007/bf02294531]
49. Satorra A, Bentler PM. Ensuring positiveness of the scaled difference chi-square test statistic. Psychometrika 2010 Jun 20;75(2):243-248 [FREE Full text] [doi: 10.1007/s11336-009-9135-y] [Medline: 20640194]
50. Cucina J, Byle K. The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. J Intell 2017 Jul 11;5(3):27 [FREE Full text] [doi: 10.3390/jintelligence5030027] [Medline: 31162418]
51. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. J Pers Assess 2010 Nov;92(6):544-559 [FREE Full text] [doi: 10.1080/00223891.2010.496477] [Medline: 20954056]

52. Reise S, Bonifay W, Haviland M. Bifactor modelling and the evaluation of scale scores. In: The Wiley Handbook of Psychometric Testing, 2 Volume Set A Multidisciplinary Reference on Survey, Scale and Test Development · Volume 1. Hoboken, New Jersey: Wiley; 2018.
53. Ajzen I. The theory of planned behavior. Organizational Behav Human Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
54. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of digital mental health for marginalized and underserved populations. Curr Treat Options Psych 2019 Jul 5;6(3):243-255. [doi: 10.1007/s40501-019-00181-z]

Abbreviations

APOI: Attitudes Towards Psychological Online Interventions
CFI: comparative fit index
DASS-21: Depression Anxiety Stress Scale-21 items
iCBT: internet-based cognitive behavioral therapy
RMSEA: root mean square error of approximation
SRMR: standardized root mean squared residual
TLI: Tucker-Lewis index
WLSMV: weighted least squares means and variance adjusted

Edited by J Torous; submitted 01.11.22; peer-reviewed by S Pardini, P Chow; comments to author 23.02.23; revised version received 05.03.23; accepted 06.03.23; published 27.04.23.

Please cite as:
Ellis DM, Anderson PL
Validation of the Attitudes Towards Psychological Online Interventions Questionnaire Among Black Americans: Cross-cultural Confirmatory Factor Analysis
JMIR Ment Health 2023;10:e43929
URL: https://mental.jmir.org/2023/1/e43929
doi:10.2196/43929
PMID:37103983

Original Paper

The Impact of a Messenger-Based Psychosocial Chat Counseling Service on Further Help-Seeking Among Children and Young Adults: Longitudinal Study

Sabrina Baldofski[1*], PhD; Elisabeth Kohls[1,2*], PhD; Zeki Efe[1], MSc; Melanie Eckert[3], MA; Shadi Saee[3], MSc; Julia Thomas[3], MA; Richard Wundrack[3], MA; Christine Rummel-Kluge[1,2], MD

[1]Department of Psychiatry and Psychotherapy, Medical Faculty, Leipzig University, Leipzig, Germany

[2]Department of Psychiatry and Psychotherapy, University Leipzig Medical Center, Leipzig, Germany

[3]krisenchat gGmbH, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Christine Rummel-Kluge, MD
Department of Psychiatry and Psychotherapy
Medical Faculty
Leipzig University
Semmelweisstr 10
Haus 13
Leipzig, 04103
Germany
Phone: 49 3419724464
Email: christine.rummel-kluge@medizin.uni-leipzig.de

Abstract

Background: Mental crises have high prevalences in adolescence. Early interventions appear to be highly important to diminish the risk of the deterioration, recurrence, or chronification of symptoms. In recent years, various providers have started offering live chat support in psychological crises. The messenger-based psychological counseling service krisenchat aims to support young people in crises and, if necessary, provide a recommendation for a referral to the health care system or to seek further help from a trusted adult person.

Objective: This study aimed to investigate the impact of using the counseling service of krisenchat on the further help-seeking behavior of young people, and to identify associated factors of further help-seeking.

Methods: This longitudinal study analyzed anonymous data from 247 individuals who used krisenchat between October 2021 and March 2022, and received a recommendation for further help-seeking. An online survey directly after the chat assessed the perceived helpfulness of the chat and well-being after the chat. After 4 weeks, further help-seeking, facilitators and barriers to help-seeking, and self-efficacy were assessed in an online follow-up survey.

Results: The most frequently recommended services or persons to seek further help from included a psychotherapist or social psychiatric service (75/225, 33.3%), a school psychologist or school social worker (52/225, 23.1%), and the user's parents (45/225, 20.0%). Of the 247 users, 120 (48.6%) indicated that they contacted the recommended service or person, and of these, 87 (72.5%) stated that they already had an appointment (or talk) with the respective service or person or that an appointment (or talk) was scheduled. The most frequently reported facilitators for further help-seeking were mental health literacy (54/120, 45.0%), improvement of self-efficacy (55/120, 45.8%), and symptom recognition (40/120, 33.3%). In users not displaying further help-seeking behavior, the most frequent barriers included stigmatization (60/127, 47.2%), lack of mental health literacy (59/127, 46.5%), need for self-reliance and autonomy (53/127, 41.7%), and negative family beliefs regarding help services (53/127, 41.7%). Subgroup comparisons indicated significantly higher levels of self-efficacy in users displaying further help-seeking behavior than in those not displaying further help-seeking behavior. Both subgroups did not differ in gender, age, recommended service or person, chat topics, perceived helpfulness, and well-being.

Conclusions: The findings of this study indicate that children and young adults receiving counseling on krisenchat benefit in terms of seeking further help. Further help-seeking seems to be associated with higher levels of self-efficacy.

Trial Registration: Deutsches Register Klinischer Studien DRKS00026671; https://tinyurl.com/4fm5xe68

(JMIR Ment Health 2023;10:e43780) doi:10.2196/43780

KEYWORDS

online intervention; e-mental health; online chat; hotline; text-based; children; adolescents; young adults; psychopathology; help-seeking

Introduction

Several mental health conditions develop or have their onset at a young age, and their prevalence increases dramatically during adolescence [1,2]. The estimated worldwide prevalence of mental health problems of 10% to 20% among adolescents shows that this age group is affected more than any other age group [3,4]. Mental health problems thus represent an important health concern for young people [4,5]. Untreated symptoms are likely to persist and result in a compromised mental health status, for example, untreated affective disorders like depression are directly associated with suicidal behavior [6]. Besides the fact that the mortality rate of suicide increases with progressing teenage years, suicide is one of the leading causes of death among young people aged 10 to 24 years [6-8]. Therefore, early intervention to diminish the risk of the deterioration, recurrence, or chronification of symptoms appears to be especially important [9,10]. Nevertheless, help-seeking behavior and treatment use are low among the majority of adolescents and young adults with mental conditions. Thus, oftentimes, adolescents do not receive professional treatment [11-13].

The help-seeking process can be described as a stage model involving (1) an awareness of symptoms and an appraisal that support or treatment is needed, (2) the expression of symptoms and the need for help, (3) an awareness of available and accessible sources and services of help, and (4) the willingness to disclose concerns and problems to the selected source [14]. Besides the mere help-seeking process, it might be necessary to address young people's willingness and readiness to reach out to a help service that is most suited to their needs [15]. This seems especially important as studies have shown that young people are overall more hesitant to seek help from mental health services [16,17]. Additionally, the more severe the symptoms, the less likely young people are to seek help [16,17]. For example, only about 1 in 5 German students with diagnosed anxiety or mood disorders had ever used mental health services in the past [18].

The low help-seeking behavior could potentially be explained by barriers to accessing mental health services. Systematic reviews have identified key barriers to help-seeking for mental health problems in young people, including the fear of stigma, family reactions or negative family beliefs toward mental health services, a lack of mental health literacy, the need for self-reliance and autonomy, concerns about confidentiality, and other barriers such as structural factors like access, time, transport, or cost [13,19].

Considering these barriers, mental health literacy might be a key factor in initiating help-seeking behavior. Mental health literacy refers to the ability to recognize, manage, and prevent mental health problems and includes the awareness of available sources of help [19,20]. Therefore, mental health literacy is an important step in adaptive coping and getting access to appropriate health care for mental health concerns [16,21]. Further, studies show that mental health services often provide information about mental health, leading to an increased mental health literacy [22]. In turn, people with an increased mental health literacy show a lower self-stigmatization regarding mental health problems, are more capable of recognizing their mental health status, and are more likely to seek help from appropriate and professional help services [23-26]. Moreover, the literature suggests that interventions focusing on mental health literacy significantly increase the readiness, motivation, and intention to seek help [27].

Young people prefer to solve their problems on their own, thus expressing a high need for self-reliance and autonomy [16,28]. The common use of digital media indicates that young people use the internet to find their own way to solve problems, and studies also show an increase in the self-reliant search for information or help to cope with mental health problems [14,29,30]. Regarding the above-mentioned barriers to help-seeking, internet-based services have many advantages, such as no geographical boundaries, usually free access, and anonymity and privacy [31], and could thus present an opportunity for low-threshold access to encourage further help-seeking behavior. Despite the increasing number of these services, there is a lack of systematic evaluations in the current literature.

One of the services offered is *krisenchat*, a German messenger-based psychological counseling service for children, adolescents, and young adults. The first cross-sectional evaluation showed a high acceptance, feasibility, and user satisfaction of this service [32]. The results indicated a high need for a 24/7 accessible, anonymous, and low-threshold online help service to offer rapid stress relief in acute crises such as suicidality [33]. Besides its goal to increase the mental health literacy of users, *krisenchat* aims to refer users in need to professional help services or personal contacts [32].

In most of the previous studies evaluating online mental health services, further help-seeking was not considered the primary outcome (eg, [24,34,35]). Thus, it seems highly relevant to evaluate the impact of online services on further help-seeking behavior in adolescents with mental health concerns [36].

This study aimed to investigate the impact of counseling received from *krisenchat* on the further help-seeking behavior of young people, and to identify associated factors of further help-seeking, using anonymous data of chat users, including an anonymous online follow-up survey. Specifically, this study aimed to (1) determine the percentage of users of *krisenchat*

who followed the recommendation for a referral to the health care system or to seek further help from a trusted adult person within 4 weeks; (2) investigate the facilitators and barriers to further help-seeking in *krisenchat* users who did and did not implement the recommendation, respectively; and (3) identify the factors associated with further help-seeking.

Methods

Participants and Procedure

krisenchat [37] is a German-speaking messenger-based psychosocial crisis chat counseling service for children, adolescents, and young adults under the age of 25 years. Counselors are volunteers with an education in psychology, psychotherapy, education, or social work and have received a minimum of 2 months of training in chat-based counseling and in screening for a potential clinical indication [32]. The aim of *krisenchat* is to listen to, calm, and comfort users during acute crises. Additionally, if deemed necessary, users also receive a recommendation to seek further help, either within the health care system (eg, professional local support services) or from a trusted adult person.

Usually, a chat counseling session consists of the following 5 phases: (1) building a trusting relationship, (2) understanding the problem, (3) clarifying the goal for the counseling session, (4) finding a common solution, and (5) properly ending the counseling session. For example, a female user aged 20 years and currently living with her violent parents may start out by mentioning an argument with her father (phase 1), but open up about being physically harmed after some empathic and supportive messages (phase 2). The counselor and user then agree that they want to do something about the situation involving outside help (phase 3). Since the user is not a minor anymore and not previously known to the youth office, this is not a case of child welfare endangerment. After asking the user what help she would like to receive and where she currently lives, the counselor informs her about her rights and options, and researches a local point of contact (in this case, an association called “Frauen Gegen Gewalt e.V.” [“Women Against Violence Association”]) (phase 4). Finally, the counselor and user agree on how to get in touch with the association and consensually end the chat after making sure that no open questions remain (phase 5).

Within the *krisenchat* counseling process, the first step toward a recommendation to a health service provider or a trusted adult person includes an assessment of age, gender, living conditions, crisis topic, (symptom) duration and severity, social resources, and existing affiliations to and previous experiences with help service providers. Usually, *krisenchat* users in need of further help are recommended to involve a trusted adult person from their private circle of acquaintances as a first low-threshold option for minor problems and conflicts. In cases where it seems necessary to eventually involve authorities (eg, youth welfare service), users are first recommended to involve a professional trusted adult person like a school psychologist. Other specialized professional counseling services with expertise in specific topics (eg, eating disorders, LGBTQ+, and drug abuse) are recommended when professional help seems necessary. Further, some users are recommended to seek medical help (eg, general practitioner and emergency ward) or psychotherapy treatment. When there are clear indicators of child welfare endangerment, contacting the youth welfare service is recommended.

For the purpose of this study, anonymous data from all chat users between October 1, 2021, and March 28, 2022, were extracted from the operational database. Data contained information collected by the counselors. In addition to this information, 2 online surveys were conducted: a feedback survey shortly after a chat session and a follow-up survey after 4 weeks.

For the feedback survey, each user received an automatically generated invitation via WhatsApp or SMS text messaging, which included a link to an online survey 6 hours after the first counseling session, if the chat session had at least 20 messages and the user was not considered at risk for child welfare endangerment by the psychological team. A minimum of 20 messages was established as an inclusion criterion based on the experiences of *krisenchat* with (1) the counseling process: after exchanging 20 messages counselors have usually gained enough information like age, gender, and the (primary) chat topic, which are necessary for a meaningful scientific evaluation of the chat; and (2) the user experience: after exchanging 20 messages the user has usually gained enough information to justify a feedback survey regarding their experience with the service.

For the follow-up survey, users received an automatically generated invitation via WhatsApp or SMS text messaging after 4 weeks, which included a link to an online survey, if the following criteria applied: the user had completed the feedback survey and had given their consent to be contacted again for the follow-up survey; a recommendation for a referral to the health care system or to seek further help from a trusted adult person had been given by the counselor and the counseling process was completed; and the user was not considered at risk for child welfare endangerment. Being considered at risk for child welfare endangerment was again an exclusion criterion in this step of the data collection process as these cases are often not labeled as cases of child welfare endangerment during the first counseling session, but only later in the counseling process. Children and young adults in these kinds of situations often contact *krisenchat* first with a minor problem to test if they can trust the service before opening up in a later counseling session about their real concern. If users attended more than one session, they received the link to the follow-up survey after the first recommendation for further help-seeking was made.

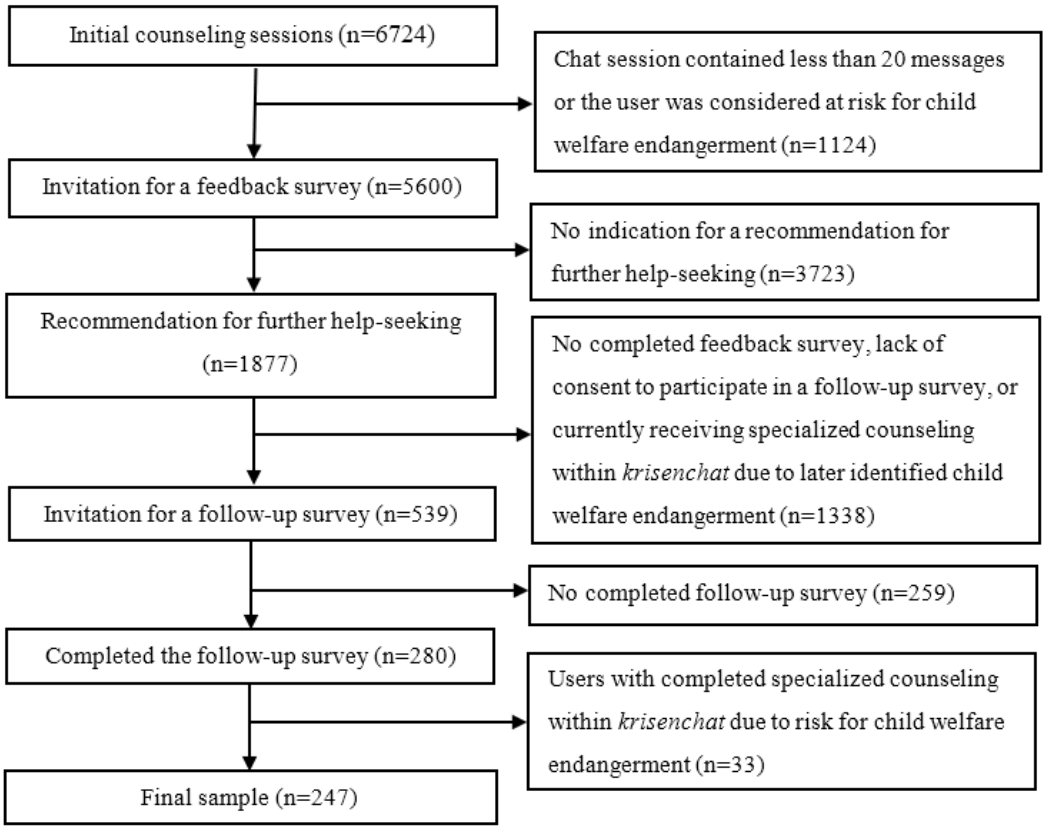
According to the definition used in a previous evaluation of *krisenchat* [32], a chat session was defined as a series of messages with no gaps between messages of more than 12 hours. It is possible that *krisenchat* users chat with more than one counselor during the course of one session. Users may or may not be made aware of the change in counselors, depending on the specific case.

Both surveys were set up in German using *typeform* [38]. At the beginning of each survey, it was offered to conduct the survey in simplified language in case users had limited German literacy.

Between October 1, 2021, and March 28, 2022, *krisenchat* received 10,614 requests in total. Of these, 3890 (36.6%) were excluded for one of the following reasons: users did not consent to the terms and conditions of *krisenchat*, they left the chat, or they did not receive a full counseling session due to a lack of capacity because of high demand. This left 6724 (63.4%) initial counseling sessions, of which 1124 (10.6%) were excluded as the chat session contained less than 20 messages (eg, when a user's issue could be resolved within less than 20 messages) or the user was considered at risk for child welfare endangerment. The remaining 5600 (52.8%) users received an invitation for the feedback survey (see Figure 1 for the data collection process).

Of the 5600 users who received an invitation for the feedback survey, 539 (9.6%) later received an invitation for the follow-up survey. The other 5061 (90.4%) users were excluded from the follow-up survey as they met the exclusion criteria for the follow-up survey. In total, 280 (5.0%) users completed the follow-up survey, and of these, 33 (0.6%) had to be excluded due to technical circumstances, as they had completed specialized counseling within *krisenchat* due to being considered at risk for child welfare endangerment and had received the follow-up survey, although they met an exclusion criterion (ie, being considered at risk for child welfare endangerment). This resulted in a final sample of 247 users with completed feedback and follow-up surveys.

Figure 1. Flowchart of the data collection process.



Ethics Approval

Informed consent was provided online via an opt-in function before participating in each survey. Ethics approval was granted by the Ethics Committee of the Medical Faculty, University of Leipzig, on August 3, 2021 (file reference: 372/21-ek).

Measures

Sociodemographic Information and Chat Use

Information regarding age and gender was collected by counselors if users disclosed them during a session. Further, counselors collected information on the topics of the counseling session. It was possible for a user to report several concerns at the same time, and counselors identified more than one chat topic for these users. For this analysis, the reported topics were aggregated into 4 categories: psychiatric symptoms, psychosocial distress, emotional distress, and violence (Table 1).

Table 1. Chat categories and topics.

Category	Included topics
Psychiatric symptoms	Anxiety, panic attacks, depressive symptoms or depression, obsessive-compulsive behavior, trauma (eg, flashbacks), depersonalization or derealization, substance abuse, behavioral addiction, symptoms related to borderline personality disorder, suicidality, nonsuicidal self-injury, psychotic symptoms, eating disorder symptoms, sleeping disorders, and psychosomatic complaints
Psychosocial distress	Conflicts in the family, at school, at work, in a relationship, or with friends; pressure to perform or high performing expectations (at school or at work, from oneself, or from others); difficulties adjusting to a new situation or environment; fear regarding the future; bullying; stress and overburden; online relationships; dealing with a relative's or friend's accident or physical illness; and dealing with a relative's or friend's suicide, nonsuicidal self-injury, mental disorder, or substance abuse
Emotional distress	Loneliness or isolation, lovesickness, feelings of guilt, low self-esteem, grief, difficulties in becoming independent from parents, emotional neglect, and anger or aggression
Violence	Being an offender or a victim of a criminal act, violence within the family or a relationship, and being a victim of physical or psychological violence

Feedback Survey

In the feedback survey, user well-being directly after the chat was assessed using 1 item ("How did you feel after the chat?"), answered on a 4-point Likert scale (1="better," 2="rather better," 3="rather worse," and 4="worse"). A second item assessed the perceived helpfulness of the chat ("Was *krisenchat* able to help you with your concerns?") on a 4-point Likert scale (1="definitely," 2="rather yes," 3="rather no," and 4="not at all").

Follow-up Survey

Further Help-Seeking, Facilitators, and Barriers

Users were asked to indicate which recommendation for a referral to the health care system or to seek further help from a trusted adult person they had received, including the following options: specialized professional counseling service, school psychologist or school social worker, social services (family assistance, educational assistance, or youth welfare service), local youth center, psychotherapist or social psychiatric service, general practitioner, calling an ambulance (to the hospital), parents, or another trusted adult person. If a user had received more than one recommendation, they were asked to indicate the option most relevant to them.

The actual further help-seeking behavior was assessed with the dichotomous item "After the counseling with *krisenchat*, did you contact the person or professional help service you were referred to?" Users who confirmed this item, indicating having followed the recommendation they received, were considered as displaying further help-seeking behavior. They were further asked if an appointment with the recommended service (or a talk with the recommended person) had already taken place, and if this was the case, it was assessed whether further appointments (or talks) were scheduled. The latter items were designed to measure the specific further help-seeking behavior based on the individual recommendation a user had received during the counseling process.

If users gave a positive answer ("yes") on the item assessing actual further help-seeking behavior, 10 items on potential facilitators for help-seeking behavior were assessed in the following part. Likewise, in the case of a negative answer ("no") on the item assessing actual further help-seeking behavior, 15 items on potential barriers for help-seeking behavior were assessed. Multiple choices were possible on each of the latter item sets. The respective items on facilitators and barriers to help-seeking were constructed based on a recent systematic review on help-seeking facilitators and barriers [19], and the practical experience of the counselors. Facilitators were assessed using 10 items grouped into 5 categories as follows: mental health literacy (3 items), symptom recognition (3 items), social support (1 item), improvement of self-efficacy through counseling (2 items), and an "other" category (1 item to indicate if reasons not covered by the other items applied). Barriers were assessed using 15 items grouped into the following 5 categories: stigmatization (2 items), negative family beliefs regarding help services (2 items), lack of mental health literacy (3 items), self-reliance and autonomy (3 items), structural factors (2 items), and an "other" category (3 items).

Self-efficacy

The General Self-efficacy Short Scale (German: "Allgemeine Selbstwirksamkeit Kurzskala [ASKU]" [39]) was used to assess the general efficacy beliefs of users. Self-efficacy relies on the definition by Albert Bandura, which describes people's beliefs about their capability to execute behaviors that influence events affecting their lives [40]. The scale consists of 3 items representing single statements (eg, "I am able to solve most problems on my own") with a 5-point Likert scale (1="does not apply at all" to 5="fully applies"), with higher mean scores indicating higher levels of self-efficacy. In this study, the scale showed an acceptable reliability with Cronbach α=.74. In other studies, the scale had a high reliability of ω=0.81 to 0.86 [39]. The mean reference values were 3.8 for low, 4.0 for medium, and 4.3 for high levels of self-efficacy [39].

Statistical Analysis

Descriptive statistics were used for sociodemographic variables, recommendations (for a referral to the health care system or to seek further help from a trusted adult person), further help-seeking behavior, and facilitators and barriers to help-seeking. A chi-square test was used to analyze further help-seeking behavior depending on the recommended service or person. To this end, the number of categories of recommended services or persons was reduced by grouping

them into the following 6 categories: psychotherapist or social psychiatric service, school psychologist or school social worker, parents, trusted adult person, counseling service, and other.

To analyze the associated factors of further help-seeking, the 2 subgroups of users displaying further help-seeking behavior and those not displaying further help-seeking behavior were compared regarding categorical variables (gender and session topic) using chi-square tests. For continuous variables (age, well-being after the chat, perceived helpfulness of the chat, and self-efficacy), both subgroups were compared using Mann-Whitney *U* tests owing to nonnormally distributed data, as indicated by Shapiro-Wilks tests ($P<.05$). Additionally, a logistic regression analysis was performed to analyze the association between various predictor variables (gender, age, session topic, well-being after the chat, perceived helpfulness of the chat, and self-efficacy) and further help-seeking behavior.

Bonferroni correction was used to account for multiple testing. To estimate effect sizes for chi-square tests, the φ coefficient was used, while Cramér *V* (φ_c) was used when the contingency table was larger than 2×2, with φ, $\varphi_c=0.10$ indicating a small effect, φ, $\varphi_c=0.30$ indicating a medium effect, and φ, $\varphi_c=0.50$ indicating a large effect [41]. Effect sizes for Mann-Whitney *U* tests were interpreted as small ($r<0.30$), medium ($r<0.50$), and large ($r>0.50$) [41]. Statistical analyses were performed using IBM SPSS Statistics version 27 (IBM Corp). A 2-tailed α value of .05 was applied to statistical testing.

Results

Sociodemographic Characteristics

In the final sample of 247 users, most users were female (female users: 199/242, 82.2%; male users: 11/242, 4.5%; diverse users: 32/242, 13.2%; the data for 5 users were missing), and the mean user age was approximately 17 years (mean 17.31, SD 3.29; range 12-25 years). Nearly half of all users in each survey (feedback survey: 124/247, 50.2%; follow-up survey: 108/247, 43.7%) chose to conduct the survey in simplified language. The main reasons for contacting *krisenchat* were psychiatric symptoms (184/247, 74.5%), and psychosocial (136/247, 55.1%) or emotional distress (71/247, 28.7%). The mean self-efficacy score of *krisenchat* users was 2.82 (SD 0.86).

Help-Seeking Behavior

The most frequently recommended services or persons to seek further help from included a psychotherapist or social psychiatric service (75/225, 33.3%), a school psychologist or school social worker (52/225, 23.1%), and the user's parents (45/225, 20.0%). Nearly half of all users (120/247, 48.6%) indicated that they contacted the recommended service or person. A chi-square test indicated no significant differences in help-seeking behavior depending on the recommended service or person ($\chi^2_{5,225}=7.87$; $P=.16$; $\varphi_c=0.19$).

Of the 120 users who indicated having contacted the recommended service or person, 102 (85.0%) gave further information. Specifically, a large proportion stated that they already had an appointment (or talk) with the respective service or person (70/102, 68.6%) or that an appointment or talk was scheduled (17/102, 16.7%), while only 12 (11.8%) reported that there were no appointments available, 2 (2.0%) reported that they had contacted but not yet reached the recommended service or person, and 1 (1.0%) indicated being referred to another service. Of the 87 users who already had an appointment (or talk with a trusted person) or had scheduled an appointment (or talk), 75 (86.2%) provided further information on possible follow-up appointments. Specifically, the majority stated that a follow-up appointment (or talk) was already scheduled (48/75, 64.0%) or was planned but not yet scheduled (12/75, 16.0%), while 5 (6.7%) reported that they had already received sufficient help and another appointment or talk was not necessary and 1 (1.3%) indicated being referred to another service after the first appointment. The remaining 9 (12.0%) users reported that no follow-up appointments were available or that they did not want to schedule another appointment as they did not feel comfortable or had not received the help they needed.

Facilitators and Barriers to Help-Seeking

A detailed description of the reported facilitators and barriers to further help-seeking is provided in Table 2. Users displaying further help-seeking behavior reported facilitators to help-seeking in the categories of mental health literacy (54/120, 45.0%), improvement of self-efficacy (55/120, 45.8%), symptom recognition (40/120, 33.3%), and social support (20/120, 16.7%). Users not displaying further help-seeking behavior reported barriers to help-seeking in the categories of stigmatization (60/127, 47.2%), lack of mental health literacy (59/127, 46.5%), self-reliance and autonomy (53/127, 41.7%), negative family beliefs regarding help services (53/127, 41.7%), structural factors (13/127, 10.2%), and other factors (eg, someone else has offered help; 24/127, 18.9%).

Table 2. Facilitators and barriers to help-seeking behavior (N=247).

Variable	Users displaying further help-seeking behavior (n=120), n (%)[a]	Users not displaying further help-seeking behavior (n=127), n (%)[a]
Facilitators		
Mental health literacy[b]	54 (45.0)	N/A[c]
Got new information	31 (25.8)	N/A
Knowledge on where to seek further help	24 (20.0)	N/A
Got a specific plan on how to seek further help	15 (12.5)	N/A
Self-efficacy[b]	55 (45.8)	N/A
Feeling encouraged to seek further help	53 (44.2)	N/A
Feeling able to change the situation	10 (8.3)	N/A
Symptom recognition[b]	40 (33.3)	N/A
Seeing the problem more clearly	16 (13.3)	N/A
Understood that what happened was not right	15 (12.5)	N/A
Knowing how to talk about my problem/concern	17 (14.2)	N/A
Social support[b]	20 (16.7)	N/A
Feeling that someone is supporting me	20 (16.7)	N/A
Other	0 (0.0%)	N/A
Barriers		
Stigmatization[b]	N/A	60 (47.2)
Fearing the reactions of others	N/A	52 (40.9)
Not daring to talk with a stranger about serious concerns	N/A	23 (18.1)
Lack of mental health literacy[b]	N/A	59 (46.5)
Concerns will not be taken seriously	N/A	24 (18.9)
Not deserving help, because problems are one’s own fault	N/A	31 (24.4)
Fear of being admitted to a psychiatric ward	N/A	32 (25.2)
Self-reliance and autonomy[b]	N/A	53 (41.7)
Sought help elsewhere	N/A	15 (11.8)
Can handle it alone	N/A	27 (21.3)
Recommendation does not fit	N/A	14 (11.0)
Family beliefs[b]	N/A	53 (41.7)
Fear that parents will find out	N/A	42 (33.1)
Professional help services are not trustworthy	N/A	20 (15.7)
Structural factors[b]	N/A	13 (10.2)
Referred help service is too far away or difficult to reach	N/A	2 (1.6)
Have not had time yet	N/A	12 (19.4)
Other[b]	N/A	24 (18.9)
Someone else has offered help	N/A	6 (4.7)
No help needed after all, already feeling better	N/A	8 (6.3)
Other reasons, not specified	N/A	10 (7.9)

[a]Percentage is calculated from valid cases.

[b]Multiple answers were possible.

[c]N/A: not applicable.

Associated Factors of Help-Seeking

Subgroup comparisons indicated significantly higher levels of self-efficacy (ASKU) in users displaying further help-seeking behavior than in those not displaying further help-seeking behavior (*P*=.01, small effect; Table 3). Moreover, users with further help-seeking behavior reported a higher perceived helpfulness of the chat than users without further help-seeking behavior (*P*=.02, small effect); however, this effect was not significant after Bonferroni correction. No significant group differences were found for gender, age, chat topics, and well-being after the chat (all *P*>.05). An overall logistic regression model for further help-seeking behavior was not statistically significant (χ^2_{10}=15.04; *P*=.13; Nagelkerkes R^2=0.09). Only perceived helpfulness of the chat was significantly associated with further help-seeking behavior (B=0.56; *P*=.03; odds ratio 1.75, 95% CI 1.06-2.88). All other predictors in the model were not significant (all *P*>.05).

Table 3. Group comparison between users with and those without further help-seeking behavior (N=247).

Variable	Users displaying further help-seeking behavior (n=120)	Users not displaying further help-seeking behavior (n=127)	Chi-square (*df*)	*U*	*P* value	Effect size
Gender, n (%)[a]			2.94 (2,242)	N/A[b]	.23	φ_c[c]=0.11
Female	94 (79.0)	105 (85.4)				
Male	8 (6.7)	3 (2.4)				
Diverse	17 (14.3)	15 (12.2)				
Age (years), mean (SD)	17.68 (3.62)	16.97 (2.91)	N/A	6759.50	.22	*r*=0.08
Topic: Psychiatric symptoms, n (%)[a]	86 (71.7)	98 (77.2)	0.98 (1,247)	N/A	.32	φ=0.06
Topic: Psychosocial distress, n (%)[a]	62 (51.7)	74 (58.3)	1.09 (1,247)	N/A	.30	φ=0.07
Topic: Emotional distress, n (%)[a]	32 (26.7)	39 (30.7)	0.49 (1,247)	N/A	.48	φ=0.05
Topic: Violence, n (%)[a]	4 (3.3)	3 (2.4)	0.21 (1,247)	N/A	.65	φ=0.03
Well-being score, mean (SD)	2.39 (0.65)	2.25 (0.65)	N/A	6029.50	.10	*r*=0.11
Perceived helpfulness score, mean (SD)	2.41 (0.68)	2.22 (0.68)	N/A	6303.50	.02	*r*=0.15
Self-efficacy score (ASKU[d]), mean (SD)	2.97 (0.82)	2.69 (0.88)	N/A	5797.00	.01	*r*=0.17

[a]Percentage is calculated from valid cases.

[b]N/A: not applicable.

[c]φ_c: Cramér *V*.

[d]ASKU: Allgemeine Selbstwirksamkeit Kurzskala (“General Self-efficacy Short Scale”).

Discussion

Summary

Overall, the results imply that children and young adults using *krisenchat* experience benefits. The 24/7 online counseling service may facilitate seeking further professional help within the health care system or support from others.

It is important to highlight that the referral of users to professional help services or personal contacts is only one of the aims of *krisenchat* (only in 27.9% of all *krisenchat* users), along with listening, calming, and comforting in an acute crisis; engaging in collaborative problem-solving; emphasizing and supporting chat users to find their own solutions; and improving mental health literacy [32]. Therefore, the current results only refer to a subgroup of all chat users (for more details on the overall use of *krisenchat* see [32]).

Help-Seeking Behavior

The results of this study show that children and young adults using *krisenchat* benefit in terms of seeking further help. Nearly half of all users (120/247, 48.6%) indicated that they contacted the recommended service or person after the professional counseling. This percentage appears to be very high and shows *krisenchat* to be an important gatekeeper service for young people and adolescents. Nevertheless, owing to novelty, there are at present limited studies on the evaluation of online chat counseling or crisis text line interventions. The few available studies are not comparable to our study owing to different foci or differing measurement approaches (eg, evaluation of users’ perceptions of effectiveness right after the chat or focus on suicide preventive aspects only [42-44]).

Nevertheless, more than half of the users who took part in the follow-up survey reported that they did not seek further help. There were no differences in help-seeking behavior depending on the recommended service or person. Other potential reasons for not displaying further help-seeking behavior remain unclear, and further research is needed in this area.

As mentioned earlier, little is known about the effects or benefits of online counseling or helplines, since studies are methodologically challenging not least because of the anonymity of the services. In a recent systematic review on the state of youth helplines, which is at least partly comparable to the *krisenchat* service, the authors concluded that helplines may provide a beneficial service to youth and that psychosocial concerns are the main reasons for contacting these services [45]. This result is in line with previous evaluations of *krisenchat* [32,33]. The authors also concluded from the results of the review that there is a lack of literature owing to a lack of controlled trials on the one hand, and complex methodological/ethical barriers preventing such trials on the other [45]. Few studies have investigated long-term outcomes of crisis hotlines in the past, mainly focusing on suicide rates within a population [46,47].

Facilitators and Barriers to Help-Seeking

Nearly half of the users who sought further help reported mental health literacy to be an important facilitator, and reported the improvement of self-efficacy and symptom recognition as an effect of using the chat counseling service. In users who reported no further help-seeking behavior, the most frequent barriers were stigmatization, lack of mental health literacy, the need for self-reliance and autonomy, and negative family beliefs regarding help services. In line with the literature, the results also indicate that improving mental health literacy seems to be a key strategy for improving help-seeking in adolescents and young adults [19,48-50]. There are different approaches for improving mental health literacy in young people, including whole-of-community campaigns; community campaigns aimed at a youth audience; school-based interventions teaching help-seeking skills, mental health literacy, or resilience; and programs training individuals to better intervene in a mental health crisis [51,52].

Overall, it is important to highlight again that previous studies revealed that *krisenchat* may lower the barriers to access help, that is, about 50% of those contacting *krisenchat* reported that they never had contact with the professional health care system before [32].

Associated Factors of Help-Seeking

The analysis on group differences revealed that users displaying further help-seeking behavior reported higher levels of self-efficacy compared with those not displaying further help-seeking behavior, but both subgroups did not differ in gender, age, recommended service or person, chat topics, perceived helpfulness of the chat, and well-being after the chat. Generally, the mean self-efficacy score of all *krisenchat* users was 2.82 (SD 0.86), which is considered to be very low [39]. In the logistic regression analysis, only a higher perceived helpfulness of the chat was associated with further help-seeking behavior; however, the interpretability of this result is limited as the overall model was not significant.

Surprisingly, in this study, central aspects of the *krisenchat* user (age, gender, and well-being) and service usage (chat topic and recommended service or person) seemed to be irrelevant for predicting further help-seeking. This means, in turn, that based on recent data, the only factor relevant for further help-seeking is self-efficacy. In line with the recent literature, this result supports the claim for more research regarding help-seeking in youth and adolescents, in order to be able to target interventions to improve and facilitate help-seeking behavior [19].

Strengths and Limitations

This study is the first to examine the impact of using *krisenchat* on the further help-seeking behavior of young people, and to identify associated factors of further help-seeking. However, a number of limitations must be considered. The final sample used for this analysis consisted of only 13.2% of all users of the service, who received a recommendation for a referral to the health care system or to seek further help from a trusted adult person, with completed feedback and follow-up surveys. There might be further associated factors of help-seeking or different patterns in the help-seeking behavior of users, which could not be detected due to this attrition rate. Additionally, there might be key differences between those who filled in the survey and those who did not (eg, regarding motivation, perceived helpfulness of the chat, chat topics, and other aspects). Further, several other variables, such as cultural background, fluency in the German language, education level, and socioeconomic status, which could not be assessed due to the anonymity of *krisenchat*, might influence help-seeking behavior and factors associated with further help-seeking. A further limitation includes the use of self-report data for assessing the recommendation for further help-seeking.

Nevertheless, given the fact that there are limited studies on the impact or benefit of 24/7 online counseling, the sample size of 247 is reasonable and can serve as a basis for stimulating further research and advances in measurements.

Conclusion

The results of this study suggest that children and young adults using *krisenchat* may benefit in terms of seeking further help. Further help-seeking seems to be associated with higher levels of self-efficacy. Further research is necessary to better understand how young people can further benefit from 24/7 online counseling. Longitudinal research seems especially crucial to understand who benefits when and from what kind of service and to assess which factors are associated with help-seeking behavior.

Acknowledgments

We acknowledge support from Leipzig University for open access publishing. This work was supported by The Federal Ministry of Health (ZMI1-2521FEP001). We thank Lukas Guenthner, MSc, for support in data preparation.

Authors' Contributions

SB, EK, and CRK designed the study. SB and ZE performed the statistical analysis. SB, EK, and ZE drafted the article. ME, SS, JT, and RW set up the data collection, prepared the data set, and edited previous versions of the manuscript. SB, EK, ZE, ME, SS, JT, RW, and CRK discussed the results and contributed to the final manuscript. All authors have approved the final manuscript.

Conflicts of Interest

SB, EK, and ZE confirm no conflicts of interest. ME, SS, JT, and RW are paid employees of krisenchat gGmbH. CRK received lecture honoraria from Recordati and Servier outside and independent of the submitted work.

References

1. Lewinsohn PM, Clarke GN, Seeley JR, Rohde P. Major depression in community adolescents: age at onset, episode duration, and time to recurrence. J Am Acad Child Adolesc Psychiatry 1994 Jul;33(6):809-818. [doi: 10.1097/00004583-199407000-00006] [Medline: 7598758]
2. Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. Dialogues in Clinical Neuroscience 2022 Apr 01;11(1):7-20. [doi: 10.31887/dcns.2009.11.1/krmerikangas]
3. Gore FM, Bloem PJN, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: a systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]
4. Mental health of adolescents. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2023-05-08]
5. Slade T, Johnston A, Oakley Browne MA, Andrews G, Whiteford H. 2007 National Survey of Mental Health and Wellbeing: methods and key findings. Aust N Z J Psychiatry 2009 Jul;43(7):594-605. [doi: 10.1080/00048670902970882] [Medline: 19530016]
6. David-Ferdon C, Crosby AE, Caine ED, Hindman J, Reed J, Iskander J. CDC Grand Rounds: Preventing Suicide Through a Comprehensive Public Health Approach. MMWR Morb Mortal Wkly Rep 2016 Sep 02;65(34):894-897 [FREE Full text] [doi: 10.15585/mmwr.mm6534a2] [Medline: 27584004]
7. Suicide. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/suicide [accessed 2023-05-08]
8. Soole R, Kõlves K, De Leo D. Suicide in Children: A Systematic Review. Arch Suicide Res 2015;19(3):285-304. [doi: 10.1080/13811118.2014.996694] [Medline: 25517290]
9. Fryers T, Brugha T. Childhood determinants of adult psychiatric disorder. Clin Pract Epidemiol Ment Health 2013;9:1-50 [FREE Full text] [doi: 10.2174/1745017901309010001] [Medline: 23539489]
10. Pine DS, Fox NA. Childhood antecedents and risk for adult mental disorders. Annu Rev Psychol 2015 Jan 03;66:459-485. [doi: 10.1146/annurev-psych-010814-015038] [Medline: 25559116]
11. Chavira DA, Garland A, Yeh M, McCabe K, Hough RL. Child anxiety disorders in public systems of care: comorbidity and service utilization. J Behav Health Serv Res 2009 Oct;36(4):492-504 [FREE Full text] [doi: 10.1007/s11414-008-9139-x] [Medline: 18688725]
12. Eisenberg D, Hunt J, Speer N, Zivin K. Mental health service utilization among college students in the United States. J Nerv Ment Dis 2011 May;199(5):301-308. [doi: 10.1097/NMD.0b013e3182175123] [Medline: 21543948]
13. Radez J, Reardon T, Creswell C, Lawrence PJ, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021 Feb;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
14. Rickwood DJ, Mazzer KR, Telford NR. Social influences on seeking help from mental health services, in-person and online, during adolescence and young adulthood. BMC Psychiatry 2015 Mar 07;15:40 [FREE Full text] [doi: 10.1186/s12888-015-0429-6] [Medline: 25886609]
15. Sanci L, Lewis D, Patton G. Detecting emotional disorder in young people in primary care. Curr Opin Psychiatry 2010 Jul;23(4):318-323. [doi: 10.1097/yco.0b013e32833aac38] [Medline: 20540178]
16. Rickwood D, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. Australian e-Journal for the Advancement of Mental Health 2014 Dec 17;4(3):218-251. [doi: 10.5172/jamh.4.3.218]
17. Sawyer MG, Sawyer ACP, La Greca AM. Help-seeking by young people with depressive disorders. Prevention Researcher 2012;19(4):11-14.
18. Essau CA. Frequency and patterns of mental health services utilization among adolescents with anxiety and depressive disorders. Depress Anxiety 2005;22(3):130-137. [doi: 10.1002/da.20115] [Medline: 16175563]

19. Aguirre Velasco A, Cruz ISS, Billings J, Jimenez M, Rowe S. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. BMC Psychiatry 2020 Jun 11;20(1):293 [FREE Full text] [doi: 10.1186/s12888-020-02659-0] [Medline: 32527236]
20. Kutcher S, Wei Y, Coniglio C. Mental Health Literacy: Past, Present, and Future. Can J Psychiatry 2016 Mar;61(3):154-158 [FREE Full text] [doi: 10.1177/0706743715616609] [Medline: 27254090]
21. Bowers H, Manion I, Papadopoulos D, Gauvreau E. Stigma in school-based mental health: perceptions of young people and service providers. Child Adolesc Ment Health 2013 Sep;18(3):165-170. [doi: 10.1111/j.1475-3588.2012.00673.x] [Medline: 32847251]
22. Jorm AF, Griffiths KM. Population promotion of informal self-help strategies for early intervention against depression and anxiety. Psychol Med 2006 Jan;36(1):3-6. [doi: 10.1017/S0033291705005659] [Medline: 16356291]
23. Collin PJ, Metcalf AT, Stephens-Reicher JC, Blanchard ME, Herrman HE, Rahilly K, et al. ReachOut.com: The role of an online service for promoting help-seeking in young people. Advances in Mental Health 2014 Dec 17;10(1):39-51. [doi: 10.5172/jamh.2011.10.1.39]
24. Horgan A, Sweeney J. Young students' use of the Internet for mental health information and support. J Psychiatr Ment Health Nurs 2010 Mar;17(2):117-123. [doi: 10.1111/j.1365-2850.2009.01497.x] [Medline: 20465756]
25. Nicholas J. The role of internet technology and social branding in improving the mental health and wellbeing of young people. Perspect Public Health 2010 Mar;130(2):86-90. [doi: 10.1177/1757913908101797] [Medline: 20455488]
26. Wright A, Jorm AF, Harris MG, McGorry PD. What's in a name? Is accurate recognition and labelling of mental disorders by young people associated with better help-seeking and treatment preferences? Soc Psychiatry Psychiatr Epidemiol 2007 Mar;42(3):244-250. [doi: 10.1007/s00127-006-0156-x] [Medline: 17450404]
27. Gulliver A, Griffiths KM, Christensen H. Barriers and facilitators to mental health help-seeking for young elite athletes: a qualitative study. BMC Psychiatry 2012 Sep 26;12:157 [FREE Full text] [doi: 10.1186/1471-244X-12-157] [Medline: 23009161]
28. Wilson CJ, Bushnell JA, Caputi P. Early access and help seeking: practice implications and new initiatives. Early Interv Psychiatry 2011 Feb;5 Suppl 1:34-39. [doi: 10.1111/j.1751-7893.2010.00238.x] [Medline: 21208389]
29. Fukkink RG, Hermanns JMA. Children's experiences with chat support and telephone support. J Child Psychol Psychiatry 2009 Jun;50(6):759-766. [doi: 10.1111/j.1469-7610.2008.02024.x] [Medline: 19207634]
30. Gowen LK. Online Mental Health Information Seeking in Young Adults with Mental Health Challenges. Journal of Technology in Human Services 2013 Apr;31(2):97-111. [doi: 10.1080/15228835.2013.765533]
31. Stephens-Reicher J, Metcalf A, Blanchard M, Mangan C, Burns J. Reaching the hard-to-reach: how information communication technologies can reach young people at greater risk of mental health difficulties. Australas Psychiatry 2011 Jul;19 Suppl 1:S58-S61. [doi: 10.3109/10398562.2011.583077] [Medline: 21878021]
32. Eckert M, Efe Z, Guenthner L, Baldofski S, Kuehne K, Wundrack R, et al. Acceptability and feasibility of a messenger-based psychological chat counselling service for children and young adults ("krisenchat"): A cross-sectional study. Internet Interv 2022 Mar;27:100508 [FREE Full text] [doi: 10.1016/j.invent.2022.100508] [Medline: 35242589]
33. Kohls E, Guenthner L, Baldofski S, Eckert M, Efe Z, Kuehne K, et al. Suicidal Ideation Among Children and Young Adults in a 24/7 Messenger-Based Psychological Chat Counseling Service. Front Psychiatry 2022;13:862298 [FREE Full text] [doi: 10.3389/fpsyt.2022.862298] [Medline: 35418889]
34. Harris K, McLean J, Sheffield J. Solving suicidal problems online: Who turns to the Internet for help? Australian e-Journal for the Advancement of Mental Health 2014 Dec 17;8(1):28-36. [doi: 10.5172/jamh.8.1.28]
35. Neal DM, Campbell AJ, Williams LY, Liu Y, Nussbaumer D. “I did not realize so many options are available”: Cognitive authority, emerging adults, and e-mental health. Library & Information Science Research 2011 Jan;33(1):25-33. [doi: 10.1016/j.lisr.2010.07.015]
36. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014 Mar 04;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
37. krisenchat. URL: https://krisenchat.de/en [accessed 2023-05-11]
38. Typeform. URL: https://www.typeform.com/ [accessed 2023-05-11]
39. Beierlein C, Kemper CJ, Kovaleva A, Rammstedt B. Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen – Allgemeine Selbstwirksamkeit Kurzskala (ASKU). In: GESIS Working Papers. Köln, Germany: GESIS; 2012.
40. Bandura A. Self Efficacy: The Exercise of Control. New York, NY: W. H. Freeman; 1997.
41. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
42. Gould MS, Pisani A, Gallo C, Ertefaie A, Harrington D, Kelberman C, et al. Crisis text-line interventions: Evaluation of texters' perceptions of effectiveness. Suicide Life Threat Behav 2022 Jun;52(3):583-595 [FREE Full text] [doi: 10.1111/sltb.12873] [Medline: 35599358]
43. Pisani AR, Gould MS, Gallo C, Ertefaie A, Kelberman C, Harrington D, et al. Individuals who text crisis text line: Key characteristics and opportunities for suicide prevention. Suicide Life Threat Behav 2022 Jun;52(3):567-582 [FREE Full text] [doi: 10.1111/sltb.12872] [Medline: 35615898]

44. Szlyk HS, Roth KB, García-Perdomo V. Engagement With Crisis Text Line Among Subgroups of Users Who Reported Suicidality. Psychiatr Serv 2020 Apr 01;71(4):319-327 [FREE Full text] [doi: 10.1176/appi.ps.201900149] [Medline: 31795857]
45. Mathieu SL, Uddin R, Brady M, Batchelor S, Ross V, Spence SH, et al. Systematic Review: The State of Research Into Youth Helplines. J Am Acad Child Adolesc Psychiatry 2021 Oct;60(10):1190-1233. [doi: 10.1016/j.jaac.2020.12.028] [Medline: 33383161]
46. Mishara BL, Chagnon F, Daigle M, Balan B, Raymond S, Marcoux I, et al. Comparing models of helper behavior to actual practice in telephone crisis intervention: a Silent Monitoring Study of Calls to the U.S. 1-800-SUICIDE Network. Suicide Life Threat Behav 2007 Jun;37(3):291-307. [doi: 10.1521/suli.2007.37.3.291] [Medline: 17579542]
47. Coveney CM, Pollock K, Armstrong S, Moore J. Callers' experiences of contacting a national suicide prevention helpline: report of an online survey. Crisis 2012;33(6):313-324 [FREE Full text] [doi: 10.1027/0227-5910/a000151] [Medline: 22759662]
48. Pretorius C, Chambers D, Coyle D. Young People's Online Help-Seeking and Mental Health Difficulties: Systematic Narrative Review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
49. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
50. Tomczyk S, Muehlan H, Freitag S, Stolzenburg S, Schomerus G, Schmidt S. Is knowledge "half the battle"? The role of depression literacy in help-seeking among a non-clinical sample of adults with currently untreated mental health problems. J Affect Disord 2018 Oct 01;238:289-296. [doi: 10.1016/j.jad.2018.05.059] [Medline: 29902732]
51. Kelly CM, Jorm AF, Wright A. Improving mental health literacy as a strategy to facilitate early intervention for mental disorders. Med J Aust 2007 Oct 01;187(S7):S26-S30. [doi: 10.5694/j.1326-5377.2007.tb01332.x] [Medline: 17908021]
52. Evans-Lacko S, Hahn JS, Peter L, Schomerus G. The impact of digital interventions on help-seeking behaviour for mental health problems: a systematic literature review. Curr Opin Psychiatry 2022 May 01;35(3):207-218. [doi: 10.1097/YCO.0000000000000788] [Medline: 35579875]

Abbreviations

ASKU: Allgemeine Selbstwirksamkeit Kurzskala (“General Self-efficacy Short Scale”)

Edited by J Torous; submitted 25.10.22; peer-reviewed by AM Duncan, JC Perez, M Bulgheroni; comments to author 03.12.22; revised version received 21.12.22; accepted 22.03.23; published 17.05.23.

Please cite as:
Baldofski S, Kohls E, Efe Z, Eckert M, Saee S, Thomas J, Wundrack R, Rummel-Kluge C
The Impact of a Messenger-Based Psychosocial Chat Counseling Service on Further Help-Seeking Among Children and Young Adults: Longitudinal Study
JMIR Ment Health 2023;10:e43780
URL: https://mental.jmir.org/2023/1/e43780
doi:10.2196/43780
PMID:37195747

Original Paper

Exploring User Perspectives of and Ethical Experiences With Teletherapy Apps: Qualitative Analysis of User Reviews

Eunkyung Jo[1], MS; Whitney-Jocelyn Kouaho[1], MS; Stephen M Schueller[1,2], PhD; Daniel A Epstein[1], PhD

[1]Department of Informatics, University of California, Irvine, CA, United States

[2]Department of Psychological Science, University of California, Irvine, CA, United States

Corresponding Author:
Daniel A Epstein, PhD
Department of Informatics
University of California
6093 Donald Bren Hall
Irvine, CA, 92697
United States
Phone: 1 949 824 4197
Email: epstein@ics.uci.edu

Abstract

Background: Teletherapy apps have emerged as a promising alternative to traditional in-person therapy, especially after the COVID-19 pandemic, as they help overcome a range of geographical and emotional barriers to accessing care. However, the rapid proliferation of teletherapy apps has occurred in an environment in which development has outpaced the various regulatory and ethical considerations of this space. Thus, researchers have raised concerns about the ethical implications and potential risks of teletherapy apps given the lack of regulation and oversight. Teletherapy apps have distinct aims to more directly replicate practices of traditional care, as opposed to mental health apps, which primarily provide supplemental support, suggesting a need to examine the ethical considerations of teletherapy apps from the lens of existing ethical guidelines for providing therapy.

Objective: In this study, we examined user reviews of commercial teletherapy apps to understand user perceptions of whether and how ethical principles are followed and incorporated.

Methods: We identified 8 mobile apps that (1) provided teletherapy on 2 dominant mobile app stores (Google Play and Apple App Store) and (2) had received >5000 app reviews on both app stores. We wrote Python scripts (Python Software Foundation) to scrape user reviews from the 8 apps, collecting 3268 user reviews combined across 2 app stores. We used thematic analysis to qualitatively analyze user reviews, developing a codebook drawing from the ethical codes of conduct for psychologists, psychiatrists, and social workers.

Results: The qualitative analysis of user reviews revealed the ethical concerns and opportunities of teletherapy app users. Users frequently perceived unprofessionalism in their teletherapists, mentioning that their therapists did not listen to them, were distracted during therapy sessions, and did not keep their appointments. Users also noted technical glitches and therapist unavailability on teletherapy apps that might affect their ability to provide continuity of care. Users held varied opinions on the affordability of those apps, with some perceiving them as affordable and others not. Users further brought up that the subscription model resulted in unfair pricing and expressed concerns about the lack of cost transparency. Users perceived that these apps could help promote access to care by overcoming geographical and social constraints.

Conclusions: Our study suggests that users perceive commercial teletherapy apps as adhering to many ethical principles pertaining to therapy but falling short in key areas regarding professionalism, continuity of care, cost fairness, and cost transparency. Our findings suggest that, to provide high-quality care, teletherapy apps should prioritize fair compensation for therapists, develop more flexible and transparent payment models, and invest in measures to ensure app stability and therapist availability. Future work is needed to develop standards for teletherapy and improve the quality and accessibility of those services.

(JMIR Ment Health 2023;10:e49684) doi:10.2196/49684

KEYWORDS

teletherapy; therapy; ethical guidelines; ethics; qualitative research; digital mental health; mobile phone

Introduction

Background

In recent years, the increasingly pervasive use of mobile devices and smartphones has begun to redefine how mental health is managed. Teletherapy apps—stand-alone platforms that specifically conduct synchronous therapy sessions with clients who are not physically colocated with their therapists [1]—have emerged as a potential alternative to traditional in-person therapy, especially after the COVID-19 pandemic [2-4]. Prior work has demonstrated that teletherapy addresses a range of geographical and emotional barriers to accessing care by eliminating the need for individuals living in rural areas to travel [4-7] and by offering anonymity for stigmatized populations [7,8]. In this climate, teletherapy apps have garnered substantial investment and attention as a means to expand access to scale mental health care [9]. Ethical standards guide traditional face-to-face psychotherapy, but their application in digital practice has not been well addressed. In this study, we examined teletherapy apps from the viewpoint of user reviews to understand how ethical principles were followed and incorporated.

The rapid proliferation of teletherapy apps has occurred in an environment in which development has outpaced the various regulatory and ethical considerations of this space. At present, no single regulatory authority oversees the practice of teletherapy apps as professional issues are regulated by state licensing boards, and the Food and Drug Administration regulates apps that provide specific treatments under their regulation of software as medical devices (eg, apps that provide computerized behavioral therapy for psychiatric conditions, such as reSET or Somryst) [10]. Moreover, the Food and Drug Administration relaxed the regulation of mental health apps to facilitate the use of these tools during the COVID-19 pandemic [11,12], which could lead to the widespread adoption of unproven and potentially low-quality mental health apps. Researchers have raised concerns about the ethical implications and potential risks of mental health apps given the lack of regulation and oversight [4,13-15], particularly regarding technical glitches [16-20], business models [19,21,22], and professionalism [16]. However, prior work has focused on general mental health apps more broadly rather than on the ethical aspects of teletherapy apps specifically. As teletherapy apps distinctly aim to more directly replicate practices of traditional care compared with mental health apps, which primarily provide supplemental mental health support, there is a greater need to specifically examine the extent to which teletherapy apps and the care provided through them uphold the ethical standards applied to traditional care.

Thus, we analyzed user reviews of commercial teletherapy apps through the lens of ethical codes of conduct for therapists. All licensed therapists are bound by the ethics code of their profession, such as the Ethical Principles of Psychologists and Code of Conduct [23], Code of Ethics of the National Association of Social Workers [24], and Principles of Medical Ethics [25], and state laws and regulations relevant to their status as licensed professionals. The ethical codes of conduct can serve as good metrics for high-quality care as they provide therapists with general guidelines and expectations for their professional conduct [26]. Prior work has used ethical frameworks to examine existing mental health technology [17,22,27-29], suggesting some important ethical considerations for mental health apps, such as privacy [22,28,30], transparency [17,22], and access to care [22,29].

Although variations exist across the ethical codes of conduct for psychologists [23], social workers [24], and psychiatrists [25], six common themes are central to the ethical practice of therapy:

1. *Professionalism*: ethical guidelines for both psychologists and psychiatrists state that they should uphold professional standards of conduct [23,25].
2. *Continuity of care*: ethical guidelines for both psychologists and social workers suggest that they should make efforts to ensure continuity of care when care is interrupted by different factors. Ethical guidelines require psychologists to "plan for facilitating services in the event that psychological services are interrupted by factors such as...unavailability, or relocation [23]." Guidelines further underline that care should be taken when terminating therapy services. Guidelines for social workers state that they should "avoid abandoning clients who are still in need of services [24]."
3. *Affordability of access*: for example, guidelines for social workers state the following: "Consideration should be given to clients' ability to pay [24]."
4. *Cost fairness*: guidelines for social workers mention that they should "ensure that the fees are fair, reasonable, and commensurate with the services performed [24]."
5. *Cost transparency*: guidelines for psychologists highlight that the costs involved in therapy should be accurate and transparent, urging them not to "misrepresent their fees [23]."
6. *Access to care*: for example, guidelines for psychiatrists state that they "shall support access to medical care for all people [25]."

Studies have pointed out a lack of research exploring user experiences of publicly available apps for mental health in spite of the critical need to understand users' real-world experiences to ensure the quality of care provided by mental health apps [22]. Prior work has also pointed out that the evaluation of such apps has primarily focused on professionals, such as therapists, rather than users [16,31-33]. Recent studies have leveraged user reviews of commercial mental health apps as a way to understand user perspectives [16,18-22,34]. User reviews provide a credible source of information for other users to learn about the apps' benefits and shortcomings [16,34], which could also provide researchers with insights into user perspectives [18,20,21]. Specifically, user reviews can provide concrete examples of the ethical challenges that users face with mental health apps [22]. For example, users have reported that technical glitches interfere with the mental health apps' ability to provide timely and continuous support during crises [16,22], which can lead to serious emotional consequences for susceptible populations [18]. Users have also expressed concerns about the business models of mental health apps, perceiving those apps

as exploitative of population vulnerability [19,21] or nontransparent in the payment process [16,22]. Users further feel that therapists matched through mental health apps are not as professional as in-person therapists [16]. However, there is a lack of understanding of how users perceive teletherapy apps specifically in enacting ethical guidelines, which is crucial for enhancing the quality of the mental health care services provided through these apps.

Objectives

Examining users' experiences with teletherapy apps through the lens of ethical codes of conduct can allow us to better understand how users perceive the quality of care that they receive and what considerations are needed to improve the quality of care through such technology. To that end, we examined user reviews of teletherapy apps from app stores through the lens of ethical codes of conduct for therapists. Through our qualitative analysis of the user reviews of teletherapy apps, we unpacked user perspectives and their ethical experiences with teletherapy apps, focusing on whether and how they perceive that such technology enacts the ethical codes of conduct in practice. On the basis of the findings, we discuss the ethical considerations that should be met by teletherapy apps, which will help create standards for how these apps should be designed and how health care policies should be implemented to support individuals' mental health care.

Methods

Data Collection

We searched for potential apps that provide teletherapy services on 2 dominant mobile app stores (Google Play for Android and Apple App Store for iOS) using the search terms "therapy" and "teletherapy." Teletherapy apps differ from teletherapy that is delivered via other web-based platforms (such as Zoom, Doxy.me, Mend, or SecureVideo) in that, in teletherapy apps, the app company acts as the intermediary of care, whereas on other web-based platforms, the therapist, if in private practice, or health systems remain the intermediaries. For example, on teletherapy apps, the companies provide the terms of service, collect payments, and manage data about patients. Given these differences, we focused on mobile teletherapy apps excluding apps that only function within traditional care facilities (eg, the Anthem Sydney Health app and the Kaiser Ginger app). We carefully read the app descriptions to confirm whether each app provided teletherapy services. We only included apps that had received at least 5000 app reviews on both Google Play and the Apple App Store. The number of reviews is a useful proxy for the quality and impact of apps, allowing us to focus on high-quality apps that are widely used [21]. These criteria yielded 8 apps: BetterHelp, Cerebral, Talkspace, Simple Practice, MDLIVE, Amwell, Doctor On Demand, and Teladoc Health.

We wrote Python scripts (Python Software Foundation) to scrape user reviews from the 8 apps. To extract recent critical user feedback on the apps, we applied the following 3 inclusion criteria for filtering. First, we only included reviews that were written in English. Second, we included reviews of ≥50 characters. Filtering out fake reviews [16] and selecting informative reviews [21] are challenges in using user reviews for research purposes. Excluding shorter reviews helped us address these challenges by improving data integrity and uncovering deeper insights in longer reviews [16]. Third, we limited the reviews to the most recent 500 reviews of each app from Google Play and Apple App Store as of December 6, 2022. Aligned with prior work [16,21], focusing on recent reviews allowed us to concentrate on relevant user experiences of the most up-to-date versions of the apps. Finally, we noticed that some apps (Cerebral, MDLIVE, Amwell, Doctor On Demand, and Teladoc) provide medication management as well as therapy and are mostly focused on providing the former. Therefore, for those apps, we decided to only include reviews that mentioned "therapy" or "therapists." We read 10% of the collected data and confirmed that the 2 keywords effectively identified reviews that were relevant to teletherapy. As a result, the final data set included 3268 user reviews. The breakdown of the number of reviews analyzed for each app is provided in Multimedia Appendix 1.

Data Analysis

We used thematic analysis to qualitatively analyze user reviews [35]. In the first phase, the first author (EJ) open coded 1300 user reviews and searched for themes by reviewing and grouping codes into potential themes. We then generated a preliminary codebook with 6 codes (usability, customer support, business model, accessibility, therapist match, and app features) and 17 subcodes. When the entire research team reviewed the preliminary codebook, we noticed that the themes were closely related to ethical principles for mental health professionals, such as billing procedures and access to care. Therefore, in the second phase, we revised the codebook in light of the common themes to the ethical codes of conduct for psychologists, psychiatrists, and social workers. The revised codebook included six parent codes and 13 child codes (the full codebook can be found in Multimedia Appendix 2): (1) professionalism (therapists' professionalism and therapists' unprofessionalism), (2) continuity of care (therapist unavailability, service termination, and technical glitch), (3) cost transparency (concerns about cost transparency and appreciation of cost transparency), (4) cost fairness (billing inaccuracy and access-based billing), (5) access to care (overcoming geographical constraints and overcoming social constraints), and (6) affordability of access (affordability and unaffordability). The first (EJ) and second (WK) authors then coded 150 (approximately 5% of the whole data set of 3268) reviews to calculate the interrater reliability, with the κ value for 6 child codes being >0.8 (unaffordability, professionalism, unprofessionalism, overcoming geographical constraints, billing inaccuracy, and cost transparency) and the κ value for 4 child codes being between 0.6 and 0.78 (affordability, overcoming social constraints, service termination, and technical glitch). We had a low agreement for access-based billing (κ=0.3) and therapist unavailability (κ=0.43) in our initial pass. Therefore, we discussed the differences to resolve ambiguities and ensure consistency, which led to more detailed definitions of those codes. Afterward, the 2 authors coded the rest of the user reviews.

Ethical Considerations

Our research aimed to analyze public app reviews available on commercial app stores. As such, our study did not involve direct interaction with human participants, and all data used for the analysis were publicly available through the app stores. Given that our study solely relied on preexisting and publicly accessible data, it falls under the category of secondary research involving nonidentifiable information [36]. Therefore, we did not seek ethics board review or approval for this study.

Results

Overview

The qualitative analysis of app reviews revealed the ethical experiences of teletherapy app users. In this section, we present the findings of the study, focusing on whether and how users' perspectives and experiences align with the ethical guidelines for mental health professionals. Table 1 summarizes the 6 themes that we identified through the user review analysis. The user review alias indicates which teletherapy app a review was posted for (eg, BetterHelp or Talkspace) on which app store (eg, Google Play or Apple App Store). For example, B-A-1 indicates that the review was for BetterHelp on the Apple App Store.

Table 1. Frequency of codes identified across the app reviews and example quotes (N=3268).

Theme and code	n (%)	Example quote
Professionalism		
Therapists' professionalism	1136 (34.76)	"My therapist constantly checks up on me. I love that you can message them at any time of the day and they usually get back to you fairly fast." [B-A-162]
Therapists' unprofessionalism	479 (14.66)	"My therapist has canceled her appointment with me 4 times now. This is ridiculous." [C-G-247]
Continuity of care		
Technical glitch	540 (16.52)	"The video cuts on and off many times throughout the sessions. Every therapy session had interruptions due to video connection." [A-G-15]
Therapist unavailability	185 (5.66)	"I can't schedule a meeting with my therapist for MONTHS because there are no available time slots available." [TD-A-33]
Service termination	6 (0.18)	"I was literally responding to the therapist's question when I was blocked out of sending anything else because I needed to renew the next month. Seriously? That's how you do your 'patients'? You cut them out of your app as if it's a Netflix subscription?" [B-A-55]
Affordability of access		
Affordability	167 (5.11)	"I am not employed and they have provided me a subscription for low-income people and it has really helped me through this difficult time in my life." [B-G-208]
Unaffordability	151 (4.62)	"Another downside is that [the app] does not accept insurance. So if you have a lower income (like I do), this is not the most affordable option." [B-A-163]
Cost fairness		
Access-based billing	159 (4.86)	"The therapists cancel appointments on the same day with no explanation. ...I paid $500+ for one month of service and got literally no help at all." [T-A-186]
Billing inaccuracy	78 (2.39)	"They also continue to charge my credit card despite me canceling the membership multiple times." [B-A-30]
Cost transparency		
Concerns about cost transparency	75 (2.29)	"I wish you had told me the price before you took all that information from me. Kind of feel manipulated." [B-G-205]
Access to care		
Overcoming geographical constraints	159 (4.86)	"I'd been looking for therapy in my area for years with no luck. It's a small town with limited options so there was never any opening. Thanks to [this app], I've finally had the access I've been needing and getting the assistance I need." [T-G-220]
Overcoming social constraints	24 (0.73)	"I have been using this app for a couple of months now and as someone who suffers from social anxiety and finds it uncomfortable just to go out, it has helped me out tremendously." [TD-A-242]

Professionalism

In our study context, professionalism encompasses aspects including therapists' quality, reliability, and commitment to the therapeutic process. Our findings demonstrate that many users perceived that their therapists met the standards of professionalism as the most frequent code that came up in the reviews was therapists' professionalism (1136/3268, 34.76%; Table 1). Users often mentioned positive qualities of their therapists, such as being "a good listener" (A-A-457), "non-judgemental" (B-G-472), "understanding" (B-A-208), "compassionate" (S-A-249), and "insightful" (B-A-373).

However, the third most frequent code was therapists' unprofessionalism (479/3268, 14.66%; Table 1), suggesting that a substantial number of users still perceived that their therapists did not meet the standards of professionalism. Users often perceived that their therapists did not listen to them and their concerns, describing them as "reading from a script" (C-G-130), giving "copied and pasted responses" (T-G-317), and sounding "like robots" (T-A-8). They also criticized the fact that therapists only gave basic and generic responses, describing the quality of therapists' responses as "fortune cookie grade" (T-A-454). Users further criticized that their therapists were distracted during therapy sessions. Users noticed that their therapists were driving during therapy sessions and were distracted:

> *It seemed like my therapist had me on speakerphone while she was driving. I definitely knew she was driving because, at one point, she yelled at either another car or pedestrians. She was distracted the entire call.* [B-A-235]

Users pointed out other unprofessional behaviors from therapists, such as "taking a phone call" (T-A-57) or "falling asleep" (B-A-289) during the sessions. Users developed their own hypotheses regarding the reasons for such unprofessional behaviors from therapists. A user posited that such behaviors resulted from a lack of policies to hold therapists accountable:

> *[The app] will only guarantee to match you with a therapist. They have no policies in place to ensure the therapist is accountable for actually showing up for an appointment.* [B-A-44]

A user suspected that therapists had low commitment as the apps were their side jobs:

> *Every therapist I was paired with made it very obvious that this was their side hustle and they had their own practice that they actually cared about, but not [this app's] patients.* [T-A-489]

In addition, users often pointed out that their therapists were not reliable in keeping their appointments and were not available on the timelines that the app promised. Users pointed out that the therapists sometimes did not deliver on the app's promise of guaranteed response time:

> *[The app] always says 24-hour guaranteed response window, but the therapists never text back.* [T-A-7]

They also faced situations in which their therapists repeatedly canceled appointments:

> *My therapist has canceled her appointment with me 4 times now. This is ridiculous.* [C-G-247]

A user similarly said the following:

> *Made two appointments for counseling in a severely distressing time. Both appointments were canceled just minutes before they were to occur.* [A-A-12]

Users felt that these behaviors were "extremely unprofessional and unethical" (C-G-247) and "absolutely ludicrous" (B-G-103) as it led them to waste their time and money.

Continuity of Care

Users noted technical glitches, therapist unavailability, and abrupt service termination that affected the teletherapy apps' ability to provide continuity of care. The app reviews revealed that technical glitches often interfered with users' ability to engage in continuous care, which was the second most frequent code that came up (540/3268, 16.52%; Table 1). Users frequently mentioned audio or video issues that occurred during their therapy sessions. A user said the following:

> *The video cuts on and off many times throughout the sessions. Every therapy session had interruptions due to video connection.* [A-G-15]

Thus, users had to spend time fixing the issues, which shortened the time they were supposed to spend on therapy. A user stated the following:

> *The constant cutting in and out of the audio takes away much-needed time that could have instead been used to make actual progress in therapy.* [S-A-30]

A user similarly noted the following:

> *My therapist nor I could hear each other. I'm really upset about it because two of my appointments were just spent trying to fix this.* [S-A-269]

Other users mentioned that the therapy sessions ended abruptly because of technical issues. A user said the following:

> *The calls and live sessions drop so much that it's incredibly frustrating.* [B-G-483]

A user similarly noted the following:

> *We had to disconnect and reconnect multiple times, which causes a disturbance in the flow of conversation.* [B-A-229]

A user pointed out that such a technical glitch in apps for mental health was irresponsible:

> *Before you offer an application that's primary purpose is supporting mental health needs, PLEASE be responsible and ensure it works properly.* [T-A-471]

User reviews further demonstrated that therapist unavailability compromised the apps' ability to support continuous care. Users often mentioned that the therapists with whom they had been matched often had limited to no availability in their schedules, which made it challenging for them to engage in weekly therapy as the apps promised. A user said the following:

> *None of my therapists have been available to meet with me weekly. They are often overbooked or unavailable, and I end up seeing them one or two times per month.* [B-A-291]

A user similarly noted the following:

> *I can't schedule a meeting with my therapist for MONTHS because there are no available time slots available.* [TD-A-33]

A user also reported that apps failed to provide quick access to care because of therapist unavailability:

> *I was matched with a therapist who was on time off, which is extremely frustrating because there was a very acute reason for signing up and if I had known that I would have to wait two weeks I'd just go to a normal provider. That's kind of the whole point of [this app], to be convenient and fast, but it failed at this job massively.* [T-G-105]

We also found that users experienced a lack of continuity because of the apps' subscription model. Users mentioned that they lost access to their therapists as soon as their subscription ended without any reminders or warnings. A user said the following:

> *The app immediately cuts you off from communicating with your therapist once your billing is due. There are no reminders or anything to keep you on your toes they just immediately cut you off.* [B-A-17]

A user similarly noted the following:

> *The way they handle payments for subscriptions is manipulative. There is no warning whatsoever that the subscription will expire soon, so you can be prepared to pay to renew it, and once your subscription expires, it will automatically cancel your appointment and lock you out of communicating with your therapist until payment is processed.* [B-A-485]

A user described how difficult it was when they lost access to their therapist without an opportunity to properly terminate care:

> *The worst part is if you have an emergency financial hardship, you immediately are unable to inform/speak to your counselor because everything locks if you miss a payment. I really valued the relationship my counselor that I had and she helped me through a lot. The fact that I couldn't just say goodbye myself was very hard. There should be something against this. I am still battling with anxiety and depression, and this was such a stressful situation not being able to speak to the person who was helping me for months.* [B-A-288]

A user felt that such practice might be justifiable for generic apps but not for apps that provide mental health care:

> *I was literally responding to the therapist's question when I was blocked out of sending anything else because I needed to renew the next month. Seriously? That's how you do your "patients"? You cut them out of your app as if it's a Netflix subscription?* [B-A-55]

Affordability of Access

Some teletherapy app users perceived that apps provide affordable mental health care by reducing prices compared with traditional in-person therapy. A user mentioned that teletherapy apps provided more affordable options compared with mental health care that she could find where she lived:

> *I can usually have 3 or 4 sessions in a month for around $260, which you really can't find that anywhere else in [the city I live].* [B-A-336]

A user also appreciated that the app provided an affordable option for individuals without insurance:

> *For the folks like me whose insurance doesn't cover therapy, the service provided by this app at $65/week is phenomenal. Average [in-person] therapists will cost about $150/week.* [B-A-242]

Users also valued the fact that the apps provided discounts based on individuals' financial circumstances even when the apps did not accept insurance. For example, a user described the following:

> *I am not employed and they have provided me a subscription for low-income people and it has really helped me through this difficult time in my life.* [B-G-208]

Similarly, a user said the following:

> *My income looks good, but I'm working through a lot of debt; I e-mailed them and they understood and gave me a bigger discount for a 3-month period.* [B-A-119]

Users further appreciated that some apps accepted insurance and made care affordable. T-G-220 appreciated that therapy was "covered by my insurance at 100%." A user similarly mentioned the following:

> *I put in my insurance member ID, turns out the sessions are FREE.* [D-A-434]

In contrast, other users perceived that teletherapy apps fell short of their promises of affordability of access as they did not accept insurance. Users thought that the lack of ability to use insurance made teletherapy apps unaffordable, particularly for low-income individuals. A user explained the following:

> *Another downside is that [the app] does not accept insurance. If you have a lower income (like I do), this is not the most affordable option. This is unfortunate because a lot of people with lower incomes need access to convenient quality mental health care. It is very difficult for those with low income to get therapy because they might not have time due to work or they don't have a means of transportation to go to face-to-face counseling. If [the app] could make the prices more affordable or at least accept insurance, they would be catering to an untapped market.* [B-A-163]

A user similarly mentioned the following:

> *The app is not affordable if you have a minimum wage job. The cheapest plan is over $200 a month. ...I*

> *understand this is someone's career, and they are providing a service. At the same time, the people have severe mental health issues and it's really unfair that money is the way of us getting help.* [T-A-404]

Therefore, they desired that apps accept insurance. A user said the following:

> *I wish that BetterHelp would have a representative for each state to negotiate a way to allow recipients of Medicaid/Medicare to receive a reduced rate. Many of these state-sponsored insurances offer behavioral health coverage. This would open up so much opportunity for some individuals.* [B-A-402]

They further desired that apps provide flexibility in the payment model rather than sticking to a monthly subscription model with weekly sessions as it would allow them to have more affordable options. A user said the following:

> *To me, the cost was a HUGE deal breaker. I understand that the therapists need to get paid; however, the fee for unemployed people is $87 per week. It would be nice if it would have given you the option to pay per use and not set up a subscription.* [B-G-256]

A user similarly noted the following:

> *They will force you to take weekly sessions, and their so-called discounted price for those weekly sessions is $87 per session. There is no way I could ever afford weekly sessions. At this point, I can only afford monthly or maybe bi-weekly sessions. The fact that they refuse to allow that option is just wrong.* [B-G-452]

Cost Fairness

Our findings suggest that users often perceive that teletherapy apps fail to fulfill their commitment to ensuring cost fairness. Users frequently brought up that the subscription model adopted by some apps resulted in them paying fees that were not commensurate with the services that they received. All the apps that primarily provided teletherapy (BetterHelp, Talkspace, and Simple Practice) were based on a monthly recurring billing model that promised weekly therapy sessions. However, because of the aforementioned challenges in ensuring therapists' professionalism and continuity of care, apps often could not deliver on their promises, which influenced how users perceived the fairness of their cost models. A user pointed out the following:

> *Three different therapists that either no showed or canceled in the first 2 weeks. So that's like 2 weeks wasted, $200 wasted.* [B-G-257]

A user similarly noted the following:

> *The therapists canceled appointments on the same day with no explanation. ...I paid $500+ for one month of service and got literally no help at all.* [T-A-186]

Users encountered unfair situations in which they had to pay despite not receiving any care because of therapist unavailability. Users frequently mentioned realizing that therapists had no availability for a few weeks after they had been matched, but the apps often refused to issue refunds citing that users had "access" to mental health care during that time. A user said the following:

> *I was matched with a counselor that had zero availability for two weeks, yet I was still charged the same.* [B-A-133]

A user similarly noted the following:

> *I went to her schedule and there was no availability for the upcoming 3 weeks. Why charge me for a service I don't even have access to? Now I'm out of $256, with no type of refund in sight.* [B-A-397]

Users criticized these billing practices, describing them as "unethical" or "predatory." A user stated the following:

> *I am paying hundreds of dollars for video sessions that are expiring because therapists aren't taking appointments. ...It is so unethical to take mentally ill people's money while they're already struggling and give them nothing in return.* [T-A-192]

A user also criticized the following:

> *Predatory payment style. You pay monthly regardless of if they meet with you or not. My therapist had COVID and couldn't meet with me for multiple weeks. I still got charged for all the time I didn't get any therapy.* [C-G-9]

Furthermore, users reported that they were charged although they could not receive care because of the platform's technical glitches. A user explained the following:

> *The app nor the website works. I have been charged for 2 months of services and have not gotten the 1st consultation.* [B-G-416]

A user similarly illustrated the following:

> *I've been charged for a session where the app wouldn't let me join the session.* [T-G-47]

Cost Transparency

The app reviews often manifested users' concerns about the cost transparency of teletherapy apps, suggesting that the apps were falling short of abiding by the guidelines. Users frequently mentioned feeling frustrated as the cost was not disclosed until they finished the sign-up process, wishing pricing information was clear up front. A user explained the following:

> *I wish you had told me the price before you took all that information from me. Kind of feel manipulated out of my personal details just to have the help put behind a paywall that isn't really that affordable.* [B-G-205]

A user also illustrated the following:

> *Cost isn't revealed until the end of the questionnaire. If you can't afford the therapy, you've filled out the whole questionnaire in detail all for nothing, like I just did.* [B-G-227]

Users further suggested that the apps provided unclear information about which insurances covered them. A user stated the following:

> *The app is very wishy-washy about how much your insurance will cover upfront. They should tell you what your insurance covers instead of having people guessing.* [T-A-339]

A user perceived that it was deceiving for apps to advertise that their services were fully covered by insurance when they were not:

> *They charge you everything at once without telling you exactly what insurance covers. They make it out to be like insurance covers the whole cost.* [T-A-400]

Access to Care

Users perceived that teletherapy apps helped promote access to care by allowing them to overcome geographical and social constraints. They appreciated that the apps helped them find therapists when they struggled to find local therapists. Some users mentioned that they were facing difficulties as their local therapists were fully booked:

> *I had searched for local counselors or therapists for months and was on a number of waiting lists when I finally decided to try [the app].* [B-A-3]

They appreciated that the apps matched them with therapists promptly:

> *I tried for a few days calling a local therapist, and they were all booked up weeks in advance. I signed up for [the app], and the next day I was already matched up with a therapist.* [B-A-167]

Users also valued that the apps helped people living in small towns access care:

> *I'd been looking for therapy in my area for years with no luck. It's a small town with limited options so there was never any opening. Thanks to [the app], I've finally had the access I've been needing and getting the assistance I need.* [T-G-220]

Users also appreciated that the apps eliminated the need to travel for therapy:

> *I love this online counseling because the stress to appear in person is reduced tenfold. For me to meet face to face is a stressful 2-hour car ride and a two-hour drive home.* [B-G-145]

They also liked that teletherapy apps enabled them to continue therapy during the COVID-19 pandemic when in-person therapy was not feasible—"Despite COVID-19 social distancing—when in-person locations became scarce—[the app] made it possible to find a therapist and continue improving my mental health." (B-A-430)—describing them as a "life saver" (S-A-176) or "lifeline" (S-G-343) during the pandemic.

Some users also felt that teletherapy apps helped them overcome their social anxiety about therapy. These users mentioned their previous struggles with traditional therapy because of their social anxiety:

> *My intense social anxiety of having to interact with someone in person made me cancel (in-person therapy) frequently. With [the app], I am much less likely to cancel because I can speak to my therapist from the comfort of my own home.* [B-A-140]

Thus, they valued that teletherapy apps let them engage in therapy without the need to go out:

> *I have been using this app for a couple of months now and someone who suffers from social anxiety and finding it uncomfortable just to go out it has helped me out tremendously.* [TD-A-242]

Discussion

Principal Findings

Our findings revealed that users' perceptions of commercial teletherapy apps aligned with some ethical principles of therapy, particularly in terms of promoting access to care and affordability for individuals considered economically disadvantaged. However, users perceived that the apps aligned poorly with other ethical principles such as professionalism, continuity of care, cost fairness, and cost transparency. Some of these problems mirror challenges often experienced in mental health care more generally, such as insurance companies providing ghost networks of therapists (ie, listing therapists who are no longer in the network, are not accepting clients, or have closed their practice) and instances in which therapists display a lack of attentiveness during sessions. However, the unique characteristics of teletherapy apps, such as the remote provision of care, dependence on technology, and existing outside traditional health care facilities, may have exacerbated the challenges in their implementation. Some of these issues could be addressed through straightforward measures such as improving cost transparency. Others require substantially reconfiguring aspects of how teletherapy is conducted, how patients and therapists are matched, and how services are billed in ways that may or may not be possible in current health care systems. In this section, we unpack some ethical implications of our findings and provide some insights for teletherapy apps to ensure the delivery of high-quality care.

Considerations to Enhance Quality of Care

Our findings suggest that, although teletherapy apps offer the potential for providing affordable care to individuals considered economically disadvantaged, these apps may compromise the quality of care by not adequately compensating therapists at market rates. In our study, we found that many users perceived unprofessionalism among therapists who provided services through teletherapy apps as they often gave basic responses, were distracted by other activities during the therapy sessions, and were not reliable in keeping their appointments. This finding is consistent with prior work indicating that users perceive therapists matched through mental health apps as not as professional and qualified as their in-person therapists as they only provide generic responses and show little interest in their clients [16]. We posit that economic reasons may explain these differences in care. As many teletherapy apps provide cheaper rates compared with in-person therapy settings, therapists tend to receive lower financial incentives on those apps. Although

therapists make approximately US $36 per hour on average in the United States [37], therapists on teletherapy apps (eg, BetterHelp and Talkspace) only earn US $14 to US $30 per hour [38]. Therapists in certain states or areas might receive even higher rates, which further reduces the incentive for them to provide high-quality care on these apps. Failure to pay market rates for therapists likely limits their ability to devote sufficient time to meet the expectations that these apps set and compromises the quality of care provided through these apps. Our findings suggest that achieving ethical guidelines of professionalism likely requires a minimum cost to ensure that these therapists are compensated fairly for such work. To ensure quality care, it is critical for teletherapy apps to offer fair compensation. Prior work on apps for depression has highlighted the need to negotiate ways to cover the cost of mental health apps with insurance providers [22]. Our findings similarly pointed to the value of teletherapy apps contracting with insurance providers, Medicaid, and Medicare as it not only would help provide fair compensation for therapists but also might make therapy services available to those who might not otherwise receive care. This might be especially important in areas with mental health care shortages where a sufficient number of providers are not available to meet the demands of those in need. In addition, streamlining operational costs, such as optimizing their technology infrastructure and automating routine administrative tasks, could help balance fair compensation for therapists while ensuring the affordability of teletherapy apps.

Of course, our findings draw from user reviews, which inherently overlook the perspectives of the therapists. Relatively little research has examined therapists' perspectives on teletherapy apps and the services they provide on them, which is critical for understanding their motivations for participating and perceptions of the quality of care they are able to provide in the setup. Although a few popular press articles have interviewed therapists providing services on these platforms and represented their perspectives, their views may not be representative of the broader views of therapists on these platforms [39,40]. An article noted that the compensation structure for therapists on these platforms often resulted in the need to maintain large caseloads for sufficient pay, which might lead therapists to resort to simplistic responses during their therapy sessions [40]. Furthermore, it was noted that the platform itself may continue to push new clients onto therapists even when their caseloads are full. Another press article highlights how the app companies exert considerable control over therapists' schedules and engagement with clients and require the use of scripted responses in text chat therapy sessions under certain circumstances [39]. These practices might limit therapist autonomy and lead to practices that might reflect more on the app company and its policies than the care that individual therapists are able to provide. A counterpoint to this view would be that schedules, engagement practices, and scripted responses might increase the quality of care if aligned with best practices, but it is unclear whether this is always the case. Ultimately, more research that centers on the therapist view would be useful, and these articles suggest that app platforms might play a strong role in shaping therapists' activities on them. Our findings, in conjunction with these insights, underscore the need for future research to delve deeper into the reasons behind therapists' behaviors on these platforms to develop platform design considerations and potentially even policy that help support care that meets practice standards.

Developing Flexible Payment Models for Cost Fairness

Our findings also highlight the ethical challenges of using a subscription model for teletherapy apps, particularly regarding cost fairness. In our study, users noted that subscription models of teletherapy apps are potentially problematic, especially when provider availability and activity do not match advertised services included in those subscriptions. All teletherapy apps that were included in our analysis charged their users monthly with the promise of weekly therapy sessions, but they often could not deliver on those promises because of technical glitches and therapists' unavailability. Although payment models are an important ethical consideration for designing, using, and vetting mental health apps, few studies have paid attention to users' perceptions of the payment strategies used by those apps [19,21]. Recent work has pointed to the ethical challenges of the freemium payment model of mental health apps, which combines free basic features with advanced subscription services [21]. Such a payment model may exploit vulnerable populations as the complexity of the freemium payment model can result in unexpected charges for users who are experiencing mental health crises. In addition, the limited-term offers associated with the freemium payment model could lead users with ongoing mental health needs to receive incomplete treatments. Aligned with prior work, our findings indicate the ethical challenges of a prevalent payment model used in teletherapy apps, specifically with regard to subscription models.

This subscription model commonly used in teletherapy apps is quite different from that of traditional health care services, which most often follow a fee-for-service model in which users are charged for each service that they receive. A subscription model can be beneficial for the teletherapy app company as it provides a predictable revenue stream based on the number of subscribers. This can help with budgeting and planning for the app's maintenance and growth to ensure quality care. Subscription models can also encourage users to engage more regularly with the apps as they have already paid for ongoing access to services, which can help improve the effectiveness of the therapy. However, subscription models make it challenging for teletherapy apps to address unexpected circumstances that impede the provision of care. For example, subscription models may not allow for the flexibility needed to address unexpected circumstances that affect therapists' ability to provide care. If a therapist experiences a personal emergency and cannot keep their appointments, users end up missing out on part of their access to the services that they have paid for. Similarly, if technical issues interrupt a therapy session and a makeup session cannot be scheduled, users also lose partial access to the services that they have paid for.

Prior work has highlighted the need for creating flexible and fair payment models for mental health apps [19,22]. Our study further reinforces the need for teletherapy apps to develop more flexible payment models that account for unexpected circumstances that may affect the provision of care. For

example, subscription models could be adjusted to allow for missed appointments or unexpected technical difficulties that interrupt therapy sessions. In addition, a promising way to ensure continuity of care by accounting for unexpected circumstances is value-based care models, or health care delivery models in which providers are paid based on patient health outcomes rather than based on the volume of health care services that they deliver [39-41]. However, these models require contracting with health care systems, insurers, or potentially even states. Innovative projects in this area have been attempted, such as Reno, Nevada, United States, contracting with Talkspace to provide services to its residents [42,43]. Future research on the impact of different payment models (eg, fee-for-service, subscription, freemium, and value-based care) of teletherapy apps on user experience would be beneficial.

Ensuring App Stability and Therapist Availability

Our findings demonstrate that teletherapy apps provide significant benefits to individuals who face challenges in accessing in-person therapy, including difficulty in finding local therapists, limited ability to travel because of health concerns, and social anxiety. This suggests that teletherapy apps are, to some extent, meeting the expectations set by previous research [13,44] and adhering the ethical guidelines of promoting access to care. However, despite these benefits, our study also revealed several challenges that teletherapy apps face in terms of continuity of care. A significant challenge is the risk of interruptions in care owing to technical glitches. This is consistent with prior work that identified technical issues as a leading cause of negative user experiences with mental health apps [19,20,22,34,45]. Such disruptions can have negative impacts on the mental health of vulnerable populations, who often seek mental health support for urgent reasons [16,18,22]. Our findings suggest a need for developers of teletherapy apps to test their apps more thoroughly than other types of apps to ensure stability and reliability given the sensitive nature of mental health [18]. In addition, prioritizing investment in measures to ensure app stability, high-quality audio and video output, and technical support resources can help mitigate the impact of technical glitches.

Another challenge that we identified was therapist unavailability, as highlighted by prior work on mental health apps [16]. Our findings showed that users frequently realized that their therapists were unavailable for extended periods only after they signed up, which impeded the provision of continuous care. To prevent such situations, teletherapy apps should check users' schedules and therapist availability *before* charging them. This proactive approach can help ensure that users can receive the care that was promised in their subscriptions. In addition, implementing measures to facilitate prompt matching between therapists and users would be beneficial. It is crucial for teletherapy apps to continuously monitor the user-to-therapist ratio to maintain a proportionate number of available therapists in relation to the number of users who sign up for the service. Maintaining a surplus of available therapists beyond the expected number required to meet the needs of users can also help promote continuity of care. Furthermore, it is important for teletherapy apps to better understand the demographics of their platform to make sure that there are sufficient therapists that meet the needs of their users. Future research is needed to better understand the reasons why teletherapy apps experience a shortage of therapists and how these factors could be addressed to improve the continuity of care.

Our study demonstrated some of the strengths and limitations of teletherapy apps when examined through the lens of ethical guidelines. Future research could be expanded to mental health apps more broadly to consider where these apps conform to or fall short of ethical guidelines as well. Professional organizations and policy makers would benefit from a better understanding of the teletherapy app space as it offers the potential to help expand care to those in need but also has the potential to significantly disrupt practice as we know it. Some of these challenges are easily anticipated, such as expanding considerations for interstate practice with lack of national licensure. Other challenges might be harder to anticipate, such as the growing influence that other advances such as large language models, digital phenotyping, and machine learning might have on these teletherapy apps. Our study used the lens of current ethical guidelines, but it will be important to consider how such guidelines might need to be reconsidered or at least clarify which changes in practice teletherapy apps might generate.

Limitations

All the selected teletherapy apps were based in the United States. However, how health care systems are designed affects people's perceptions of and engagement with teletherapy apps as well as traditional in-person mental health care. Therefore, our findings may not be generalizable to other countries with health care systems that are designed differently. In addition, other countries might have different ethical codes for mental health practitioners, which limits our findings to the context of the United States. Further work investigating users' experiences with teletherapy apps in different countries will also help understand the impact of differences in health care systems and regulation policies on the influence of teletherapy care provided.

Although publicly available user reviews on app stores provided us with insightful data, such a data collection method offered no direct engagement with users. Unlike traditional qualitative methods such as interviews or focus groups, our approach did not allow for further probing of user statements or validation of interpretations of those statements. Furthermore, the analysis of user reviews has inherent biases in the data collected. App reviews are typically provided by individuals who voluntarily decide to share their experiences, which may introduce selection bias as those who choose to write reviews may have more extreme opinions or specific motivations [46]. Specifically in the context of mental health, individuals facing mental health crises might prioritize seeking immediate help and support over writing web-based reviews. As a result, the pool of reviews may not be fully representative of the perspectives of the entire user base. Future work on interviewing a diverse range of teletherapy app users will allow researchers to better capture their ethical experiences with the technology and gain a more comprehensive understanding of user perceptions and experiences. In addition, using app reviews ignores the perspective of the therapists or the app companies, which are important parties to consider

when offering guidance on the implementation and regulation of teletherapy apps.

Conclusions

Our study shed light on users' perceptions of the services provided in commercial teletherapy apps and aligned these perceptions with ethical guidelines for therapy. Although users appreciated the access to care and affordability of teletherapy apps, they also reported concerns regarding the professionalism of therapists, technical glitches, and therapist unavailability leading to interrupted care and challenges in ensuring cost fairness and transparency under subscription models. Our findings suggest that teletherapy apps should prioritize fair compensation for therapists, develop more flexible payment models, and invest in measures to ensure app stability and therapist availability to provide high-quality care. We hope that this work contributes to ongoing efforts to develop standards for technology-driven mental health care and improve the quality and accessibility of those services.

Acknowledgments

The authors thank Jacquelyn Powell for writing the scripts for data collection. This project was partially supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through grant UL1 TR001414. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

SMS serves on the Scientific Advisory Board for Headspace, for which he receives compensation, and has received consulting payments from Boehringer Ingelheim, K Health (Trusst), and Otsuka Pharmaceuticals for unrelated work.

Multimedia Appendix 1
The number of reviews used for analysis from Android Google Play and the Apple App Store for 8 teletherapy apps.
[PDF File (Adobe PDF File), 35 KB - mental_v10i1e49684_app1.pdf]

Multimedia Appendix 2
Themes from ethical guidelines for psychiatrists, psychologists, and social workers.
[PDF File (Adobe PDF File), 83 KB - mental_v10i1e49684_app2.pdf]

References

1. Robledo Yamamoto F, Voida A, Voida S. From therapy to teletherapy: relocating mental health services online. Proc ACM Hum Comput Interact 2021 Oct 18;5(CSCW2):1-30 [FREE Full text] [doi: 10.1145/3479508]
2. Mitchell E. "Much more than second best": therapists' experiences of videoconferencing psychotherapy. Eur J Qual Res Psychother 2020;10:121-135 [FREE Full text]
3. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
4. Mazziotti R, Rutigliano G. Tele-mental health for reaching out to patients in a time of pandemic: provider survey and meta-analysis of patient satisfaction. JMIR Ment Health 2021 Jul 29;8(7):e26187 [FREE Full text] [doi: 10.2196/26187] [Medline: 34114956]
5. Madigan S, Racine N, Cooke JE, Korczak DJ. COVID-19 and telemental health: benefits, challenges, and future directions. Can Psychol 2021 Feb 26;62(1):5-11 [FREE Full text] [doi: 10.1037/cap0000259]
6. Schriger SH, Klein MR, Last BS, Fernandez-Marcote S, Dallard N, Jones B, et al. Community Mental Health Clinicians' Perspectives on Telehealth During the COVID-19 pandemic: mixed methods study. JMIR Pediatr Parent 2022 Mar 03;5(1):e29250 [FREE Full text] [doi: 10.2196/29250] [Medline: 35023839]
7. Cipolletta S, Mocellin D. Online counseling: an exploratory survey of Italian psychologists' attitudes towards new ways of interaction. Psychother Res 2018 Nov;28(6):909-924. [doi: 10.1080/10503307.2016.1259533] [Medline: 28068875]
8. Sunjaya AP, Chris A, Novianti D. Efficacy, patient-doctor relationship, costs and benefits of utilizing telepsychiatry for the management of post-traumatic stress disorder (PTSD): a systematic review. Trends Psychiatry Psychother 2020 Jan;42(1):102-110 [FREE Full text] [doi: 10.1590/2237-6089-2019-0024] [Medline: 32321088]
9. Landi H. In record-setting quarter, these 5 telehealth startups raked in $1.6B in funding. Fierce Healthcare. 2021 Jul 30. URL: https://www.fiercehealthcare.com/tech/record-setting-quarter-these-5-telehealth-startups-raked-1-6b-funding [accessed 2023-08-30]
10. DeSalvo KB, Samuels J. Examining oversight of the privacy and security of health data collected by entities not regulated by HIPAA. US Department of Health and Human Services. 2016 Jul 19. URL: https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf [accessed 2023-08-30]

11. Brauser D. COVID-19: dramatic changes to telepsychiatry rules and regs. Medscape. 2020 Mar 26. URL: https://www.medscape.com/viewarticle/927556 [accessed 2023-08-30]
12. Al-Faruque F. FDA relaxes regs for COVID-19 mental health apps. Medtech Insight. 2020 Apr 15. URL: https://medtech.pharmaintelligence.informa.com/MT126593/FDA-Relaxes-Regs-For-COVID-19-Mental-Health-Apps [accessed 2023-08-30]
13. Wykes T, Schueller S. Why reviewing apps is not enough: Transparency for Trust (T4T) principles of responsible health app marketplaces. J Med Internet Res 2019 May 02;21(5):e12390 [FREE Full text] [doi: 10.2196/12390] [Medline: 31045497]
14. Pereira-Sanchez V, Adiukwu F, El Hayek S, Bytyçi DG, Gonzalez-Diaz JM, Kundadak GK, et al. COVID-19 effect on mental health: patients and workforce. Lancet Psychiatry 2020 Jun;7(6):e29-e30 [FREE Full text] [doi: 10.1016/S2215-0366(20)30153-X] [Medline: 32445691]
15. Martinez-Martin N, Kreitmair K. Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Ment Health 2018 Apr 23;5(2):e32 [FREE Full text] [doi: 10.2196/mental.9423] [Medline: 29685865]
16. Haque MR, Rubya S. "For an app supposed to make its users feel better, it sure is a joke" - an analysis of user reviews of mobile mental health applications. Proc ACM Hum Comput Interact 2022 Nov 11;6(CSCW2):1-29 [FREE Full text] [doi: 10.1145/3555146]
17. Bowie-DaBreo D, Sünram-Lea SI, Sas C, Iles-Smith H. Evaluation of treatment descriptions and alignment with clinical guidance of apps for depression on app stores: systematic search and content analysis. JMIR Form Res 2020 Nov 13;4(11):e14988 [FREE Full text] [doi: 10.2196/14988] [Medline: 33185566]
18. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. J Med Internet Res 2018 Jun 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]
19. Oyebode O, Alqahtani F, Orji R. Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews. IEEE Access 2020;8:111141-111158 [FREE Full text] [doi: 10.1109/access.2020.3002176]
20. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]
21. Eagle T, Mehrotra A, Sharma A, Zuniga A, Whittaker S. "Money doesn't buy you happiness": negative consequences of using the freemium model for mental health apps. Proc ACM Hum Comput Interact 2022 Nov 11;6(CSCW2):1-38 [FREE Full text] [doi: 10.1145/3555155]
22. Bowie-Dabreo D, Sas C, Iles-Smith H, Sünram-Lea S. User perspectives and ethical experiences of apps for depression: a qualitative analysis of user reviews. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA p. 1-24 URL: https://dl.acm.org/doi/abs/10.1145/3491102.3517498 [doi: 10.1145/3491102.3517498]
23. Ethical principles of psychologists and code of conduct. American Psychological Association. 2016. URL: https://www.apa.org/ethics/code [accessed 2023-08-30]
24. Code of ethics. National Association of Social Workers. 2021. URL: https://www.socialworkers.org/About/Ethics/Code-of-Ethics/Code-of-Ethics-English#:~:text=Ethical%20Principle%3A%20Social%20workers%20respect,'%20socially%20responsible%20self%2Ddetermination [accessed 2023-08-30]
25. The principles of medical ethics with annotations especially applicable to psychiatry. Am J Psychiatry 1973 Sep;130(9):1058-1064. [Medline: 11643545]
26. Barnett JE. The ethical practice of psychotherapy: easily within our reach. J Clin Psychol 2008 May;64(5):569-575. [doi: 10.1002/jclp.20473] [Medline: 18384116]
27. Jones MD, Anderson Z, Häkkilä J, Cheverst K, Daiber F. HCI outdoors: understanding human-computer interaction in outdoor recreation. In: Proceedinngs of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI EA '18; April 21-26, 2018; Montreal, QC p. 1-8 URL: https://dl.acm.org/doi/10.1145/3170427.3170624 [doi: 10.1145/3170427.3170624]
28. Karcher NR, Presser NR. Ethical and legal issues addressing the use of mobile health (mHealth) as an adjunct to psychotherapy. Ethics Behav 2016 Sep 03;28(1):1-22 [FREE Full text] [doi: 10.1080/10508422.2016.1229187]
29. Schueller SM, Torous J. Scaling evidence-based treatments through digital mental health. Am Psychol 2020 Nov;75(8):1093-1104 [FREE Full text] [doi: 10.1037/amp0000654] [Medline: 33252947]
30. Jones N, Moffitt M. Ethical guidelines for mobile app development within health and mental health fields. Prof Psychol Res Pr 2016 Apr;47(2):155-162 [FREE Full text] [doi: 10.1037/pro0000069]
31. Martinez de Alva FE, Wadley G, Lederman R. It feels different from real life: users' opinions of mobile applications for mental health. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction. 2015 Presented at: OzCHI '15; December 7-10, 2015; Parkville, Australia p. 598-602 URL: https://dl.acm.org/doi/10.1145/2838739.2838806 [doi: 10.1145/2838739.2838806]
32. Whitton AE, Proudfoot J, Clarke J, Birch MR, Parker G, Manicavasagar V, et al. Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015 Mar 04;2(1):e3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]

33. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
34. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
36. University of California Irvine, Office of Research. URL: https://research.uci.edu/human-research-protections/do-you-need-irb-review/self-exempt/ [accessed 2023-09-01]
37. Therapist average salary in the USA, 2023. talent.com. URL: https://www.talent.com/salary?job=therapist#:~:text=How [accessed 2023-08-30]
38. Perry D. How much do online therapy platforms pay their therapists? TherapyReviews. 2022. URL: https://therapy-reviews.com/blog/comparisons/how-much-do-online-therapy-platforms-pay-their-therapists/ [accessed 2023-02-05]
39. NEJM Catalyst. What is value-based healthcare? NEJM Catal 2017 Jan;3(1) [FREE Full text] [doi: 10.1056/CAT.17.0558]
40. Teisberg E, Wallace S, O'Hara S. Defining and implementing value-based health care: a strategic framework. Acad Med 2020 May;95(5):682-685 [FREE Full text] [doi: 10.1097/ACM.0000000000003122] [Medline: 31833857]
41. Scott DM. United States health care system: a pharmacy perspective. Can J Hosp Pharm 2016 Jul;69(4):306-315 [FREE Full text] [doi: 10.4212/cjhp.v69i4.1585] [Medline: 27621491]
42. Conrad B, Davis J. City's talkspace proposal draws mixed reactions. This Is Reno. 2020 Nov 30. URL: https://thisisreno.com/2020/11/citys-talkspace-proposal-draws-mixed-reactions/ [accessed 2023-08-30]
43. Areàn PA, Hoa Ly K, Andersson G. Mobile technology for mental health assessment. Dialogues Clin Neurosci 2016 Jun;18(2):163-169 [FREE Full text] [doi: 10.31887/DCNS.2016.18.2/parean] [Medline: 27489456]
44. Hull SC, Oen-Hsiao JM, Spatz ES. Practical and ethical considerations in telehealth: pitfalls and opportunities. Yale J Biol Med 2022 Sep;95(3):367-370 [FREE Full text] [Medline: 36187411]
45. Waycott J, Kelly R, Baker S, Barbosa NB, Thach K, Lederman R. The role of staff in facilitating immersive virtual reality for enrichment in aged care: an ethic of care perspective. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA p. 1-17 URL: https://dl.acm.org/doi/10.1145/3491102.3501956 [doi: 10.1145/3491102.3501956]
46. Fu B, Lin J, Li L, Faloutsos C, Hong J, Sadeh N. Why people hate your app: making sense of user feedback in a mobile app store. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013 Presented at: KDD '13; August 11-14, 2013; Chicago, IL p. 1276-1284 URL: https://dl.acm.org/doi/10.1145/2487575.2488202 [doi: 10.1145/2487575.2488202]

Edited by J Torous; submitted 05.06.23; peer-reviewed by A AL-Asadi, S Chokshi; comments to author 28.07.23; revised version received 08.08.23; accepted 10.08.23; published 22.09.23.

Please cite as:
Jo E, Kouaho WJ, Schueller SM, Epstein DA
Exploring User Perspectives of and Ethical Experiences With Teletherapy Apps: Qualitative Analysis of User Reviews
JMIR Ment Health 2023;10:e49684
URL: https://mental.jmir.org/2023/1/e49684
doi:10.2196/49684
PMID:37738085

Original Paper

Training Staff Across the Veterans Affairs Health Care System to Use Mobile Mental Health Apps: A National Quality Improvement Project

Pearl McGee-Vincent[1,2], PsyD; Margaret-Anne Mackintosh[1], PhD; Andrea L Jamison[1], PhD; Katherine Juhasz[1,3], MSc; Colleen Becket-Davenport[1,4], PsyD; Jeane Bosch[1], MPH, PhD; Timothy J Avery[1,5], PsyD; Lauren Glamb[1,6], PsyD; Shilpa Hampole[1], BSc

[1]Dissemination and Training Division, National Center for PTSD, Veterans Affairs Palo Alto Health Care System, Menlo Park, CA, United States

[2]Office of Mental Health and Suicide Prevention, Veterans Health Administration, Menlo Park, CA, United States

[3]National Training Division, Education Service, Veterans Benefits Administration, Washington, DC, United States

[4]Department of Psychiatry, Weill Cornell Medical College, New York, NY, United States

[5]Peninsula Vet Center, Readjustment Counselling Services, US Department of Veterans Affairs, Menlo Park, CA, United States

[6]Veterans Affairs Pacific Islands Health Care System, Honolulu, HI, United States

Corresponding Author:
Pearl McGee-Vincent, PsyD
Office of Mental Health and Suicide Prevention
Veterans Health Administration
795 Willow Rd (334-NCPTSD)
Menlo Park, CA, 94025
United States
Phone: 1 650 353 8847
Email: pearl.mcgee-vincent@va.gov

Abstract

Background: The National Center for PTSD, within the Department of Veterans Affairs (VA), has developed a suite of free, publicly available, evidence-informed apps that can reach an increasing number of veterans and bridge gaps in care by providing resources to those who are not engaged in mental health treatment. To expand the reach of these apps, staff across VA service lines learned about these apps, their features and limitations, and how to introduce them to veterans.

Objective: This study aimed to develop, disseminate, and evaluate a training for multidisciplinary staff as part of a national quality improvement project to increase the reach of mobile mental health apps as a resource for veterans.

Methods: Sites from all of VA's 18 geographic regions enrolled in this project. At each site, a minimum of 25 VA staff members who had direct contact with veterans, including staff from the mental health service line and all other service lines, were recruited to participate. Training included a 3-hour multidisciplinary *core* module, and a 1-hour *clinical integration* module designed specifically for mental health clinicians. Owing to the COVID-19 pandemic, the trainings were adapted to a live, web-based format. Pre- and posttraining surveys assessed program reach (ie, participants enrolled per site), satisfaction, and effectiveness of the training as measured by changes in knowledge, basic skills, and behavioral intentions to use apps with veterans.

Results: A total of 1110 participants representing 34 disciplines at 19 VA sites completed the training. Overall, 67% (743/1109) of participants were mental health staff members. Sites averaged 58.4 participants (SD 36.49, median [IQR] 51). Most (961/1024, 93.85%) participants were satisfied with the training and reported that they (941/1018, 92.44%) would recommend it to others. App knowledge scores significantly increased from pretraining (mean 80.8% correct, SD 15.77%) to posttraining (mean 91.1% correct, SD 9.57%; $P<.001$). At posttraining, participants also reported greater confidence in their ability to show veterans how to *download* ($z=-13.86$; $P<.001$) and *use* VA mental health apps ($z=-15.13$; $P<.001$). There was near universal endorsement by staff for their intentions to recommend apps to veterans as well as their ability to think of at least one specific veteran to whom they could recommend an app. Staff also reported a strong motivation to encourage other VA staff to share apps with veterans.

Conclusions: The training far exceeded the initial goals for staff recruitment and training for all three metrics. Overall, 33% (366/1109) of participants came from service lines outside of mental health, indicating the feasibility of introducing these mental health resources during medical appointments and in other contexts.

(JMIR Ment Health 2023;10:e41773) doi:10.2196/41773

KEYWORDS

mental health; mobile apps; digital health; technology; veterans; training

Introduction

To expand access to mental health self-management resources, the Department of Veterans Affairs' (VA) National Center for PTSD develops free mental health apps (VA MH apps) that provide psychoeducation, symptom tracking, and coping tools for PTSD and other related mental health concerns [1,2]. Although these self-care apps (eg, PTSD Coach, Insomnia Coach, and Mindfulness Coach) do not replace treatment, they have the potential to bridge gaps in care by making mental health resources, including information for accessing professional care, available to veterans who experience barriers to treatment [3]. In addition, for veterans participating in evidence-based psychotherapy, treatment companion apps (eg, PE Coach for prolonged exposure therapy) can facilitate aspects of treatment protocols (eg, access to educational materials, between-session homework, and session recordings) [4].

With an estimated 10,000 mental health apps available [5] and minimal quality control or gatekeeping in the app marketplaces [6], it is important that VA staff recommend apps that are evidence-informed and do not compromise patient privacy. VA MH apps are free to download, do not require an account or log-in, do not collect or store identifiable data, and are developed with subject matter expertise [2]. Furthermore, VA MH apps are required to comply with Section 508 of the Rehabilitation Act [7], meaning that they are accessible to individuals with disabilities. Although VA MH apps are available to any member of the public, content is designed to be acceptable to a veteran user base. In fact, VA MH apps have been well received by VA patients and staff [8] and are feasible to use [9-11]. Although not designed to constitute stand-alone treatment interventions, VA MH apps contain active components of evidence-based psychotherapies [12]. Naturalistic use data indicate that the PTSD Coach app, in particular, has promising potential as a public health resource [13,14]. In addition, PTSD Coach users who received clinician support when using the app in primary care were more likely to accept a mental health referral and attend PTSD treatment [15].

Despite the potential benefits of VA MH apps and the increasing number of veterans with access to a smartphone or tablet, a minority of veterans engaged in VA care have heard of or used these apps [16]. To increase awareness and use of a similar suite of apps, Armstrong et al [17] implemented a training program over a 3-year period (2014-2017) in which 760 mental health clinicians, primarily in the Department of Defense (DoD), were trained on how to integrate VA and DoD apps into mental health treatment. Given that many VA patients with mental health needs do not engage in formal mental health treatment [18], training staff outside mental health settings to use these apps could be an important way to reach such patients. As VA MH apps for self-care can be introduced by any VA staff member (eg, chaplains, primary care physicians, and peer support specialists), a training program was developed to meet the learning needs of staff who are part of the mental health service line (MHSL) as well as all other service lines (AOSLs). The training program was designed to increase participants' (1) knowledge about VA MH apps, (2) confidence in their ability to download and use these apps with veterans, and (3) intention to use VA MH apps with veterans in the months following the training. Originally designed to be delivered in-person, the training was modified to a live, web-based format due to the COVID-19 pandemic and the resulting travel restrictions.

This training was part of a national quality improvement project to increase veterans' access to VA MH apps as a resource. As training is important but often insufficient to implement new practices [19], the project team used Implementation Facilitation [20] to support the implementation and sustainment of VA MH apps as an integrated part of care at participating sites. Implementation Facilitation is a widely used set of practices that leverages trained Facilitators along with tailored organizational strategies to help health care organizations overcome barriers and increase the adoption of evidence-based or promising practice [20]. This paper will focus primarily on the evaluation of the training, and a subsequent manuscript will elaborate on the results of the implementation activities that followed.

Methods

Formative Evaluation

Formative evaluation (FE) is a process of evaluating and modifying the content and design of training as it is developed [21]. As part of the FE, interviews were conducted with VA staff to identify learner needs, tailor training objectives, and adapt content from existing training materials. For example, FE helped the project team to determine how to modify the training content to accommodate the needs of staff from different disciplines. This process was ongoing, meaning that the training was modified over time, as new information was collected from participants. This information was obtained formally (via surveys) and informally (via verbal participant feedback), as well as through interviews with Facilitators and members of the project team who delivered the training and served as the primary point of contact with participating sites.

Recruitment

Site-Level

To maximize the reach of the project, sites from each of VA's 18 geographic regions that span the entire United States were invited to apply with the goal of enrolling one participating site per region. These sites were identified, via FE interviews, as having a site champion (defined below), leadership buy-in, and interest in participation from mental health, and at least one other service line.

Sites interested in enrolling were required to (1) identify a site champion, or "mHealth Specialist," to lead the initiative locally,

including coordinating with Facilitators for training and subsequent implementation activities; (2) identify at least 25 staff participants; (3) provide time for enrolled staff to participate in a live 4-hour training; (4) identify and engage key stakeholders (eg, managers and leaders from across programs) to participate in a site visit to include leadership briefings and an implementation planning session; and (5) agree to participate in the 3-month implementation initiative following the training.

Participant-Level

All staff members at enrolled sites who had direct contact with veterans were eligible to participate, pending supervisor approval. In addition to those who provide direct clinical care, other staff members such as medical support assistants, librarians, and peer support specialists were also eligible. Facilitators provided tailored support and recruitment materials (eg, information packets and flyers) to help sites identify and recruit staff to participate in the project. Ultimately, the sites used several recruitment strategies. For example, mHealth Specialists and service-line leadership at most sites disseminated a project recruitment flyer via email with an electronic enrollment link to all service line staff. Other sites worked with a public affairs officer to email the flyer to all staff. At most sites, the participants were self-selected. However, in some instances, supervisors from individual teams nominated representatives to participate in the training. Occasionally, the entire team or service line was directed by local leadership to enroll in the project.

Design of the VA MH Apps Training

Development of the Training

The training curriculum was adapted from training by Armstrong et al [17] and trainings developed previously by the project team. These include conference workshops on VA MH apps [22], a web-based course [23], and a book chapter describing how to integrate apps into care [24]. These materials were integrated with information collected during FE interviews and adapted to be relevant for both MHSL and AOSL participants with varying levels of skills and experience using the VA MH apps.

The project team created a 3-hour *core* training for all staff, regardless of discipline, and a 1-hour *clinical integration* module designed specifically for staff who provide mental health treatment. Topics for the core training included the rationale for recommending VA MH apps for self-care, live demonstrations of self-care apps, practical suggestions for introducing apps to veterans, and an overview of related resources for veterans and VA staff. The live app demonstrations featured the 2 most downloaded VA MH apps: PTSD Coach and Mindfulness Coach. In addition, the digital Safety Plan, a newly added feature of the PTSD Coach app, was designed to provide a readily accessible, mobile option for the safety planning for suicide intervention worksheet. The training provided practical examples for introducing apps in a range of settings and scenarios to help participants determine how to best fit these resources into their workflows. The 1-hour clinical integration module then focused on apps as an adjunct to mental health treatment. Topics relevant to clinicians (eg, education on obtaining informed consent when integrating apps into care and demonstration of treatment companion apps) were covered in this section. All training sessions were delivered by Facilitators.

Consistent with FE, the project team continued to refine the training based on systematic documentation of feedback, including participant input during training, feedback on posttraining surveys, and posttraining interviews with Facilitators. Some examples of changes made during the course of the project included the addition of a live demonstration of the Anger & Irritability Management Skills (AIMS) app during the clinical integration hour, rearranging material to improve flow, and the creation of flexible breakout groups in the clinical integration module, in which participants could choose the treatment companion app they would like to see (eg, PE Coach vs CPT Coach). Training modifications were often larger in scope following the first few training sessions; however, small modifications were made as required through the course of the training period.

In addition to increasing participants' awareness of VA MH apps and their ability to download and use them competently with veterans in accordance with their scope of practice, the third objective of the training was to increase participants' intentions to apply what they had learned. As part of the training, participants were encouraged to share VA MH apps with veterans and other staff over the course of the 3 months following the training and beyond.

Site Visits

The trainings were designed to take place during an in-person site visit led by 2 Facilitators at each participating site. Sites elected to offer 1 or 2 staff training days, depending on the number of participants enrolled, the capacity of the conference room (for in-person visits), and staff schedules. The participants completed the surveys (described below) before and after each training. In addition to the trainings, site visits included an implementation planning session and leadership briefings, which were designed to support the application of training and the sustainment of new practices.

Adaptations Due to the COVID-19 Pandemic

Site visits, including live training sessions, were scheduled to take place on a rolling basis over 12 months, from January to December 2020, at a rate of approximately 3 visits every 2 months. The first 3 site visits occurred in person before the COVID-19 pandemic. Several in-person site visits scheduled for the spring were postponed and ultimately converted to a live, web-based format. After a pause and recalibration, the first web-based site visit occurred in June 2020, and the remaining 15 site visits occurred via web from July to December.

Instead of each visit spanning 1 to 3 days on-site, web-based site visit activities could be spread out over 1 to 2 weeks to accommodate multiple schedules. Trainings were also adapted to make them more engaging in web-based formats [25,26]. For example, breakout groups, multimedia elements, and polls were added to foster audience participation. Facilitators encouraged participants to interact in the chat by offering prompts and answering questions in real time. The breakout groups discussed

pertinent training topics and promoted the participants' engagement. Finally, Facilitators included an interactive exercise after each live demonstration of an app, which allowed participants to gain hands-on experience navigating through the app to find answers to prompts. QR codes were also added to the training slides and other project materials so that participants could easily access them during the web-based training.

Measures

Survey of Participant Characteristics, Attitudes, and Use of Technology

Before training, participants completed web-based surveys regarding their demographics and role within VA, as well as their attitudes toward, skills, comfort with, and use of VA MH apps. Items measuring basic skills, as well as knowledge about and behavioral intentions to use VA MH apps, were assessed again immediately posttraining. Individual items from these sources were used to describe the sample and compare the background characteristics between MHSL and AOSL staff. Participants who reported serving both the MHSL and AOSL were included in the MHSL group.

Training Program Reach

To assess the reach of the training, the metrics used were (1) a minimum of 25 participants per site and (2) participation from both MHSL and AOSL staff.

Training Satisfaction

Two items in the posttraining survey were used to assess the participants' satisfaction with the program. The first item asked how satisfied each person was with the training and the second evaluated whether they would recommend the training to others. Each item was rated on a 5-point scale ranging from *strongly agree* to *strongly disagree* with a midpoint of *neither agree nor disagree*. We reported the percentage of participants who either *strongly agreed* or *agreed* with the 2 questions as our measure of program satisfaction.

Training Effectiveness

Three domains were used to assess the effectiveness of the training program: (1) increase in knowledge about VA MH apps, (2) acquisition of the basic skills needed to demonstrate how to *download* and *use* VA MH apps with veterans, and (3) participants' behavioral intentions to use VA MH apps with veterans and other staff in the 3 months following training.

Knowledge Check

Knowledge about using VA MH apps was assessed using a 16-item knowledge test created specifically for this project. The item content was developed by Facilitators and other subject matter experts and was based on the information highlighted during the training. The percentage of correctly answered items was used as a knowledge metric.

Basic Skills

Possession of the basic skills to use VA MH apps with veterans was measured using 2 items assessing confidence in one's "knowledge and skills necessary to demonstrate to Veterans how to *download* VA apps" and "knowledge and skills necessary to demonstrate to Veterans how to *use* VA apps." Each item was rated on a 5-point scale ranging from *strongly agree* to *strongly disagree* with *neither agree nor disagree* as the midpoint.

Behavioral Intentions

Behavioral intentions to use VA MH apps with veterans were measured using 3 items adapted from Kim and Park, 2012 [27], which are grounded in the Theory of Planned Behavior by Ajzen [28]. The questions were (1) "I will recommend VA apps to Veterans in the next 3 months," (2) "I can think of at least one Veteran I plan to recommend or use a VA app within the next month," and (3) "I will encourage other VA staff to recommend or use VA apps with Veterans." Each item was scored on a 4-point scale: *definitely will*, *probably will*, *probably will not*, and *definitely will not*.

Statistical Analysis

Analyses focused on descriptive statistics detailing the composition and characteristics of the key groups and primary outcomes. Inferential statistics, including 2-sided *t* tests and general linear models, were used to analyze continuous outcomes. Chi-square tests were used to compare independent groups, and Wilcoxon signed-rank tests were used to compare paired groups when analyses included nominal and ordinal outcomes. In addition to assessing the statistical significance of the results, we provided effect size estimates. For *t* tests, standardized mean differences (Cohen *d*) are presented using the conventional interpretative guidelines of 0.20=small, 0.50=medium, and 0.80=large effects [29]. Effect sizes for the general linear models are summarized using partial η^2 with interpretative guidelines for 0.01=small, 0.09=medium, and 0.25=large effects [29]. All the available data were used for each analysis. Because multiple statistical tests were conducted, family-wise error was controlled using the Holm step-down method, which yielded a criterion of .001 for unadjusted *P* values [30]. We report unadjusted *P* values, and those equal to or smaller than this criterion (*P*=.001) were considered statistically significant. All analyses were conducted using SPSS (version 24.0; IBM Corp).

Missing data rates varied from less than 1% for many of the background variables (eg, service lines and disciplines), approximately 25% (about 278/1110) for basic skills and behavioral intention items, and approximately 30% (about 333/1110) for sociodemographic items (eg, age, gender, and education), which appeared at the end of the survey. Of the 1110 participants enrolled in the project, 862 (77.66%) started the pretraining survey and 1023 (92.16%) started the posttraining survey. As most statistical procedures used (ie, chi-square analyses and Wilcoxon rank sum tests) do not have easily accessible routines to pool the results required to use modern multiple imputation procedures, sensitivity analyses were conducted using 20 multiply imputed data sets, in which (1) the analyses described were replicated using the 20 imputed sets, and (2) analyses using similar but more complex statistical procedures answering similar questions (eg, binary or ordinal regression) were explored. The patterns of the results remained unchanged across the analyses. Thus, initial analyses are provided.

Ethics Approval

Ethics approval was obtained from the Stanford Institutional Review Board (Protocol #60207), and this project was determined not to qualify as human subjects research.

Results

Sample Description

The sample consisted of 1110 participants and included 42.42% (467/1101) staff members who reported that they provided mental health services or psychotherapy. In addition, 53.18% (569/1070) of participants reported conducting safety planning interventions for suicide prevention. Tables 1 and 2 present the staff work characteristics and sociodemographic information of MHSL staff (743/1109, 67%) and AOSL staff (366/1109, 33%, one participant's service line could not be identified). A total of 39 participants reported serving both on the mental health and another service line and were counted in the MHSL group. The 2 staff types were similar in gender, disability status, years of experience working with veterans, and average number of hours per week spent interacting with veterans. The 2 groups differed on two background characteristics. First, MHSL staff members were slightly younger (mean 44.2, SD 10.16) compared with AOSL staff members (mean 46.0, SD 10.9), $t_{743}=2.11$, $P=.04$, Cohen $d=0.17$. Second, MHSL staff reported higher educational attainment compared with AOSL staff, $\chi^2_5=73.41$, $P<.001$.

The 1110 participants represented a range of service lines with 67% (743/1109) from MHSL, 14.92% (165/1106) from primary care, 14.29% (158/1106) from Medical Specialty service lines, and 7.96% (88/1106) from other service lines. Across service lines, 68.17% (754/1106) of the participants worked in outpatient programs, 20.80% (230/1106) reported working across 2 or more program types, 9.13% (101/1106) worked in inpatient or residential programs, and 1.90% (21/1106) reported working in research or administrative roles. The participants represented 34 different disciplines (Table 3).

Table 1. Participants' age and work characteristics (N=862).

Characteristics	MHSL[a] staff (n=602[b])		AOSL[c] staff (n=260[b])		Statistical results		
	n	mean (SD)	n	mean (SD)	*t* test (*df*)	*P* value	Cohen *d*[d]
Age (years)	521	44.2 (10.2)	234	46.0 (10.9)	2.1 (753.0)	.04	0.2
Experience working with veterans (years)	549	8.4 (5.6)	243	8.0 (6.4)	−0.87 (413.5)	.41	−0.1
Hours interacting with veterans per week	545	24.2 (11.0)	239	23.5 (13.8)	−0.76 (376.6)	.49	−0.1

[a]MHSL: mental health service line.

[b]862 of the 1110 participants (77.66%) enrolled in the project responded to the pretraining survey.

[c]AOSL: all other service lines.

[d]Standardized mean difference representing between-group effect size as Cohen *d*, guideline for effect sizes: small=0.20, medium=0.50, and large=0.80.

Table 2. Participants' sociodemographic descriptors (N=862).

Characteristics	MHSL[a] staff (n=602[b]), n (%)	AOSL[c] staff (n=260[b]), n (%)	Statistical results	
			χ^2 (*df*)	*P* value
Gender			7.4 (3)	.06
Woman	397 (71.8)	193 (79.1)		
Man	144 (26)	46 (18.9)		
Nonbinary	0 (0)	1 (0.4)		
Prefer not to say	12 (2.2)	4 (1.6)		
Race or ethnicity[d]				
American Indian or Alaska Native	11 (1.3)	3 (0.3)	0.5 (1)	.47
Asian or Asian American	37 (4.3)	23 (2.7)	2.0 (1)	.16
Black or African American	75 (8.7)	43 (5)	2.5 (1)	.11
Hispanic, Latino, or Spanish	50 (5.8)	20 (2.3)	0.1 (1)	.76
Native Hawaiian or other Pacific Islander	5 (0.8)	4 (0.5)	0.9 (1)	.35
White	408 (47.4)	165 (19.2)	1.6 (1)	.21
Another race or ethnicity	6 (0.7)	6 (0.7)	2.3 (1)	.13
Education			73.4 (5)	<.001
High school or General Educational Development	6 (1.1)	2 (0.8)		
Some college or Associate's Degree	35 (6.3)	57 (23.9)		
Bachelor's Degree	47 (8.5)	34 (14.4)		
Some graduate school	11 (2)	13 (5.5)		
Master's degree	245 (44.4)	79 (33.1)		
Doctorate, medical degree, or equivalent	208 (37.7)	53 (22.3)		
Disabilities reported[e]			1.8 (2)	.40
None	483 (86.3)	205 (83.7)		
1	65 (11.6)	31 (12.7)		
≥2 (2-5)	12 (2.1)	9 (3.7)		

[a]MHSL: mental health service line.

[b]862 of the 1110 participants (77.7%) enrolled in the project responded to the pretraining survey.

[c]AOSL: all other service lines.

[d]Participants could endorse multiple races or ethnicities. Therefore, race and ethnicity were analyzed separately.

[e]Self-reported disabilities included hearing impairment (33/805, 4.1%), visual impairment (31/805, 3.9%), mobility impairment (17/805, 2.1%), learning disability (12/805, 1.5%), and any other disability or impairment (58/805, 7.2%). Overall, 59/805 (7.3%) stated that they preferred not to report their disability status.

Table 3. Participants' disciplines (N=1107).

Disciplines	Values, n (%)
Social Workers	256 (23.1)
Nurses	240 (21.7)
Psychologists	232 (21.0)
Peer Support Specialists	61 (5.5)
Chaplains	41 (3.7)
Medical Support Assistants	40 (3.6)
Psychiatrists	39 (3.5)
Medical Doctors	33 (3)
Audiologists	31 (2.8)
Other disciplines[a]	134 (12.1)

[a]Other disciplines (reported as n out of 1107): Vocational Rehabilitation Specialist (19/1107, 1.72%), Pharmacists (15/1107, 1.36%), Licensed MH Counselor (12/1107, 1.08%), Psychology Technicians (9/1107, 0.81%), Physician Assistants (8/1107, 0.72%), Administrative Personnel (7/1107, 0.63%), Recreational Therapist (7/1107, 0.63%), Program Support Specialists (6/1107, 0.54%), Communication Specialists (5/1147, 0.45%), Dietitians (5/1107, 0.45%), Occupational Therapists (5/1107, 0.45%), Patient Advocates (5/1107, 0.45%), Coaches (4/1107, 0.4%), Medical Technicians (4/1107, 0.4%), Physical Therapists (4/1107, 0.36%), Addiction Therapists (3/1107, 0.27%), Librarians (3/1107, 0.27%), Volunteers (3/1107, 0.27%), Music Therapists (2/1107, 0.18%), Program Coordinators (2/1107, 0.18%), Learning Resource/Education Staff (2/1107, 0.18%), Dental Assistant (1/1107, 0.09%), Outreach Specialist (1/1107, 0.09%), Police Officer (1/1107, 0.09%), and Research Assistant (1/1107, 0.09%).

Technology-Related Participant Characteristics

Table 4 summarizes the participants' ownership, attitudes, and skills using VA MH apps broken down by staff type at baseline. Both groups reported a high degree of personal smartphone ownership (828/862, 96.06%). The vast majority (753/830, 90.72%) were either *very comfortable* or *quite comfortable* using apps, with most spending at least 1 hour per day using them (719/844, 85.19%). Most participants reported being confident in their skills to show veterans how to *download* apps (655/841, 77.88%) and showing veterans how to *use* VA MH apps (593/839, 70.68%) at baseline.

The 2 staff types differed in 3 ways related to technology. First, MHSL staff (204/602, 33.89%) were more likely than AOSL staff (51/260 19.62%) to have government-issued smartphones, $\chi^2_1=17.75$, $P<.001$. Second, MHSL staff had higher mean percentage correct scores on the knowledge check at pretraining (mean 82.1% correct, SD 16%) compared with the AOSL staff, who averaged 76.5% correct (SD 15.22%), $t_{829}=-4.74$, $P<.001$, Cohen $d=-0.36$. Finally, MHSL staff (550/574, 95.82%) were more likely to report that they *definitely* or *probably* could think of at least one veteran who they planned to use apps with compared with AOSL staff (215/252, 85.32%), $\chi^2_2=29.55$, $P<.001$.

Table 4. Participants' technological skills and intentions for Veterans Affairs mental health (VA MH) apps use at baseline (N=862).

	MHSL[a] staff (n=602[b]), n (%)	AOSL[c] staff (n=260[b]), n (%)	Statistical results	
			χ^2 (*df*)	*P* value
Owns a personal smartphone			0.7 (1)	.39
Yes	576 (95.7)	252 (96.9)		
No	26 (4.3)	8 (3.1)		
Has a government-issued smartphone			17.8 (1)	<.001
Yes	204 (33.9)	51 (19.6)		
No	398 (66.1)	209 (80.4)		
Comfort using mobile devices			2.9 (2)	.24
Not at all or A little comfortable	45 (8.1)	30 (11.9)		
Quite comfortable	204 (35.4)	87 (34.4)		
Very comfortable	326 (56.5)	136 (53.8)		
Hour per day spent using apps			6.0 (4)	.20
I do not use apps	13 (2.2)	4 (1.6)		
<1 hour per day	76 (12.9)	32 (12.5)		
1-2 hour per day	238 (40.5)	84 (32.8)		
3-4 hour per day	168 (28.6)	88 (34.4)		
>4 hour per day	93 (15.8)	48 (18.7)		
Skills to demonstrate how to *download* apps			0.8 (4)	.93
Strongly agree	204 (34.9)	83 (32.3)		
Agree	251 (43)	117 (45.5)		
Neither Agree nor Disagree	71 (12.2)	30 (11.7)		
Disagree	49 (8.4)	22 (8.6)		
Strongly disagree	9 (1.5)	5 (1.9)		
Skills to demonstrate how to *use* apps			1.5 (4)	.83
Strongly agree	162 (27.8)	69 (27)		
Agree	245 (42.1)	117 (45.7)		
Neither Agree or Disagree	95 (16.3)	36 (14.1)		
Disagree	72 (12.3)	29 (11.3)		
Strongly disagree	9 (1.5)	5 (2)		
Recommend VA MH apps to veterans (BI[d] item 1)			11.1 (2)	.004
Definitely will	356 (61.7)	143 (56.3)		
Probably will	211 (36.6)	96 (37.8)		
Probably will not or Definitely will not	10 (1.7)	15 (5.9)		
Think of at least 1 Veteran I plan to use apps with (BI item 2)			29.6 (2)	<.001
Definitely will	335 (58.4)	121 (48)		
Probably will	215 (37.5)	94 (37.3)		
Probably will not or Definitely will not	24 (4.2)	37 (14.7)		
Encourage other VA staff to use apps (BI item 3)			1.1 (2)	.58
Definitely will	307 (53.1)	142 (55.9)		
Probably will	237 (41)	101 (39.8)		

	MHSL[a] staff (n=602[b]), n (%)	AOSL[c] staff (n=260[b]), n (%)	Statistical results	
			χ^2 (*df*)	*P* value
Probably will not or Definitely will not	34 (5.9)	11 (4.3)		

[a]MHSL: mental health service line.

[b]Of the 1110 participants enrolled in the project, 862 (77.66%) responded to the pretraining survey.

[c]AOSL: all other service lines.

[d]BI: behavioral intentions.

Training Program Reach

The program exceeded the goals set for participant recruitment. Across the 19 sites, 1110 staff attended the training program with an average of 58.4 participants per site (SD 36.49, median [IQR] 51). This was 247% of the minimum goal of 25 participants from 18 locations. All sites met the goal of recruiting both MHSL and AOSL staff members to participate in the program.

Training Satisfaction

Table 5 summarizes the data from the posttraining survey, which included measures of training satisfaction. There were no significant differences based on staff type for training satisfaction, χ^2_3=7.93, *P*=.047 or for willingness to recommend training to others, χ^2_3=7.71, *P*=.052. The vast majority (961/1024, 93.85%) of participants were satisfied with the training and 92.44% (941/1018) would recommend the training to others.

Table 5. Participants' training satisfaction, basic technology skills, and intentions for Veterans Affairs mental health (VA MH) apps use from the posttraining survey (N=1023).

	MHSL[a] staff (n=686[b]), n (%)	AOSL[c] staff (n=337[b]) , n (%)	Statistical results	
			χ^2 (*df*)	*P* value
Satisfaction with training			7.9 (3)	.047
Strongly agree	393 (57.3)	203 (60.1)		
Agree	241 (35.1)	124 (36.7)		
Neither agree nor disagree	35 (5.1)	9 (2.7)		
Disagree or strongly disagree	17 (2.5)	2 (0.6)		
Would recommend training			7.7 (3)	.052
Strongly agree	421 (61.5)	206 (61.7)		
Agree	201 (29.4)	113 (33.8)		
Neither agree nor disagree	43 (6.3)	11 (3.3)		
Disagree or strongly disagree	19 (2.8)	4 (1.2)		
Skills to demonstrate how to *download* apps			3.0 (4)	.55
Strongly agree	384 (56.2)	182 (54.3)		
Agree	278 (40.7)	136 (40.6)		
Neither agree or disagree	13 (1.9)	12 (3.6)		
Disagree	4 (0.6)	3 (0.9)		
Strongly disagree	4 (0.6)	2 (0.6)		
Skills to demonstrate how to *use* apps			1.7 (4)	.80
Strongly agree	344 (50.1)	160 (47.5)		
A	314 (45.8)	161 (47.8)		
Neither agree or disagree	19 (2.8)	10 (3)		
D	4 (0.6)	4 (1.2)		
Strongly disagree	5 (0.7)	2 (0.6)		
Recommend apps to veterans (BI[d] item 1)			11.8 (2)	.003
Definitely will	577 (84.5)	255 (75.9)		
Probably will	96 (14.1)	76 (22.6)		
Probably will not or Definitely will not	10 (1.5)	5 (1.5)		
Think of at least 1 Veteran I plan to use VA MH apps with (BI item 2)			29.3 (2)	<.001
Definitely will	556 (81.3)	225 (67)		
Probably will	116 (17)	92 (27.4)		
Probably will not or Definitely will not	12 (1.7)	19 (5.6)		
Encourage other VA staff to use VA MH apps (BI item 3)			1.7 (2)	.43
Definitely will	486 (71.3)	235 (70.4)		
Probably will	178 (26.1)	94 (28.1)		
Probably will not or Definitely will not	18 (2.6)	5 (1.5)		

[a]MHSL: mental health service line.

[b]Of the 1110 participants enrolled in the project, 1023 (92.2%) responded to the posttraining survey.

[c]AOSL: all other service lines.

[d]BI: behavioral intentions.

Training Effectiveness

Knowledge Check

For knowledge check scores, a generalized linear model was used to evaluate the between-group effect for staff type (MHSL vs AOSL staff) and the repeated measures effect of time (pre- and posttraining) as well as their interaction. Controlling for family-wise error, the interaction was not statistically significant, $F_{1,758}$=7.89, *P*=.005, partial η^2=0.010. The main effect for staff type was statistically significant, $F_{1,773}$=25.77, *P*<.001, partial η^2=0.033 (small effect). Collapsing across pre- and posttraining time points, MHSL staff had significantly higher mean scores (mean 87.2% correct, SD 12.56%) compared with AOSL staff (mean 83.1% correct, SD 7.64%). The main effect for time was statistically significant. When collapsing across staff types, the mean knowledge check scores significantly increased from pretraining (mean 80.8%, SD 15.77%) to posttraining (mean 91.1% correct, SD 9.57%), $F_{1,758}$=318.79, *P*<.001, η^2=0.296 (large effect).

Basic Skills

Wilcoxon signed-rank tests were conducted to assess the pre- to posttraining changes on the 2 items that asked participants to rate their confidence in their ability to demonstrate how to *download* and *use* apps with veterans. There were statistically significant increases in participants' confidence ratings for both how to *download* apps, *z*=−13.86, *P*<.001 and for how to *use* apps, *z*=−15.13, *P*<.001. For both items, the *strongly disagree* and *disagree* response categories were collapsed because there were only a few responses in each category. The percentage of staff who *agreed* or *strongly agreed* to having the knowledge and skills to show veterans how to *download* apps (655/841, 77.33%) and how to *use* apps (593/839, 70.68%) at baseline increased to 96.27% (980/1018) for *downloading* apps and 95.70% (979/1023) for *using* apps following training. There were no significant differences on either item based on staff type (download apps: χ^2_4=3.04, *P*=.55; use apps: χ^2_4=1.67, *P*=.80).

Behavioral Intentions

Wilcoxon signed-rank tests were used to compare pre- and posttraining changes for the three items measuring participants' behavioral intentions to use mobile apps with veterans. Statistically significant increases at posttraining compared with pretraining for intentions to *use* VA MH apps with veterans were found for all 3 items (see the Methods section > Training Effectiveness > Behavioral Intentions section of this report), item 1, *z*=−11.77, *P*<.001; item 2, *z*=−11.85, *P*<.001; and item 3, *z*=−9.94, *P*<.001. Similar to the findings on the pretraining survey for the second Behavioral Intentions item, there were statistically significant differences at posttraining between MHSL and AOSL staff on the ability to identify at least one veteran with whom they plan to use apps, χ^2_2=29.32, *P*<.001 (Table 5)**.** Specifically, compared with AOSL staff (225/336, 66.96%), MHSL staff (556/684, 81.29%) were significantly more likely to report that they *definitely* could think of at least one veteran with whom they plan to use apps following training. However, looking more broadly at this item after training, nearly all staff (989/1020, 96.96%) reported that they could *definitely* or *probably* identify a veteran with whom they planned to use apps.

Discussion

Principal Findings

A training program was implemented to facilitate VA staff's use of VA MH apps with veterans seen in mental health and non–mental health settings. The number of participants (N=1110) exceeded the project's recruitment target of 25 per site, suggesting a higher-than-anticipated staff interest in this training. Participants represented 34 disciplines, and 33% (366/1109) of participants worked in settings other than mental health, reflecting the relevance of VA MH apps in diverse settings across the health care system. Given that most VA patients do not use mental health services, with only 22% of those diagnosed with a mental illness receiving psychotherapy and over half dropping out by the second session [31], the level of participation by staff outside mental health settings is promising for veterans who might otherwise not be reached. Furthermore, the training was well received: over 90% of staff members reported being satisfied with the training and said that they would recommend it to others.

In addition to the successful recruitment and engagement of participants, measures of the effectiveness of training indicated significant increases in (1) key knowledge about VA MH apps, (2) confidence in the ability to use VA MH apps with veterans, and (3) behavioral intentions to use VA MH apps with veterans. AOSL and MHSL staff knowledge was high at pretraining, particularly for MHSL staff, but increased significantly from pre- to posttraining for both groups. In terms of skills, about 96% of participants reported being comfortable showing veterans how to both download (980/1018, 96.27%) and use (979/1023, 95.70%) VA MH apps after the training. Finally, although behavioral intentions to use VA MH apps with veterans at pretraining were high, they also increased following training. Nearly all participants endorsed intentions to recommend apps to veterans and could think of at least one specific veteran to whom they could recommend an app.

To summarize, both AOSL and MHSL staff made significant gains in their knowledge and intention to use VA MH apps. Both before and directly after the training, AOSL staff had less knowledge of VA MH apps and were less likely than their MHSL counterparts to be able to think of a Veteran to whom they could introduce a VA MH app. However, the differences between staff types were small, and both groups indicated an increased likelihood of using MH apps with veterans after training.

Limitations

This quality improvement project relied on voluntary participation of VA employees at facilities that attained leadership support for the implementation of VA MH apps. The extent to which the positive outcomes were observed could be attributed to this being a voluntary sample of early adopters. Furthermore, because we focused on VA Medical Centers, it is unclear how these findings might generalize to other health care

settings. Research is needed to understand the potential impacts of introducing VA MH apps outside mental health contexts. For example, future studies could further explore whether VA MH apps help individuals self-refer to mental health services.

Next, there were concerns about ceiling effects at baseline for some key metrics (staff members' knowledge and confidence in their skills to use VA apps). However, significantly more staff members met the goal criteria after the training. In addition, we did not deliver the same exact training across sites, as we modified our training based on feedback from each site. We also adapted to unanticipated circumstances by switching from an in-person to a live, web-based training model owing to COVID-19 restrictions. The effects of these two factors on outcomes are unclear, but they are consistent with quality improvement work.

Finally, the self-reported survey data presented in this paper were collected before and immediately after the training and did not directly address how attitudes and beliefs were translated into action. Participants reported intentions to act, which, according to the Theory of Planned Behavior [28], can inform actual behavior. This limitation will also be addressed in a subsequent manuscript that will discuss posttraining implementation efforts and, specifically, participants' adoption of VA MH apps.

Comparison With Prior Work

To our knowledge, this is the first VA MH app training program to demonstrate that lay health care staff are willing and able to be trained to introduce patient-facing mental health apps in a variety of VA settings. Armstrong et al [17] trained DoD and VA behavioral health providers on a similar suite of mental health applications. Although the authors used different measures of satisfaction, skill acquisition, and behavioral intentions, they similarly showed that participants rated the program well, had skill improvements, and intended to adopt practices following training. The results were promising: most of the 760 clinicians trained over 3 years demonstrated knowledge gains and reported immediately following the training that they believed the training was helpful and intended to apply what they learned in their clinical work. The current project adds to the extant literature by including nonmental health staff and training more staff in a shorter timeframe (1 year vs 3 years) while also increasing staff's knowledge and behavioral intentions. Importantly, this project coincided with a global pandemic and a concurrent increase in telehealth [32], which may have generated increased interest in digital health technology, such as mental health apps, that can be used remotely.

Conclusions

The VA MH apps training far exceeded the initial goals for staff recruitment and training in all three metrics: program reach, participant satisfaction, and training effectiveness. Over 33% (366/1109) of the participants came from AOSLs, which suggests the utility of VA MH apps across the health care system. Veterans' access to mental health resources is not limited to traditional referrals to mental health providers. The pathway to getting the needed help may be self-guided or initiated by a peer, chaplain, or other VA employee. By training an occupationally diverse cadre of VA staff to be capable and willing to share mental health resources with veterans, we can potentially reach a wide range of veterans who can benefit from these tools.

Acknowledgments

This study was supported by the United States Department of Defense and the Veterans Affairs Joint Incentive Fund # 19190. The REDCap (Research Electronic Data Capture) platform is developed and operated by Stanford Medicine Research IT team. The REDCap platform services at Stanford are subsidized by (1) Stanford School of Medicine Research Office and (2) the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through grant UL1 TR001085 . The authors would like to acknowledge former project staff Sharon Huang, Adrienne Heinz, and Tara Zaugg; Defense Health Agency Connected Health Branch collaborators Christina Armstrong and Julie Kinn; and the VA National Center for PTSD Mobile Apps team, which is responsible for the design and development of the Veterans Affairs (VA) mental health (MH) apps discussed in this paper.

All the authors were affiliated with the National Center for PTSD, Training, and Dissemination Division at the time of the project. PMV is currently affiliated with the VA Office of Mental Health and Suicide Prevention, KJ is currently affiliated with the National Training Division Education Service at the Veterans Benefits Administration, TJA is currently affiliated with the VA Peninsula Vet Center, and LG is currently affiliated with the VA Pacific Islands Health Care System.

Conflicts of Interest

None declared.

References

1. Gould CE, Kok BC, Ma VK, Zapata AM, Owen JE, Kuhn E. Veterans Affairs and the Department of Defense mental health apps: a systematic literature review. Psychol Serv 2019 May;16(2):196-207. [doi: 10.1037/ser0000289] [Medline: 30431306]
2. Owen JE, Kuhn E, Jaworski BK, McGee-Vincent P, Juhasz K, Hoffman JE, et al. VA mobile apps for PTSD and related problems: public health resources for veterans and those who care for them. Mhealth 2018 Jul;4:28 [FREE Full text] [doi: 10.21037/mhealth.2018.05.07] [Medline: 30148141]

3. McGee-Vincent P, Mackintosh M, Jaworski B, Blasko K. Self-care during the COVID-19 pandemic: deploying mobile mental health resources to support service members, veterans, and the public. In: Proceedings of the TMS Proceedings 2021. 2021 Presented at: TMS Proceedings 2021; Nov 3-5, 2021; Virtual. [doi: 10.1037/TMS0000111]
4. Reger GM, Browne KC, Campellone TR, Simons C, Kuhn E, Fortney JC, et al. Barriers and facilitators to mobile application use during PTSD treatment: clinician adoption of PE coach. Psychiatr Rehabil J 2017 Dec;48(6):510-517. [doi: 10.1037/pro0000153]
5. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 May 01;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
6. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 2;2(1):118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
7. Section 508 of the Rehabilitation Act. Federal Communications Commission. URL: https://www.fcc.gov/general/section-508-rehabilitation-act [accessed 2022-12-15]
8. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]
9. Koffel E, Kuhn E, Petsoulis N, Erbes CR, Anders S, Hoffman JE, et al. A randomized controlled pilot study of CBT-I Coach: feasibility, acceptability, and potential impact of a mobile phone application for patients in cognitive behavioral therapy for insomnia. Health Informatics J 2018 Mar;24(1):3-13 [FREE Full text] [doi: 10.1177/1460458216656472] [Medline: 27354394]
10. Babson KA, Ramo DE, Baldini L, Vandrey R, Bonn-Miller MO. Mobile app-delivered cognitive behavioral therapy for insomnia: feasibility and initial efficacy among veterans with cannabis use disorders. JMIR Res Protoc 2015 Jul 17;4(3):e87 [FREE Full text] [doi: 10.2196/resprot.3852] [Medline: 26187404]
11. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: a pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 May;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
12. McGee-Vincent P, Juhasz K, Jamison AL, Avery TJ, Owen JE, Jaworski BK, et al. Mobile mental health apps from the National Center for PTSD: digital self-management tools for co-occurring disorders. J Dual Diagn 2021 Jun 21;17(3):181-192. [doi: 10.1080/15504263.2021.1939919] [Medline: 34152258]
13. Hallenbeck HW, Jaworski BK, Wielgosz J, Kuhn E, Ramsey KM, Taylor K, et al. PTSD Coach version 3.1: a closer look at the reach, use, and potential impact of this updated mobile health app in the general public. JMIR Ment Health 2022 Mar 29;9(3):e34744 [FREE Full text] [doi: 10.2196/34744] [Medline: 35348458]
14. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD Coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
15. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016 Jan;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
16. Reger GM, Harned M, Stevens ES, Porter S, Nguyen J, Norr AM. Mobile applications may be the future of veteran mental health support but do veterans know yet? A survey of app knowledge and use. Psychol Serv 2022 Aug 03;19(3):480-487. [doi: 10.1037/ser0000562] [Medline: 34081527]
17. Armstrong CM, Ciulla RP, Edwards-Stewart A, Hoyt T, Bush N. Best practices of mobile health in clinical care: the development and evaluation of a competency-based provider training program. Professional Psychol Res Pract 2018 Oct;49(5-6):355-363. [doi: 10.1037/pro0000194]
18. Gupta N, Bhalla IP, Rosenheck RA. Treatment of veterans with psychiatric diagnoses nationally in the Veterans Health Administration: a comparison of service delivery by mental health specialists and other providers. Adm Policy Ment Health 2019 May 31;46(3):380-390. [doi: 10.1007/s10488-018-00920-z] [Medline: 30706163]
19. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
20. Ritchie MJ, Parker LE, Kirchner JE. From novice to expert: a qualitative study of implementation facilitation skills. Implement Sci Commun 2020;1:25 [FREE Full text] [doi: 10.1186/s43058-020-00006-8] [Medline: 32885184]
21. Brown KG, Gerhardt MW. Formative evaluation: an integrative practice model and case study. Personnel Psychol 2002 Dec;55(4):951-983. [doi: 10.1111/j.1744-6570.2002.tb00137.x]
22. Park MW. Promoting societal change: integrating traumatic stress research, practice and policy for vulnerable populations. In: Proceedings of the ISTSS 34th Annual Meeting. 2018 Presented at: ISTSS 34th Annual Meeting; Nov 8 -10, 2018; Washington, DC, USA URL: https://istss.org/ISTSS_Main/media/Documents/ISTSS-18-FP_FNL1.pdf

23. McGee-Vincent P, Juhasz K. Integrating National Center for PTSD mobile apps into veteran mental health care. National Center for PTSD Continuing Education. 2020 May 14. URL: https://www.ptsd.va.gov/professional/continuing_ed/integrating_mobileapps_tx.asp [accessed 2022-07-25]
24. Kuhn E, McGee-Vincent P. Using mobile apps in mental health practice. In: Technology and Mental Health: A Clinician's Guide to Improving Outcomes. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2020.
25. Schwartz A, Wilson J, Boden S, Moore T, Bradbury T, Fletcher N. Managing resident workforce and education during the COVID-19 pandemic: evolving strategies and lessons learned. JB JS Open Access 2020;5(2):e0045 [FREE Full text] [doi: 10.2106/JBJS.OA.20.00045] [Medline: 33117955]
26. Ritchie M, Dollar K, Miller C, Smith J, Oliver K, Kim B, et al. Using implementation facilitation to improve healthcare (version 3). Veterans Health Administration, Behavioral Health Quality Enhancement Research Initiative (QUERI). 2020. URL: https://www.queri.research.va.gov/tools/Facilitation-Manual.pdf [accessed 2020-11-02]
27. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 01;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
28. Ajzen I. From intentions to actions: a theory of planned behavior. In: Action Control. Berlin, Heidelberg: Springer; 1985.
29. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Cambridge, Massachusetts, United States: Academic Press; 1969.
30. Vickerstaff V, Omar RZ, Ambler G. Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes. BMC Med Res Methodol 2019 Jun 21;19(1):129 [FREE Full text] [doi: 10.1186/s12874-019-0754-4] [Medline: 31226934]
31. Cully JA, Tolpin L, Henderson L, Jimenez D, Kunik ME, Petersen LA. Psychotherapy in the Veterans Health Administration: missed opportunities? Psychol Serv 2008 Nov;5(4):320-331 [FREE Full text] [doi: 10.1037/a0013719] [Medline: 25177213]
32. Connolly SL, Kuhn E, Possemato K, Torous J. Digital clinics and mobile technology implementation for mental health care. Curr Psychiatry Rep 2021 May 07;23(7):38 [FREE Full text] [doi: 10.1007/s11920-021-01254-8] [Medline: 33961135]

Abbreviations

AIMS: Anger & Irritability Management Skills
AOSL: all other service lines
DoD: Department of Defense
FE: formative evaluation
MHSL: mental health service line
REDCap: Research Electronic Data Capture
VA MH apps: Veterans Affairs mental health apps
VA: Department of Veterans Affairs

Edited by J Torous; submitted 08.08.22; peer-reviewed by S Connolly, J Torous; comments to author 09.10.22; revised version received 19.10.22; accepted 20.10.22; published 12.01.23.

Please cite as:
McGee-Vincent P, Mackintosh MA, Jamison AL, Juhasz K, Becket-Davenport C, Bosch J, Avery TJ, Glamb L, Hampole S
Training Staff Across the Veterans Affairs Health Care System to Use Mobile Mental Health Apps: A National Quality Improvement Project
JMIR Ment Health 2023;10:e41773
URL: https://mental.jmir.org/2023/1/e41773
doi:10.2196/41773
PMID:36633895

Original Paper

Exposure Versus Cognitive Restructuring Techniques in Brief Internet-Based Cognitive Behavioral Treatment for Arabic-Speaking People With Posttraumatic Stress Disorder: Randomized Clinical Trial

Jana Stein[1,2], MSc; Max Vöhringer[2], MSc; Birgit Wagner[3], PhD; Nadine Stammel[1], PhD; Yuriy Nesterko[2,4], PhD; Maria Böttche[1*], PhD; Christine Knaevelsrud[1*], PhD

[1]Clinical Psychological Intervention, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

[2]Department for Transcultural and Traumatic Stress Studies, Center ÜBERLEBEN, Berlin, Germany

[3]Clinical Psychology and Psychotherapy, Medical School Berlin, Berlin, Germany

[4]Department of Medical Psychology and Medical Sociology, University of Leipzig, Leipzig, Germany

[*]these authors contributed equally

Corresponding Author:
Jana Stein, MSc
Clinical Psychological Intervention
Department of Education and Psychology
Freie Universität Berlin
Schloßstr. 1
Berlin, 12163
Germany
Phone: 49 30 303906 76
Email: j.stein@ueberleben.org

Abstract

Background: Cognitive behavioral interventions delivered via the internet are demonstrably efficacious treatment options for posttraumatic stress disorder (PTSD) in underserved, Arabic-speaking populations. However, the role of specific treatment components remains unclear, particularly in conflict-affected areas of the Middle East and North Africa.

Objective: This study aims to evaluate 2 brief internet-based treatments in terms of efficacy, including change in PTSD symptom severity during treatment. Both treatments were developed in line with Interapy, an internet-based, therapist-assisted cognitive behavioral therapy protocol for PTSD and adapted to the specific research question. The first treatment comprised self-confrontation and social sharing (exposure treatment; 6 sessions); the second comprised cognitive restructuring and social sharing (cognitive restructuring treatment; 6 sessions). The 2 treatments were compared with each other and with a waitlist control group.

Methods: In total, 365 Arabic-speaking participants from the Middle East and North Africa (mean age 25.49, SD 6.68 y) with PTSD were allocated to cognitive restructuring treatment (n=118, 32.3%), exposure treatment (n=122, 33.4%), or a waitlist control group (n=125, 34.2%) between February 2021 and December 2022. PTSD symptom severity, posttraumatic maladaptive cognitions, anxiety, depressive and somatoform symptom severity, and quality of life were assessed via self-report at baseline and after treatment or waiting time. PTSD symptom severity was also measured throughout treatment or waiting time. Treatment satisfaction was assessed after treatment completion. Treatment use and satisfaction were compared between the 2 treatment conditions using appropriate statistical tests (eg, chi-square and Welch tests). Multiple imputation was performed to address missing data and evaluate treatment-associated changes. These changes were analyzed using multigroup change modeling in the completer and intention-to-treat samples.

Results: Overall, 200 (N=240, 83.3%) participants started any of the treatments, of whom 123 (61.5%) completed the treatment. Treatment condition was not significantly associated with the proportion of participants who started versus did not start treatment (P=.20) or with treatment completion versus treatment dropout (P=.71). High treatment satisfaction was reported, with no significant differences between the treatment conditions (P=.48). In both treatment conditions, PTSD, anxiety, depressive and somatoform symptom severity, and posttraumatic maladaptive cognitions decreased, and quality of life improved significantly from baseline to the posttreatment time point (P≤.001 in all cases). Compared with the baseline assessment, overall PTSD symptom

severity decreased significantly after 4 sessions in both treatment conditions (P<.001). Moreover, both treatment conditions were significantly superior to the waitlist control group regarding overall PTSD symptom severity (P<.001) and most other comorbid mental health symptoms (P<.001 to P=.03). Differences between the 2 conditions in the magnitude of change for all outcome measures were nonsignificant.

Conclusions: Internet-based cognitive behavioral treatments for PTSD focusing primarily on either self-confrontation or cognitive restructuring are applicable and efficacious for Arabic-speaking participants.

Trial Registration: German Clinical Trials Register DRKS00010245; https://drks.de/search/de/trial/DRKS00010245

(JMIR Ment Health 2023;10:e48689) doi:10.2196/48689

KEYWORDS

posttraumatic stress; posttraumatic stress disorder; PTSD; Middle East; North Africa; Arabic; Arabic-speaking people; internet; internet-based; exposure; cognitive restructuring; randomized clinical trial; RCT; cognitive behavioral therapy; CBT; cognitive behavioral treatment

Introduction

Background

The Middle East and North Africa region is both geographically and culturally diverse. Ongoing civil wars, local conflicts, political instability, economic insecurity (ie, high unemployment rates), and high levels of displacement [1,2] have taken a toll on the mental health of the civilian population in this region [3,4]. A meta-analysis of prevalence rates of mental disorders in the Eastern Mediterranean region identified depression, generalized anxiety disorder, and posttraumatic stress disorder (PTSD) as the most prevalent disorders, with pooled current prevalence rates of 20.5%, 10.3%, and 9.5%, respectively [5]. However, despite the high prevalence of mental disorders, many individuals do not receive adequate treatment [6,7], partly because of the dearth of available professionals in the region. For instance, the World Health Organization [8] reported that only 0.7 psychologists per 100,000 population serve the Eastern Mediterranean region compared with 5.4 psychologists in the Americas.

Interventions provided via the internet may offer a solution to bridge the gap between the high demand for mental health services and the limited access to such support. By combining the advantage of high availability with the independence of the therapist's location, internet-based interventions represent a promising opportunity to provide support in regions with low access to psychotherapeutic help (eg, in regions shattered by conflict [9]). Moreover, the easy accessibility and greater visual anonymity of internet-based interventions often facilitate their use by people who have been exposed to highly stigmatizing traumatic events, who may fear prejudgment when seeking help for mental health problems, or when mobility is limited (ie, women not being permitted to leave the house without a male attendant). A survey of 503 Arabic-speaking people found that 73% were willing to try an intervention for anxiety and depression delivered via the internet [10]. Furthermore, in a previous study, >6000 Arabic-speaking people completed a screening process to participate in internet-based interventions for the treatment of depression or PTSD [11], suggesting that Arabic-speaking people have a strong interest in psychological treatments delivered via the internet.

In addition, several meta-analyses have pointed to the efficacy of internet-delivered cognitive behavioral interventions for the treatment of PTSD [12]. Cognitive behavioral interventions delivered through the internet seem to be superior to inactive control groups [12] and noninferior to cognitive behavioral treatments delivered face-to-face [13]. A meta-analysis focusing specifically on the efficacy of cognitive behavioral therapy for Arabic-speaking people with PTSD, anxiety, or depression found large effect sizes (ie, PTSD: g=2.08; depression: g=1.26; anxiety: g=1.44), and a reduction in psychopathological symptoms was reported for all included internet-based cognitive behavioral interventions (n=5 of 9 studies) [14]. Knaevelsrud et al [15] similarly found high levels of satisfaction with an internet-based trauma-focused cognitive behavioral intervention among traumatized Arabic-speaking people. Thus, cognitive behavioral interventions delivered via the internet appear to be accepted and significantly reduce distressing symptoms in different populations, including Arabic-speaking people with PTSD [11,15].

Although trauma-focused cognitive behavioral approaches are superior to cognitive behavioral approaches without a trauma focus [16] and are, therefore, the treatment of first choice for adults with a diagnosis of PTSD [17], and trauma-focused cognitive behavioral approaches delivered via the internet show promising results, it is hugely important to examine the differential effects of specific cognitive behavioral treatment components—particularly when delivered via the internet—to provide the best possible care for individuals with PTSD. In the face-to-face setting, a number of studies have investigated the specific effect of cognitive methods (ie, cognitive restructuring [CR]) on PTSD symptoms compared with exposure-based methods [18-20], but the superiority of exposure methods, cognitive methods, or a combination of the 2 could not be clearly demonstrated [20-22]. Studies on the efficacy of internet-delivered psychotherapeutic interventions with a specific focus on either exposure-based techniques [11] or cognitive methods [23,24] have revealed significant improvements in PTSD and comorbid mental health symptoms. Although exposure-based techniques delivered via the internet have been successfully implemented for Arabic-speaking people with PTSD (overall PTSD symptom improvement during treatment: d=1.13) [11], exposure might not be a suitable treatment option for all individuals with PTSD as some may be

unwilling to confront the traumatic event in detail and may drop out of the intervention before treatment gains become apparent. In particular, for people in Arabic-speaking cultures who have experienced any form of sexual violence, going through the traumatic event in detail may be a huge burden as this type of trauma is likely to be associated with great shame, loss of honor, or feelings of guilt [25]. For internet-delivered interventions combining exposure and cognitive methods for Arabic-speaking people with PTSD, dropout rates of approximately 37% have been reported [26]. Similar dropout rates were found when only providing exposure treatment for this population [11], highlighting the need for additional treatment options without a focus on exposure.

In summary, studies conducted in face-to-face psychotherapy settings have proven that both exposure and cognitive methods have beneficial effects on PTSD and comorbid mental health symptoms [19-22]. However, little research has investigated the differential effects of cognitive and exposure-based treatments by comparing the 2 treatment techniques with each other directly and with a passive control group in internet-based settings. To the best of our knowledge, no study has addressed this topic in Arabic-speaking populations. Therefore, addressing this issue is of considerable practical relevance, especially in areas with limited access to treatment.

Study Aims

The aim of this study was to evaluate 2 brief internet-based treatments—1 including CR as the main treatment component and the other including exposure—for Arabic-speaking participants with PTSD. Specifically, we sought to examine the association between the 2 treatment conditions and treatment use investigating the proportion of individuals who started treatment and the proportion who dropped out during treatment as well as treatment duration. We expected that the proportion of treatment starters and dropouts as well as treatment duration would not differ between the 2 treatment conditions. Furthermore, we compared completers' treatment satisfaction between the 2 treatment conditions and, again, did not expect any differences between conditions. Finally, changes in posttraumatic stress symptom severity, posttraumatic cognitions, anxiety, depressive and somatoform symptom severity, and quality of life during the 2 treatments were examined and compared with those of a waitlist control group. On the basis of previous research, we assumed that both treatments would lead to significant improvements in all treatment outcomes between the baseline assessment and posttreatment time point. We expected that the 2 treatments would lead to similar changes in terms of treatment outcomes and would outperform the waitlist control group.

Methods

Trial Information

The study was administered by a psychosocial center for the treatment of war and torture survivors in cooperation with the Freie Universität Berlin and the Medical School Berlin, Germany. The study was preregistered at the German Clinical Trials Register (trial DRKS00010245).

Participants

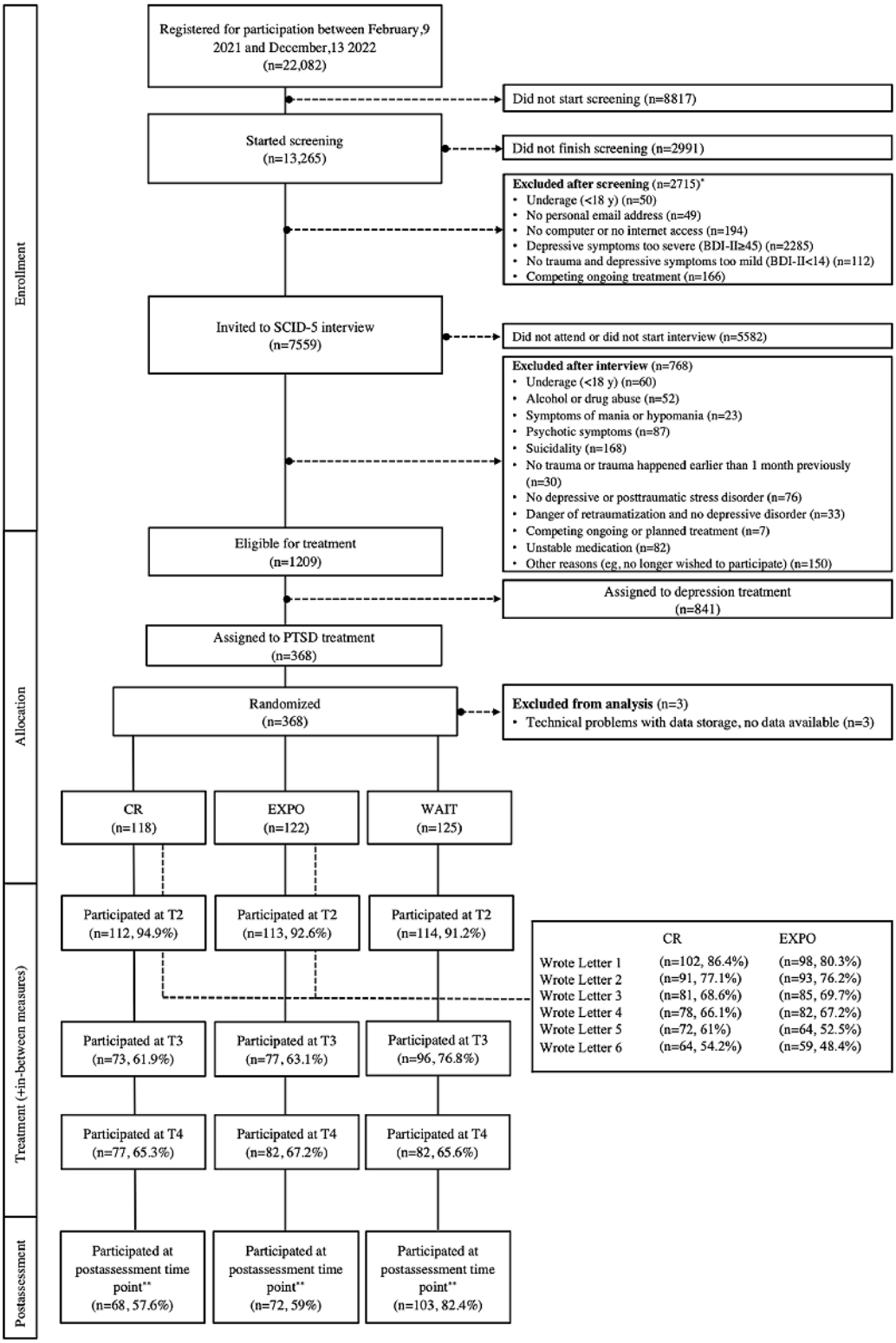
This study included Arabic-speaking adults from different countries who were seeking help via the internet for posttraumatic stress and depressive symptoms. As an inclusion criterion, all participants were required to be able to speak, read, and write standard Arabic. Individuals were excluded if they self-reported any of the following in the screening battery: age of <18 years, no private email address or access to a computer and internet, simultaneous psychotherapeutic treatment elsewhere or plans for psychotherapeutic treatment within the next 4 weeks, or severe depressive symptoms (Beck Depression Inventory–II of ≥45). After successfully passing the screening battery, participants underwent a clinical interview in which interviewers checked whether participants met the diagnostic criteria for a depressive disorder or PTSD according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), assessed using the Structured Clinical Interview for the DSM-5 (SCID-5 Clinical Version) [27] as a requirement for participation in any of the offered treatments. If the diagnostic criteria were not met, the participants were excluded. In the interview, participants were further screened for symptoms of mania or hypomania, psychotic experiences, risk of suicide, drug and alcohol use, and current risk of retraumatization (ie, still living with the perpetrator). We further excluded participants who reported psychotic tendencies, manic or hypomanic episodes, a high risk of suicide (ie, serious suicide attempts within the last 3 y or a current intent), dependency on or abuse of drugs or alcohol with current use, or a current danger of retraumatization. In addition, interviewers checked whether any participants receiving psychopharmacological treatment were on a stable dose and whether participants had completed our treatment program within the previous months. Again, participants who failed to meet these criteria were excluded.

Procedure

Recruitment took place between February 9, 2021, and December 13, 2022. Participants were recruited through the program's website (Ilajnafsy [ ], Arabic for *psychotherapy* [28]), word-of-mouth recommendation, and social media (ie, Facebook). Applicants could register on the website for free. The registration page contained information on data security and the terms of participation. Participants were required to provide informed consent (via checkboxes) to receive a confirmation link. After confirming the link, participants could access the password-protected internet portal and begin the web-based screening battery of self-report questionnaires assessing several sociodemographic characteristics, questions on trauma exposure, and clinical characteristics. After successfully completing the screening questionnaires, participants who met the aforementioned inclusion criteria booked an appointment for a clinical interview. Interviews were conducted by trained clinical interviewers by telephone or voice over IP. If further inclusion criteria (outlined previously) assessed in the interview were met, participants were assigned to PTSD or depression treatment based on their primary diagnosis. This study focused only on participants assigned to PTSD treatment. Participants who were deemed eligible for PTSD treatment were randomly allocated to the CR treatment, exposure treatment, or waitlist control group and subsequently

assigned to a counselor. Participants allocated to any of the treatments were able to begin treatment after the counselor sent a first letter. At 2 days after the interview, participants who were allocated to the waitlist condition were informed in the password-protected web portal that they would wait for 3 weeks until treatment. During treatment and the waiting time, participants regularly completed a questionnaire on posttraumatic stress symptoms. After completing the treatment or waiting time, participants filled out the set of web-based self-report questionnaires again to examine changes in clinical symptom presentation. Figure 1 shows the flow of participants through the trial.

Figure 1. Flowchart. *Some participants fulfill more than one exclusion criterion. **Includes cases that started the assessment without necessarily having completed all questionnaires; BDI-II: Beck Depression Inventory–II; CR: cognitive restructuring treatment; EXPO: exposure treatment; PTSD: posttraumatic stress disorder; SCID-5: Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; T2: assessment immediately before starting treatment or waiting time; T3: assessment after 2 letters or 1 week of waiting; T4: assessment after 4 letters or 2 weeks of waiting; WAIT: waitlist control condition.



Randomization and Blinding

Randomization was performed using block randomization with variable block sizes of 6, 9, and 12. The allocation schedule was created using the R package *Blockrand* (R Foundation for Statistical Computing) [29] and was embedded in the web portal. Allocation to any of the 3 conditions was performed invisibly and automatically on the web portal itself and, thus, was concealed (ie, participants, counselors, and researchers had no previous knowledge of and, therefore, no control over the group to which a participant would be allocated). Owing to the nature

of the provided treatments, participants and counselors could not be blinded to the treatment condition received.

Study Conditions

Overview of Treatments

The 2 treatments were based on an internet-based cognitive behavioral treatment approach for PTSD (Interapy) [30]. An overview of the procedure and writing examples of both treatment conditions can be found in Table S1 in Multimedia Appendix 1. The protocols were translated into Modern Standard Arabic. To obtain linguistically and culturally appropriate protocols, we made the following changes: (1) different versions for female and male participants regarding particularities in the Arabic language, (2) strengthening of the advice to not mention real names or places involved in the traumatic event because of basic precautionary measures, (3) use of pictorial metaphors (ie, scar or wound metaphor and linen cupboard metaphor for PTSD) to explain the purpose and process of trauma treatment in a less technical way, and (4) use of an encouraging and motivational but directive writing style. If needed (eg, the participant expressed a high level of faith), counselors could include quotes from the Qur'an. In addition, the layout of the protocols and the technical descriptions within the protocols were adapted to fit the format of the web portal (eg, participants were instructed to use the web-based planner). Both treatments consisted of twice-weekly 45-minute structured writing assignments in the form of letters over a period of approximately 3 weeks (approximately 2 letters/wk). The writing sessions were planned, and participants were instructed to plan the date and hour in which they would write each letter. After receiving each letter from a participant, the counselors provided individual feedback and instructions for the next letter within 2 working days. The feedback and instructions consisted of standard examples that were tailored to the participants' individual needs and the content of the previous letters. Both treatments began with an introduction by the counselor providing information on writing treatment in general, the procedure of the treatment in detail, and psychoeducational information on PTSD. As the counselors already had knowledge of the traumatic event to be addressed in the treatment through the interview report, they could directly refer to the most distressing traumatic event or associated dysfunctional thoughts and feelings in their first letter. At the beginning of each module, psychoeducational information on the specific treatment phase was provided. In both treatments, a final letter was sent at the end of treatment in which the counselor summarized the participant's progress during treatment.

Exposure Treatment

In addition to the introductory part, the exposure treatment included 2 different phases. In the first phase (self-confrontation), participants were instructed to write 4 letters about the traumatic event and their related thoughts, fears, and physiological reactions. They were asked to describe sensory perceptions in detail and focus on the most distressing situation of the trauma. We also included a section on how traumatic events are processed and why symptoms are maintained, as well as how exposure treatment could help, to make the condition comparable in length with the CR treatment. In the second phase—the social sharing phase—participants were asked to write 2 letters to summarize their memories of the trauma and consider how they were going to deal with the trauma in the future. The social sharing phase focused on a symbolic farewell letter that participants were instructed to address to themselves or to a significant other.

CR Treatment

In addition to the introductory part, the CR treatment included 2 different phases. The first phase of CR treatment encompassed 4 letters to reflect on automatic dysfunctional cognitions and adjust unrealistic assumptions (eg, guilt). Participants were instructed to write a letter to a hypothetical friend who had experienced the same traumatic event without necessarily going into the details of the traumatic experience. Compared with the original Interapy protocol, in which the CR phase was implemented after the exposure phase, we had to adapt the CR treatment part to enable participants to begin this phase without the previous knowledge gained through the exposure phase. Therefore, detailed information was provided in advance regarding the impact of traumatic events (ie, how traumatic experiences can influence thoughts and beliefs about oneself, other people, and the world and how these unhelpful thoughts and beliefs lead to emotions such as guilt and shame). In addition, to encourage participants to identify, challenge, and modify unhelpful beliefs, numerous reflective questions (eg, why did the traumatic event occur? What evidence and counterevidence is there that your friend is responsible for what happened?) were included before starting with the writing assignments. The second phase—social sharing—was identical to that described for the exposure treatment.

Waitlist Control Group

A comparison with the waitlist control group was conducted to account for the potential influence of elapsed time and quantify the efficacy of the 2 treatments. After a waiting period of 3 weeks, the waitlist participants received an email invitation to start the program. Before starting any of the 2 treatments (to which they had been randomized in advance), they completed all symptom questionnaires.

Reminder Messages

Participants received automated emails when they were supposed to log in to the portal (eg, when there was a letter or message from the counselor or a writing assignment was due). In addition, they received automated reminders at each step of the procedure if they were inactive. During the registration and screening process, participants received automated reminders after 3 and 7 days of inactivity and were excluded after 14 days of inactivity. During the interview process, they were reminded after 3, 7, and 14 days and excluded after 21 days of inactivity. Participants who had already been included and allocated to one of the conditions received an automated email after 3 and 7 days if they did not respond to the automated invitation (waitlist control group) or if they did not complete the letters on the chosen dates (treatment groups). In addition, if the participants did not respond to the 2 reminder messages, the counselor contacted them by telephone (if possible) to encourage them to continue. If they could not be reached by telephone, a

message was sent including a deadline for a response. After 14 days of nonresponse, participants were considered dropouts.

Counselors

A total of 10 native Arabic-speaking counselors living in Egypt or Germany performed the treatments. All counselors had a diploma in psychology or psychology-related disciplines (eg, social work, counseling, and psychotherapy) or extensive work experience. Counselors received continuous training covering information about and treatment options for PTSD, the fundamentals and technical aspects of internet-based treatments, specific treatment rationales, provision of feedback, and dealing with challenging situations. Furthermore, all counselors attended regular supervision meetings held by experienced psychotherapists. Support for participants via email or telephone was limited to emergency situations (ie, in cases of suicidality or dropout), technical support, or reminders to continue treatment.

Assessment

Structured Clinical Interview

The clinical interview by telephone or voice over IP was conducted in standard Arabic. It consisted of an introductory part (ie, explaining the procedure; informing about data security; and asking about age, current treatment and current medication, and past treatment in Ilajnafsy); a suicide screening measure (suicidal scale of the Mini-International Neuropsychiatric Interview [31]); substance and alcohol screening measures (Alcohol Use Disorders Identification Test [32] and Drug Abuse Screening Test–10 [33]); and relevant parts of the SCID-5 Clinical Version [27], namely, the sections covering PTSD, mood episodes, psychotic and associated symptoms, and a final part. All interviewers completed training on the administration of the interview, attended interviews conducted by an experienced interviewer, and conducted an interview under the supervision of an experienced interviewer with subsequent feedback. In addition, they received weekly supervision.

Web-Based Assessment

Primary and secondary outcome measures were self-reported and administered via the internet in a password-protected area. Instruments that were not available in standard Arabic at the time of planning the study were translated using the forward and backward translation method. Initial translation was conducted by a native Arabic-speaking person, and back translation was carried out by a different native speaker who had no knowledge of the original version. Subsequently, the 2 versions were compared, and deviations were discussed by a team of professionals before agreeing on a final version. Moreover, the instruction texts of the original instruments were adapted to fit the web-based format if necessary. Sociodemographic characteristics and exposure to traumatic events were assessed in the screening test battery only. For this purpose, we used items from the Harvard Trauma Questionnaire [34]; the Posttraumatic Diagnostic Scale [35]; and the Life Events Checklist for DSM-5 [36], with a total of 25 items asking about exposure to various potentially traumatic events, as well as the extended version of the Life Events Checklist for DSM-5 asking about further details of the most distressing event. All outcome measures were assessed as part of the screening test battery (baseline assessment; T1) and at the end of treatment or waiting time (postassessment time point; T5). Questionnaires asking about satisfaction with the treatment were administered after participants had completed any of the treatment conditions. In addition, the Posttraumatic Stress Disorder Checklist for the DSM-5 (PCL-5) [37] was administered at 3 measurement time points during treatment (T2: assessment immediately before starting treatment; T3: assessment after 2 letters; T4: assessment after 4 letters). In the waitlist control group, intermediate measures were administered in correspondence with the treatment groups (ie, participants were invited to complete the PCL-5 every week during the waiting period; T2: immediately before starting the waiting time; T3: assessment after 1 wk; T4: assessment after 2 wk).

Primary Outcome Measure

Symptoms of posttraumatic stress in the previous month were assessed using the PCL-5 [37]. The PCL-5 is a self-report questionnaire with 20 items that correspond to the DSM-5 PTSD symptoms. Each item is rated on a 5-point scale (from 0 to 4), with higher scores indicating greater symptom severity. A total of 4 subscales that correspond to the 4 different DSM-5 PTSD symptom clusters (re-experiencing, avoidance, negative alterations in cognitions and mood, and hyperarousal) can be differentiated. For each subscale, a sum score was calculated to assess the severity of each DSM-5 symptom cluster. The PCL-5 has proven to be a valid and reliable screening instrument for traumatized Arabic-speaking populations [38]. In this study, the Cronbach α was .87 for the overall scale and ranged from .70 (hyperarousal) to .83 (avoidance) for the subscales.

Secondary Outcome Measures

Posttraumatic maladaptive beliefs about the world, others, and the self were assessed using the self-report Posttraumatic Maladaptive Beliefs Scale (PMBS) [39]. The scale encompasses 15 statements that are rated on a 7-point scale (from 1 to 7). For the sake of consistency with the PCL-5, we used a past-month timeline of inquiry for the PMBS. A sum score was calculated to assess overall posttraumatic maladaptive beliefs, with higher scores indicating higher levels of maladaptive beliefs. Sensitivity to changes that can occur during treatment has been demonstrated [39]. The Cronbach α in this study was .75.

Trauma-related guilt cognitions were assessed using the guilt cognitions scale of the self-report Trauma-Related Guilt Inventory (TRGI) [40]. Respondents rate 22 statements on a 5-point scale (from 4 to 0) to indicate the degree to which they believe the statement is true. Mean scores were calculated, with higher scores indicating higher levels of maladaptive guilt cognitions. For the sake of consistency with the PCL-5, a past-month timeline of inquiry was used for the TRGI. The Cronbach α for the guilt cognitions scale in this sample was .90.

Anxiety symptom severity was measured using the self-report Arabic version of the Generalized Anxiety Disorder–7 (GAD-7) [41]. An Arabic version of the GAD-7 was used [42]. The questionnaire asks about general anxiety symptoms using 7 items rated on a 4-point scale (from 0 to 3) referring to the

previous 2 weeks. The sum score of all items serves as an indicator of generalized anxiety. The GAD-7 has shown poorer psychometric properties in Arabic-speaking populations than in Western populations [42,43]. In this sample, the Cronbach α was .80.

Depressive symptom severity was assessed using the self-report Patient Health Questionnaire–9 (PHQ-9) [44,45]. An Arabic version of the PHQ-9 was used [42]. The PHQ-9 includes 9 items rated on a 4-point scale (from 0 to 3). A sum score was calculated to determine depressive symptoms, with higher scores indicating greater symptom severity. The PHQ-9 has shown good internal consistency in different Arabic-speaking populations [42,46] and has already been used as a treatment outcome measure in patients from an Arab immigrant population receiving internet-based interventions for depression and anxiety [47]. In this sample, the Cronbach α was .79.

Somatoform symptom severity was measured using the self-report Patient Health Questionnaire–15 [48], which assesses somatic symptoms over the previous month. A total of 15 items for women and 14 items for men are rated on a 3-point scale (from 0 to 2). A sum score was calculated, with higher scores indicating greater impairment. The Patient Health Questionnaire–15 is widely used and has shown good psychometric properties in Western samples [48]. It showed good internal consistency in a study with Saudi Arabian university students [46] and was found to be valid in studies conducted with Saudi Arabian primary care patients [43]. In this sample, the Cronbach α was .79.

Quality of life was assessed using the self-report EUROHIS Quality of Life 8-item index, an adapted version of the World Health Organization Quality of Life Questionnaire and its shorter version [49]. The EUROHIS Quality of Life 8-item index assesses markers of quality of life using 8 items rated on a 5-point scale. A general quality of life index was determined by summing all items, with higher scores indicating better quality of life. The Arabic version of the short version of the World Health Organization Quality of Life Questionnaire has demonstrated adequate psychometric properties [50]. In this sample, the Cronbach α was .66.

Posttreatment Evaluation Questions

After completing any of the treatment conditions, participants were further asked about their experience of the treatment using the following specific questions: How satisfied were you with the treatment? (5-point scale from *totally satisfied* to *unsatisfied*), Was the treatment helpful? (5-point scale from *very helpful* to *not helpful*), Would you recommend the treatment? (5-point scale from *definitely* to *definitely not*), and How do you rate the duration of the treatment? (*too short*, *sufficient*, or *too long*).

Statistical Analyses

Overview of Statistical Analyses

Analyses were conducted using the R statistical software (version 4.2.2) [51] and the Mplus statistical modeling software (version 8, Muthén and Muthén) [52]. All 3 conditions were compared regarding baseline characteristics to see whether randomization worked properly. We examined the association between treatment condition and the proportion of participants who did not start treatment and those who wrote at least one letter (nonstarters vs starters). Nonstarters and starters were further compared in terms of sociodemographic, trauma-related, and clinical characteristics reported at baseline. Similarly, we investigated the association between treatment condition and the proportion of participants who stopped treatment before completing all 6 letters and those who completed all 6 letters (dropouts vs completers). Furthermore, in both treatment conditions, dropouts and completers were compared in terms of sociodemographic, trauma-related, and clinical characteristics reported at baseline. In addition, the duration of both treatment conditions (in days) was compared between both treatment conditions. The results of the posttreatment evaluation questions as markers of treatment satisfaction were compared between the 2 treatment conditions. All the aforementioned group differences were investigated using Welch or chi-square tests. If assumptions for the Welch tests were not met, the Kruskal-Wallis or Mann-Whitney tests were applied. The Fisher exact test was used as an alternative to the chi-square test. Treatment-associated changes in primary and secondary outcome measures across different measurement time points were modeled using multigroup latent change models [53,54]. The models were estimated using the robust maximum likelihood estimator. The rate of change is determined under the assumption that the score at a specific measurement time point after the initial assessment is composed of the initial score and the difference between the initial score and the score obtained at the specific measurement time point after the initial assessment (ie, the postassessment score) [55]. Thus, the rate of change between measurement time points is directly modeled in the form of the change score. The mean of the change scores represents the average change (decrease or increase) between 2 measurements within each condition in units of the questionnaire. Between-group effects are represented by the differences between the group-specific mean change scores. Within-group effect sizes (*d*) were computed by dividing the mean change scores by their SD for each group. Between-group effect sizes (*d*) were computed by dividing the mean difference between the mean change scores of the 2 groups by the pooled SD. All the results of treatment-associated changes were pooled across multiple imputed data sets. Bonferroni correction was applied to maintain the error rate at 0.05 for within-group changes and between-group differences. Therefore, a *P* value of <.005 (adjusted for 11 treatment outcomes) was considered statistically significant for within- and between-group comparisons. To assess reliable changes in individual posttraumatic stress symptom severity between baseline and the postassessment time point in all 3 conditions, we calculated the reliable change index for each participant [56] using the test-retest reliability of r=0.82 for the PCL-5 [57] and the SD at baseline of this sample (SD of 13.28 pooled across imputed data sets). According to this calculation, changes in posttraumatic stress symptom severity were considered statistically significant if the difference between baseline and the postassessment time point exceeded 16 points in the PCL-5 (α=.05). The proportions of participants with reliable improvement (16-point decrease minimum) or deterioration

(16-point increase minimum) were calculated. Furthermore, we calculated the rates of remitted participants (ie, participants with a baseline PCL-5 value of ≥23 as an indicator of caseness [having PTSD] and a postassessment PCL-5 value of <23). The cutoff value of 23 was chosen based on a study with Arabic-speaking people [38]. The results of the study showed that the PCL-5 achieved the best balance between sensitivity and specificity in the Arabic-speaking sample when this cutoff was used. In addition, the proportions of participants who experienced both reliable and clinically significant improvement (RCSI) were determined. The association between all 3 conditions and the proportion of participants with reliable change, experience of remission, and RCSI was examined using chi-square tests, which were pooled across all imputed data sets [58]. Analyses of treatment-associated changes were conducted on the intention-to-treat (ITT) and completer samples. Completers in both treatment conditions were defined as those participants who completed all 6 letters. In the waitlist control condition, completers were defined as participants who completed all questionnaires of the postassessment time point (T5). The results of the completer analyses can be found in Tables S2-S5 in Multimedia Appendix 1.

Missing Data

At baseline, 7 participants (CR treatment: n=3; exposure treatment: n=4) did not provide data for the trauma-related questionnaires (trauma exposure questions, PCL-5, PMBS, and TRGI). Owing to the low rate of missingness, statistical comparisons between specific groups (nonstarters vs starters and dropouts vs completers) that only included baseline scores were conducted using listwise deletion. The rates of missing values at the postassessment time point in the ITT sample with respect to all primary and secondary outcome measures ranged from 42.4% to 48.3% in the CR treatment group, from 41% to 48.4% in the exposure treatment group, and from 17.6% to 19.2% in the waitlist control condition. In the completer sample, rates of missing values at the postassessment time point were lower (CR treatment: 7.8%-12.5%; exposure treatment: 0%-3.4%). To deal with missing data, multiple imputation (100 imputed data sets; 50 iterations) for primary and secondary outcome measures was performed using the R package *MICE* [59]. All outcome measures were used in the imputation model. Predictive mean matching on the level of sum scores was applied for all variables except for the overall sum score of the PCL-5. For the overall sum scores of the PCL-5 at all measurement time points, passive imputation was used to account for the dependency of the overall PCL-5 sum score on the sum scores of the symptom clusters [60]. Multiple imputation was conducted separately for each of the 3 conditions. Following recommended guidelines [61], a sensitivity analysis was conducted to investigate whether deviations from the missing-at-random assumption would affect the conclusions drawn from the results calculated under the assumption that data are missing at random. For the primary outcome measured using the PCL-5, a total of 3 different conditions were modeled for the ITT sample. Individual imputed scores at each measurement time point (after the baseline assessment) increased by 25%, 50%, and 75% for all participants (Tables S6-S8 in Multimedia Appendix 1).

Ethics Approval

The Ethics Committee of the Freie Universität Berlin approved the study (107/2016).

Results

Participants

In total, 365 Arabic-speaking participants (CR treatment: n=118, 32.3%; exposure treatment: n=122, 33.4%; waitlist control: n=125, 34.2%) were included in this study. Participants were mainly female (272/365, 74.5%), single (227/365, 62.2%), living in urban areas (327/365, 89.6%), highly educated (331/365, 90.7%), and young adults (mean age 25.49, SD 6.68; range 18-53 y). The largest shares of participants were from Egypt (96/365, 26.3%), Saudi Arabia (69/365, 18.9%), and Syria (46/365, 12.6%) and were currently residing in Egypt (100/365, 27.4%), Saudi Arabia (60/365, 16.4%), and Jordan (25/365, 6.8%). On average, participants reported 5.19 (SD 3.72) different traumatic events in the trauma exposure questionnaire, with the worst event most frequently involving sexual violence (ie, “sexual assault by family member or acquaintance” [69/358, 19.3%], “sexual contact while under the age of 18 with a person at least 5 years older” [55/358, 15.4%], and “sexual assault by a stranger” [27/358, 7.5%]). PCL-5 scores at baseline ranged between 5 and 77, with a mean of 48.1 (SD 13.31). On average, participants reported an elevated level of depressive (mean 17.55, SD 5.17; range 2-27), anxiety (mean 14.21, SD 4.4; range 2-21), and somatoform (mean 14.5, SD 5.19; range 2-29) symptoms. In addition to having PTSD, most participants (268/365, 73.4%) had a comorbid depressive disorder (current or previous depressive episode, dysthymia, or both [“double depression”]), as assessed using the SCID-5. An overview of the sociodemographic, trauma-related, and clinical characteristics of participants in the total sample and in each condition is displayed in Table 1. No significant differences were found among the conditions regarding sociodemographic or trauma-related characteristics at baseline (Table 1). With regard to outcome measures, differences in anxiety symptom severity were found, with baseline values higher in the waitlist control group than in the 2 treatment conditions (Games-Howell post hoc tests: mean difference of 1.37, 95% CI 0.12-2.61, and P=.03 for exposure treatment vs waitlist control; mean difference of 1.41, 95% CI 0.09-2.73, and P=.03 for CR treatment vs waitlist control).

Table 1. Sociodemographic, trauma-related, and clinical characteristics of the total sample and subsamples in each condition (N=365).

	Total	CR[a] (n=118)	EXPO[b] (n=122)	WAIT[c] (n=125)	*F* test (*df*)[d]	*P* value
Sociodemographic characteristics						
Age (years), mean (SD)	25.49 (6.68)	25.04 (6.47)	26.39 (7.85)	25.03 (5.51)	0.10 (2)[e]	.61
Female sex, n (%)	272 (74.5)	91 (77.1)	91 (74.6)	90 (72)	0.84 (2)[f]	.66
Marital status, n (%)					Fisher[g]	.05
Single	227 (62.2)	78 (66.1)	68 (55.7)	81 (64.8)		
Married or in a relationship	121 (33.2)	30 (25.4)	51 (41.8)	40 (32)		
Divorced	15 (4.1)	8 (6.8)	3 (2.5)	4 (3.2)		
Widowed	2 (0.5)	2 (1.7)	0 (0)	0 (0)		
Education, n (%)					2.38 (2)[f]	.30
High education (high school, university, or college diploma)	331 (90.7)	111 (94.1)	109 (89.3)	111 (88.8)		
Low education (no or intermediate school diploma)	34 (9.3)	7 (5.9)	13 (10.7)	14 (11.2)		
Type of residence, n (%)					0.13 (2)[f]	.96
Urban (metropolitan city, small town, or suburb)	327 (89.6)	106 (89.8)	110 (90.2)	111 (88.8)		
Rural (village or single farmstead)	38 (10.4)	12 (10.2)	12 (9.8)	14 (11.2)		
Trauma-related characteristics						
Number of different traumatic events (trauma exposure list), mean (SD)	5.19 (3.72)	4.8 (3.39)	5.17 (4.01)	5.59 (3.71)	3.53 (2)[e]	.17
Exposure to sexual violence during most distressing trauma (LEC-5[h]), n (%)	113 (31.6)[i]	40 (34.8)[j]	33 (28)[k]	40 (32)	1.27 (2)[f]	.53
Clinical characteristics, mean (SD)						
Posttraumatic stress symptom severity (PCL-5[l])						
Overall	48.1 (13.31)[i]	46.63 (13.11)[j]	47.77 (13.62)[k]	49.77 (13.12)	1.77 (2, 235.86)	.17
Re-experiencing	11.03 (4.68)[i]	10.5 (4.96)[j]	10.85 (4.53)[k]	11.69 (4.51)	3.09 (2)[e]	.21
Avoidance	5.12 (2.35)[i]	4.95 (2.35)[j]	5.19 (2.35)[k]	5.22 (2.37)	1.14 (2)[e]	.57
Negative alterations in cognitions and mood	18.18 (5.25)[i]	17.89 (5.1)[j]	18.08 (5.47)[k]	18.54 (5.2)	1.68 (2)[e]	.64
Hyperarousal	13.77 (4.72)[i]	13.29 (4.83)[j]	13.64 (4.77)[k]	14.32 (4.55)	1.52 (2, 235.21)	.22
Posttraumatic maladaptive beliefs (PMBS[m])	67.08 (13.33)[i]	67.5 (14.77)[j]	66.88 (12.6)[k]	66.86 (12.68)	0.08 (2, 233.72)	.93
Trauma-related guilt (TRGI[n])	1.78 (0.85)[i]	1.81 (0.86)[j]	1.76 (0.85)[k]	1.77 (0.84)	0.28 (2)[e]	.87
Anxiety symptom severity (GAD-7[o])	14.21 (4.4)	13.71 (4.81)	13.75 (4.41)	15.12 (3.84)	7.24 (2)[e]	.03
Depressive symptom severity (PHQ-9[p])	17.55 (5.17)	17.52 (5.11)	17.14 (5.41)	17.98 (5)	1.30 (2)[e]	.52
Somatoform symptom severity (PHQ-15[q])	14.5 (5.19)	14.02 (4.79)	14.38 (5.53)	15.06 (5.19)	1.36 (2, 240.85)	.26
Quality of life (EUROHIS-QOL-8[r])	12.8 (4.43)	12.94 (4.45)	12.91 (4.34)	12.57 (4.51)	0.26 (2, 241.04)	.80

[a]CR: cognitive restructuring treatment.

[b]EXPO: exposure treatment.

[c]WAIT: waitlist control group.

[d]Group comparisons among all 3 groups are based on the Welch test unless otherwise stated.

[e]Kruskal-Wallis test (test statistic H).

[f]Chi-square test of independence.

[g]Fisher exact test.

[h]LEC-5: Life Events Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (extended version).

[i]n=358 because of missing data.

[j]n=115 because of missing data.

[k]n=118 because of missing data.

[l]PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[m]PMBS: Posttraumatic Maladaptive Beliefs Scale.

[n]TRGI: Trauma-Related Guilt Inventory.

[o]GAD-7: Generalized Anxiety Disorder–7.

[p]PHQ-9: Patient Health Questionnaire–9.

[q]PHQ-15: Patient Health Questionnaire–15.

[r]EUROHIS-QOL-8: EUROHIS Quality of Life 8-item index.

Use of Treatment

Starters Versus Nonstarters

Overall, 16.7% (40/240) of the participants did not start any of the treatments (CR treatment: 16/118, 13.6%; exposure treatment: 24/122, 19.7%) after allocation to treatment. There was no significant association between treatment condition and the proportion of participants who did not start treatment versus those who wrote at least one letter (χ^2_1=1.6; *P*=.20). Concerning trauma-related and clinical characteristics, none of the comparisons between nonstarters and starters reached significance (*P*>.05 in all cases). Regarding sociodemographic characteristics, starters included a significantly higher proportion of female participants (158/200, 79%) than nonstarters (24/40, 60%; *P*=.01).

Completers Versus Dropouts

Of the 200 participants who began any of the treatments, 123 (61.5%) completed all 6 letters (CR treatment: 64/102, 62.7%; exposure treatment: 59/98, 60%). There was no significant association between treatment condition and the proportion of participants who completed the treatment versus participants who dropped out (χ^2_1=0.1; *P*=.71). Welch tests revealed significant differences between completers and dropouts in both treatment conditions regarding baseline scores of overall posttraumatic stress symptom severity, negative alterations in cognitions and mood, maladaptive posttraumatic beliefs, depressive symptom severity, and quality of life. Other comparisons were nonsignificant (*P*>.05 in all cases). Games-Howell post hoc tests indicated that, compared with participants who dropped out of the CR treatment, those who completed the CR treatment had lower baseline posttraumatic stress symptom severity (mean difference of 6.93, 95% CI 0.90-12.95; *P*=.02), negative alterations in cognitions and mood (mean difference of 2.69, 95% CI 0.33-5.06; *P*=.02), and maladaptive posttraumatic beliefs (mean difference of 8.39, 95% CI 0.59-16.20; *P*=.03). Furthermore, compared with participants who dropped out of the CR treatment, those who completed either the exposure treatment or the CR treatment had lower depressive symptom severity at baseline, with a mean difference of 2.97 (95% CI 0.47-5.48; *P*=.01) and 2.71 (95% CI 0.20-5.22; *P*=.03), respectively. Quality of life at baseline was significantly higher in participants who completed the CR treatment than in those who dropped out of the CR treatment, with a mean difference of 2.50 (95% CI 0.20-4.80; *P*=.03). Table 2 summarizes the characteristics of participants who dropped out and those who completed any of the treatments, as well as the statistical results.

Table 2. Characteristics of and comparison between participants who completed <6 letters (dropouts) and participants who completed all 6 letters (completers).

	EXPO[a]		CR[b]		*F* test (*df*)[c]	*P* value
	Dropouts (n=39)	Completers (n=59)	Dropouts (n=38)	Completers (n=64)		
Sociodemographic characteristics						
Age (years), mean (SD)	24.74 (6.47)	27.81 (8.12)	24.03 (6.11)	24.97 (6.31)	7.08 (3)[d]	.07
Female sex, n (%)	28 (72)	47 (80)	30 (79)	53 (83)	1.80 (3)[e]	.62
Marital status, n (%)					Fisher[f]	.29
Single	22 (56)	34 (58)	24 (63)	44 (69)		
Married or in a relationship	17 (44)	22 (37)	10 (26)	17 (27)		
Divorced	0 (0)	3 (5)	3 (8)	3 (5)		
Widowed	0 (0)	0 (0)	1 (3)	0 (0)		
Education, n (%)					Fisher[f]	.58
High education (high school, university or college diploma)	35 (90)	53 (90)	36 (95)	61 (95)		
Low education (no or intermediate school diploma)	4 (10)	6 (10)	2 (5)	3 (5)		
Type of residence, n (%)					Fisher[f]	.07
Urban (metropolitan city, small town, or suburb)	36 (92)	54 (92)	30 (79)	61 (95)		
Rural (village or single farmstead)	3 (8)	5 (8)	8 (21)	3 (5)		
Trauma-related characteristics						
Number of different traumatic events (trauma exposure list), mean (SD)	5.15 (4.16)	5.58 (3.99)	4.53 (3.03)	4.92 (3.5)	1.82 (3)[d]	.61
Exposure to sexual violence during most distressing trauma (LEC-5[g]), n (%)	15 (41)[h]	13 (23)[i]	17 (46)[h]	19 (31)[j]	6.57 (3)[e]	.09
Clinical characteristics, mean (SD)						
Posttraumatic stress symptom severity (PCL-5[k])						
Overall	47.43 (13.15)[h]	48.25 (12.37)[i]	52.30 (10.85)[h]	45.37 (11.37)[j]	3.02 (3, 94.5)	.03
Re-experiencing	10.70 (4.55)[h]	10.82 (4.43)[i]	11.81 (4.64)[h]	10.24 (4.70)[j]	0.87 (3, 94.8)	.46
Avoidance	5.22 (2.45)[h]	5.30 (2.20)[i]	5.30 (2.60)[h]	4.97 (2.10)[j]	1.68 (3)[d]	.64
Negative alterations in cognitions and mood	17.68 (5.42)[h]	18.37 (5.16)[i]	20.08 (4.13)[h]	17.39 (4.67)[j]	3.19 (3, 95.3)	.03
Hyperarousal	13.84 (4.72)[h]	13.75 (4.43)[i]	15.11 (4.35)[h]	12.77 (4.50)[j]	2.15 (3, 94.6)	.10
Posttraumatic maladaptive beliefs (PMBS[l])	68.76 (12.08)[h]	66.65 (13.42)[i]	72.81 (14.09)[h]	64.42 (14.69)[j]	2.83 (3, 96.3)	.04
Trauma-related guilt (TRGI[m])	1.59 (0.75)[h]	1.89 (0.81)[i]	1.92 (0.84)[h]	1.75 (0.89)[j]	1.54 (3, 96.2)	.21
Anxiety symptom severity (GAD-7[n])	14.15 (4.57)	13 (4.46)	14.18 (4.34)	13.34 (4.97)	0.82 (3, 99.8)	.49
Depressive symptom severity (PHQ-9[o])	17.36 (5.9)	16.24 (4.92)	19.21 (4.38)	16.5 (5.15)	3.78 (3, 98.7)	.01
Somatoform symptom severity (PHQ-15[p])	14.95 (4.97)	14.17 (5.97)	15.13 (4.86)	13.67 (4.53)	0.10 (3, 98.4)	.40

	EXPO[a]		CR[b]		*F* test (*df*)[c]	*P* value
	Dropouts (n=39)	Completers (n=59)	Dropouts (n=38)	Completers (n=64)		
Quality of life (EUROHIS-QOL-8[q])	13.28 (4.62)	12.93 (4.38)	11.34 (4.04)	13.84 (4.69)	2.80 (3, 99.7)	.04

[a]EXPO: exposure treatment.

[b]CR: cognitive restructuring treatment.

[c]Group comparisons based on the Welch test unless otherwise stated.

[d]Kruskal-Wallis test (test statistic H).

[e]Chi-square test of independence.

[f]Fisher exact test.

[g]LEC-5: Life Events Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (extended version).

[h]n=37 because of missing data.

[i]n=57 because of missing data.

[j]n=62 because of missing data.

[k]PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (extended version).

[l]PMBS: Posttraumatic Maladaptive Beliefs Scale.

[m]TRGI: Trauma-Related Guilt Inventory.

[n]GAD-7: Generalized Anxiety Disorder–7.

[o]PHQ-9: Patient Health Questionnaire–9.

[p]PHQ-15: Patient Health Questionnaire–15.

[q]EUROHIS-QOL-8: EUROHIS Quality of Life 8-item index.

Duration of Treatment

Participants who started the CR treatment (102/118, 86.4%) were in treatment for an average of 38.8 (SD 19.7) days. Participants who started the exposure treatment (98/122, 80.3%) were in treatment for an average of 38 (SD 18.4) days. For participants who completed the CR treatment (64/102, 62.7%), the mean treatment duration was 35.2 (SD 15) days. For participants who completed the exposure treatment (59/98, 60%), the mean treatment duration was 32.2 (SD 13.3) days. The duration of treatment did not differ between the 2 treatment conditions either regarding participants who started any of the treatments (U=4882; P=.78) or regarding participants who completed any of the treatments (U=1652; P=.23).

Treatment Satisfaction

Of all completers who answered the evaluation questions, 89% (53/59) in the CR treatment and 86% (51/59) in the exposure treatment were completely satisfied, very satisfied, or satisfied with the treatment (U=1867; P=.48). Moreover, 93% (55/59) of patients in the CR treatment and 92% (54/59) of patients in the exposure treatment experienced the treatment as very helpful, helpful, or rather helpful (U=1920; P=.31). In both treatments, 95% (56/59) of the participants would recommend the treatment to someone else (U=1787.5; P=.78). In terms of treatment duration, 58% (34/59) of the participants in the CR treatment and 64% (38/59) of the participants in the exposure treatment experienced the treatment duration as sufficient (CR treatment: n=0 too long and 25/59, 42% too short; exposure treatment: 1/59, 2% too long and 20/59, 34% too short; P=.45).

Changes in Primary Outcome (ITT Sample)

Within-Group Changes

In both treatment conditions, no significant changes in overall posttraumatic stress symptom severity emerged between baseline and T2 (directly before starting treatment) or T3 (after 2 letters). Significant changes in overall posttraumatic stress symptom severity were found between baseline and T4 (after 4 letters; CR treatment: d=−0.45; exposure treatment: d=−0.54) and between baseline and the postassessment time point (CR treatment: d=−1.03 and exposure treatment: d=−1.00) in both treatment conditions. With regard to the subscale scores, significant changes between baseline and T3 (after 2 letters) emerged for the “re-experiencing” subscale in the CR treatment (d=−0.39) and for the “negative alterations in cognitions and mood” subscale in the exposure treatment (d=−0.30). Moreover, changes on the subscales “re-experiencing,” “negative alterations in cognitions and mood,” and “hyperarousal” between baseline and T4 (after 4 letters) were statistically significant in both treatment conditions, ranging from d=−0.32 (CR treatment; “hyperarousal” subscale) to d=−0.50 (CR treatment; “re-experiencing” subscale). In both treatment conditions, changes in symptom severity from baseline to the postassessment time point were statistically significant for all subscales (effect sizes ranging from d=−0.58 [CR treatment; “avoidance” subscale] to d=−0.94 [CR treatment; “re-experiencing” subscale]). Participants in the waitlist control condition showed statistically significant changes in overall posttraumatic stress symptom severity between baseline and T3 (after 1 wk of waiting) and T4 (after 2 wk of waiting) and between baseline and the postassessment time point (d=−0.29, d=−0.34, and d=−0.31, respectively). In addition, the waitlist control participants showed significant changes in the “re-experiencing” subscale between baseline and all subsequent

measurement time points (effect sizes ranging from d=−0.28 to d=−0.43). In the "negative alterations in cognitions and mood" subscale, significant changes emerged between baseline and T4 (after 2 wk of waiting; d=−0.35). Table 3 provides further information on estimates for within-group changes in posttraumatic stress symptom severity between baseline and each subsequent measurement time point in the ITT sample.

Table 3. Estimated within-group changes in posttraumatic stress symptom severity between baseline and subsequent assessments (intention-to-treat sample)[a].

Outcome (PCL-5[b]) and group	T2[c]–T1[d]			T3[e]–T1			T4[f]–T1			T5[g]–T1		
	M[h] (SE; 95% CI)	*P* value	*d*	M (SE; 95% CI)	*P* value	*d*	M (SE; 95% CI)	*P* value	*d*	M (SE; 95% CI)	*P* value	*d*
Overall												
CR[i]	−1.06 (1.00; −3.02 to 0.90)	.29	−0.10	−2.52 (1.16; −4.79 to −0.26)	.03	−0.22	−6.78 (1.54; −9.79 to −3.76)	*<.001*[j]	−0.45	−19.11 (2.17; −23.36 to −14.86)	*<.001*	−1.03
EX-PO[k]	−1.54 (1.05; −3.61 to 0.52)	.14	−0.14	−3.81 (1.45; −6.65 to −0.98)	.008	−0.27	−7.76 (1.48; −10.65 to −4.86)	*<.001*	−0.54	−17.00 (1.91; −20.74 to −13.26)	*<.001*	−1.00
WAIT[l]	−2.11 (1.05; −4.17 to −0.06)	.04	−0.18	−3.27 (1.08; −5.40 to −1.14)	*.003*	−0.29	−4.28 (1.26; −6.74 to −1.81)	*.001*	−0.34	−3.79 (1.18; −6.09 to −1.48)	*.001*	−0.31
Re-experiencing												
CR	−0.88 (0.35; −1.58 to −0.19)	.01	−0.24	−1.52 (0.44; −2.38 to −0.67)	*<.001*	−0.39	−2.23 (0.47; −3.15 to −1.31)	*<.001*	−0.50	−5.36 (0.63; −6.59 to −4.12)	*<.001*	−0.94
EXPO	−0.47 (0.33; −1.11 to 0.17)	.15	−0.14	−0.83 (0.45; −1.71 to 0.05)	.06	−0.19	−1.89 (0.45; −2.77 to −1.01)	*<.001*	−0.43	−3.99 (0.61; −5.20 to −2.79)	*<.001*	−0.74
WAIT	−1.17 (0.38; −1.92 to −0.42)	*.002*	−0.28	−1.80 (0.41; −2.61 to −0.99)	*<.001*	−0.43	−1.86 (0.45; −2.74 to −0.98)	*<.001*	−0.43	−1.82 (0.43; −2.65 to −0.98)	*<.001*	−0.41
Avoidance												
CR	0.06 (0.22; −0.38 to 0.50)	.79	0.03	0.20 (0.26; −0.31 to 0.70)	.44	0.08	0.01 (0.25; −0.48 to 0.50)	.95	0.01	−1.66 (0.33; −2.32 to −1.01)	*<.001*	−0.58
EXPO	0.03 (0.20; −0.37 to 0.42)	.89	0.01	−0.12 (0.28; −0.67 to 0.42)	.66	−0.05	−0.49 (0.29; −1.05 to 0.07)	.09	−0.18	−1.84 (0.33; −2.49 to −1.20)	*<.001*	−0.62
WAIT	0.33 (0.22; −0.10 to 0.75)	.13	0.14	0.16 (0.22; −0.28 to 0.60)	.47	0.07	0.11 (0.22; −0.32 to 0.54)	.61	0.05	0.25 (0.22; −0.18 to 0.68)	.25	0.11
Negative alterations in cognitions and mood												
CR	−0.09 (0.47; −1.00 to 0.83)	.85	−0.02	−0.69 (0.51; −1.69 to 0.31)	.18	−0.14	−2.79 (0.69; −4.15 to −1.44)	*<.001*	−0.42	−7.33 (0.96; −9.22 to −5.44)	*<.001*	−0.91
EXPO	−1.01 (0.55; −2.09 to 0.06)	.07	−0.17	−1.90 (0.66; −3.19 to −0.61)	*.004*	−0.30	−3.20 (0.68; −4.53 to −1.87)	*<.001*	−0.47	−6.65 (0.85; −8.32 to −4.98)	*<.001*	−0.92
WAIT	−0.83 (0.43; −1.68 to 0.02)	.06	−0.18	−1.08 (0.47; −2.00 to −0.16)	.02	−0.22	−1.75 (0.51; −2.74 to −0.75)	*.001*	−0.35	−1.27 (0.51; −2.26 to −0.28)	.01	−0.24
Hyperarousal												
CR	−0.15 (0.41; −0.95 to 0.65)	.71	−0.04	−0.51 (0.46; −1.41 to 0.40)	.27	−0.11	−1.76 (0.57; −2.89 to −0.64)	*.002*	−0.32	−4.76 (0.71; −6.15 to −3.37)	*<.001*	−0.82
EXPO	−0.09 (0.36; −0.79 to 0.61)	.80	−0.02	−0.96 (0.47; −1.89 to −0.03)	.04	−0.21	−2.18 (0.51; −3.18 to −1.19)	*<.001*	−0.46	−4.51 (0.62; −5.74 to −3.29)	*<.001*	−0.80

Outcome (PCL-5[b]) and group	T2[c]–T1[d]			T3[e]–T1			T4[f]–T1			T5[g]–T1		
	M[h] (SE; 95% CI)	*P* value	*d*	M (SE; 95% CI)	*P* value	*d*	M (SE; 95% CI)	*P* value	*d*	M (SE; 95% CI)	*P* value	*d*
WAIT	−0.44 (0.38; −1.19 to 0.31)	.25	−0.10	−0.56 (0.39; −1.33 to 0.21)	.15	−0.14	−0.78 (0.47; −1.70 to 0.15)	.10	−0.17	−0.95 (0.40; −1.74 to −0.16)	.02	−0.23

[a]All estimates were pooled across 100 imputed data sets.

[b]PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[c]T2: assessment immediately before starting treatment or waiting time.

[d]T1: baseline assessment.

[e]T3: assessment after 2 letters or 1 week of waiting.

[f]T4: assessment after 4 letters or 2 weeks of waiting.

[g]T5: postassessment time point.

[h]M: change score (mean change in raw score units of the questionnaire).

[i]CR: cognitive restructuring treatment.

[j]Significant *P* values.

[k]EXPO: exposure treatment.

[l]WAIT: waitlist control group.

Between-Group Differences

There were no statistically significant differences between the 2 treatment conditions regarding the magnitude of change in posttraumatic stress symptom severity (overall or subscales) at any measurement time point. Differences between the CR treatment and waitlist control group and between the exposure treatment and waitlist control group regarding the magnitude of change in symptom severity (overall or subscales) only reached significance between baseline and the postassessment time point. The effect sizes for overall posttraumatic stress symptom severity were high between the CR treatment and waitlist control group (d=0.98) and between the exposure treatment and waitlist control group (d=0.89). The effect sizes for the subscales ranged from d=0.44 (exposure treatment vs waitlist control; “re-experiencing” subscale) to d=0.90 (CR treatment vs waitlist control; “negative alterations in cognitions and mood” subscale).

Table 4 presents further details on estimated between-group differences in mean change scores and effect sizes with regard to posttraumatic stress symptom severity between baseline and each subsequent measurement time point for the ITT sample.

Table 4. Estimated between-group differences in posttraumatic stress symptom severity between baseline and subsequent assessments (intention-to-treat sample)[a].

Outcome (PCL-5[b]) and group comparison	T2[c]–T1[d]			T3[e]–T1			T4[f]–T1			T5[g]–T1		
	ΔM[h] (SE; 95% CI)	*P* value	*d*	ΔM (SE; 95% CI)	*P* value	*d*	ΔM (SE; 95% CI)	*P* value	*d*	ΔM (SE; 95% CI)	*P* value	*d*
Overall												
CR[i] vs EXPO[j]	−0.48 (1.45; −3.31 to 2.35)	.74	−0.04	−1.29 (1.88; −4.98 to 2.40)	.49	−0.10	−0.98 (2.15; −5.20 to 3.23)	.65	−0.07	2.12 (2.87; −3.51 to 7.74)	.46	0.12
EXPO vs WAIT[k]	−0.57 (1.49; −3.49 to 2.35)	.70	−0.05	0.54 (1.80; −2.98 to 4.06)	.76	0.04	3.48 (1.94; −0.32 to 7.29)	.07	0.26	13.21 (2.27; 8.76 to 17.66)	*<.001*[l]	0.89
CR vs WAIT	−1.05 (1.45; −3.90 to 1.80)	.47	−0.10	−0.75 (1.57; −3.83 to 2.34)	.64	−0.07	2.50 (2.00; −1.43 to 6.43)	.21	0.18	15.32 (2.48; 10.47 to 20.18)	*<.001*	0.98
Re-experiencing												
CR vs EXPO	0.41 (0.48; −0.52 to 1.35)	.38	0.12	0.69 (0.63; −0.54 to 1.92)	.27	0.17	0.34 (0.66; −0.95 to 1.64)	.60	0.08	1.37 (0.85; −0.31 to 3.04)	.11	0.25
EXPO vs. WAIT	−0.70 (0.51; −1.70 to 0.30)	.17	−0.18	−0.97 (0.60; −2.15 to 0.22)	.11	−0.23	0.03 (0.64; −1.23 to 1.28)	.97	0.01	2.18 (0.76; 0.68 to 3.67)	*.004*	0.44
CR vs WAIT	−0.29 (0.53; −1.32 to 0.75)	.59	−0.07	−0.27 (0.60; −1.45 to 0.91)	.65	−0.07	0.37 (0.66; −0.92 to 1.66)	.57	0.08	3.54 (0.77; 10.47 to 20.18)	*<.001*	0.70
Avoidance												
CR vs EXPO	−0.03 (0.30; −0.62 to 0.56)	.92	−0.01	−0.32 (0.38; −1.07 to 0.43)	.40	−0.12	−0.50 (0.37; −1.24 to 0.23)	.18	−0.20	−0.18 (0.48; −1.13 to 0.77)	.71	−0.06
EXPO vs WAIT	0.30 (0.29; −0.28 to 0.88)	.31	0.13	0.28 (0.35; −0.40 to 0.97)	.42	0.11	0.60 (0.36; −0.11 to 1.31)	.10	0.24	2.10 (0.40; 1.31 to 2.89)	*<.001*	0.79
CR vs WAIT	0.27 (0.31; −0.34 to 0.87)	.39	0.11	−0.04 (0.34; −0.71 to 0.63)	.91	−0.01	0.10 (0.34; −0.56 to 0.75)	.77	0.04	1.92 (0.40; 1.12 to 2.71)	*<.001*	0.74
Negative alterations in cognitions and mood												
CR vs EXPO	−0.93 (0.72; −2.33 to 0.48)	.20	−0.17	−1.21 (0.84; −2.86 to 0.43)	.15	−0.21	−0.41 (0.98; −2.33 to 1.51)	.68	−0.06	0.68 (1.28; −1.82 to 3.19)	.59	0.09
EXPO vs WAIT	0.18 (0.70; −1.19 to 1.55)	.80	0.03	0.82 (0.81; −0.76 to 2.41)	.31	0.14	1.45 (0.85; −0.21 to 3.12)	.09	0.24	5.38 (0.99; 3.43 to 7.32)	*<.001*	0.85
CR vs WAIT	−0.74 (0.64; −1.99 to 0.50)	.24	−0.15	−0.39 (0.69; −1.74 to 0.97)	.58	−0.08	1.04 (0.86; −0.63 to 2.72)	.22	0.18	6.06 (1.08; 3.95 to 8.17)	*<.001*	0.90
Hyperarousal												

Outcome (PCL-5[b]) and group comparison	T2[c]–T1[d]			T3[e]–T1			T4[f]–T1			T5[g]–T1		
	ΔM[h] (SE; 95% CI)	*P* value	*d*	ΔM (SE; 95% CI)	*P* value	*d*	ΔM (SE; 95% CI)	*P* value	*d*	ΔM (SE; 95% CI)	*P* value	*d*
CR vs EXPO	0.06 (0.54; −1.00 to 1.12)	.91	0.01	−0.45 (0.68; −1.79 to 0.88)	.51	−0.10	−0.42 (0.77; −1.93 to 1.10)	.59	−0.08	0.25 (0.96; −1.64 to 2.14)	.80	0.04
EXPO vs WAIT	−0.35 (0.52; −1.38 to 0.68)	.50	−0.09	0.40 (0.61; −0.80 to 1.60)	.52	0.09	1.40 (0.69; 0.05 to 2.76)	.04	0.30	3.56 (0.75; 2.10 to 5.03)	*<.001*	0.72
CR vs WAIT	−0.29 (0.56; −1.39 to 0.81)	.61	−0.07	−0.05 (0.60; −1.24 to 1.13)	.93	−0.01	0.99 (0.75; −0.48 to 2.46)	.19	0.20	3.81 (0.81; 2.22 to 5.39)	*<.001*	0.76

[a]All estimates were pooled across 100 imputed data sets.

[b]PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[c]T2: assessment immediately before starting treatment or waiting time.

[d]T1: baseline assessment.

[e]T3: assessment after 2 letters or 1 week of waiting.

[f]T4: assessment after 4 letters or 2 weeks of waiting.

[g]T5: postassessment time point.

[h]ΔM: difference between group-specific means of change scores.

[i]CR: cognitive restructuring treatment.

[j]EXPO: exposure treatment.

[k]WAIT: waitlist control group.

[l]Significant *P* values.

Reliable Change, Remission, and RCSI

Table 5 summarizes the rates of reliable change, remission, and RCSI for all 3 conditions in the ITT sample. The 2 treatment conditions did not significantly differ regarding the rates of participants who experienced reliable change (*P*=.74), remission (*P*=.49), or RCSI (*P*=.52) between baseline and the postassessment time point. The rates of reliable change (*P*<.001), remission (*P*<.001), and RCSI (*P*<.001) differed significantly between the CR treatment and waitlist control group. Similarly, the exposure treatment and waitlist control group differed significantly with regard to rates of reliable change (*P*<.001), remission (*P*<.001), and RCSI (*P*<.001).

Table 5. Rates of reliable change, remission, and reliable change and significant improvement (RCSI; intention-to-treat sample)[a].

	Baseline assessment to postassessment time point (PCL-5[b]), n (%)		
	CR[c] (n=118)	EXPO[d] (n=122)	WAIT[e] (n=125)
Reliable change			
Improvement	65.56 (55.6)	68.4 (56.1)	22.68 (18.1)
No change	48.96 (41.5)	48.75 (40)	93.27 (74.6)
Deterioration	3.48 (2.9)	4.85 (4)	9.05 (7.2)
Remission	48.44 (41.1)	45.63 (37.4)	4.65 (3.7)
RCSI	43.38 (36.8)	42.04 (34.5)	3.35 (2.7)

[a]All results were averaged across imputed data sets; therefore, the counts contain decimals.

[b]PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. An increase or decrease of 16 PCL-5 points between the baseline and postassessment time points was defined as reliable deterioration or improvement, respectively. Remission was defined as a score of ≥23 at baseline and a score of <23 at the postassessment time point. RCSI was defined as experiencing both remission and reliable improvement from baseline to the postassessment time point.

[c]CR: cognitive restructuring treatment.

[d]EXPO: exposure treatment.

[e]WAIT: waitlist control group.

Changes in Secondary Outcomes (ITT Sample)

Within-Group Changes

In both the CR and exposure treatments, all changes in secondary outcome measures between the baseline and postassessment time points were significant. In the ITT sample, the effect size estimates ranged from $d=-0.41$ (exposure treatment; trauma-related guilt) to $d=-1.10$ (CR treatment; depressive symptom severity). In the waitlist control condition, the levels of depressive and anxiety symptoms as well as quality of life also changed significantly ($d=-0.50$, $d=-0.44$, and $d=0.33$, respectively). Other changes were nonsignificant. The estimated within-group changes for all secondary outcomes between baseline and the postassessment time point in the ITT sample are shown in Table 6.

Table 6. Estimated within-group changes and between-group differences for secondary outcomes between baseline and the postassessment time point (intention-to-treat sample)[a].

Outcome and group	Within-group changes			Group comparison	Between-group differences		
	M[b] (SE; 95% CI)	*P* value	*d*		ΔM[c] (SE; 95% CI)	*P* value	*d*
Posttraumatic maladaptive beliefs (PMBS[d])							
CR[e]	−15.07 (2.26; −19.50 to −10.64)	*<.001* [f]	−0.73	CR vs EXPO[g]	3.66 (3.05; −2.31 to 9.63)	.23	0.20
EXPO	−11.41 (1.93; −15.19 to −7.64)	*<.001*	−0.74	EXPO vs WAIT[h]	10.57 (2.23; 6.19 to 14.95)	*<.001*	0.77
WAIT	−0.84 (1.14; −3.08 to 1.39)	.46	−0.07	CR vs WAIT	14.22 (2.51; 9.30 to 19.15)	*<.001*	0.86
Trauma-related guilt (TRGI[i])							
CR	−0.44 (0.09; −0.62 to −0.26)	*<.001*	−0.60	CR vs EXPO	0.07 (0.14; −0.21 to 0.34)	.64	0.08
EXPO	−0.37 (0.11; −0.58 to −0.16)	*.001*	−0.41	EXPO vs WAIT	0.26 (0.12; 0.02 to 0.51)	.03	0.34
WAIT	−0.11 (0.06; −0.23 to 0.01)	.09	−0.18	CR vs WAIT	0.33 (0.11; 0.12 to 0.54)	*.002*	0.49
Anxiety symptom severity (GAD-7[j])							
CR	−5.34 (0.69; −6.70 to −3.98)	*<.001*	−0.89	CR vs EXPO	0.30 (1.01; −1.68 to 2.28)	.77	0.05
EXPO	−5.04 (0.69; −6.38 to −3.69)	*<.001*	−0.87	EXPO vs WAIT	2.95 (0.82; 1.33 to 4.56)	*<.001*	0.56
WAIT	−2.09 (0.46; −2.98 to −1.19)	*<.001*	−0.44	CR vs WAIT	3.25 (0.83; 1.63 to 4.87)	*<.001*	0.60
Depressive symptom severity (PHQ-9[k])							
CR	−7.75 (0.86; −9.43 to −6.07)	*<.001*	−1.10	CR vs EXPO	1.14 (1.16; −1.13 to 3.41)	.33	0.16
EXPO	−6.61 (0.80; −8.17 to −5.04)	*<.001*	−0.95	EXPO vs WAIT	4.06 (0.94; 2.23 to 5.90)	*<.001*	0.67
WAIT	−2.55 (0.49; −3.50 to −1.59)	*<.001*	−0.50	CR vs WAIT	5.20 (0.99; 3.25 to 7.15)	*<.001*	0.85
Somatoform symptom severity (PHQ-15[l])							
CR	−2.80 (0.64; −4.04 to −1.55)	*<.001*	−0.50	CR vs EXPO	0.44 (0.90; −1.33 to 2.21)	.62	0.08
EXPO	−2.36 (0.62; −3.58 to −1.13)	*<.001*	−0.43	EXPO vs WAIT	2.12 (0.75; 0.65 to 3.58)	.005	0.44
WAIT	−0.24 (0.40; −1.02 to 0.55)	.55	−0.06	CR vs WAIT	2.56 (0.77; 1.06 to 4.06)	*.001*	0.52
Quality of life (EUROHIS-QOL-8[m])							
CR	5.18 (0.74; 3.72 to 6.64)	*<.001*	0.83	CR vs EXPO	−1.19 (1.10; −3.35 to 0.98)	.28	−0.19
EXPO	3.99 (0.74; 2.53 to 5.45)	*<.001*	0.66	EXPO vs WAIT	−2.63 (0.85; −4.29 to −0.96)	*.002*	−0.51
WAIT	1.36 (0.41; 0.56 to 2.17)	*.001*	0.33	CR vs WAIT	−3.81 (0.85; −5.48 to −2.15)	*<.001*	−0.72

[a]All estimates were pooled across 100 imputed data sets.

[b]M: change score (mean change in raw score units of the questionnaire).

[c]ΔM: difference between group-specific means of change scores.

[d]PMBS: Posttraumatic Maladaptive Beliefs Scale.

[e]CR: cognitive restructuring treatment.

[f]Significant *P* values.

[g]EXPO: exposure treatment.

[h]WAIT: waitlist control group.

[i]TRGI: Trauma-Related Guilt Inventory.

[j]GAD-7: Generalized Anxiety Disorder–7.

[k]PHQ-9: Patient Health Questionnaire–9.

[l]PHQ-15: Patient Health Questionnaire–15.

[m]EUROHIS-QOL-8: EUROHIS Quality of Life 8-item index.

Between-Group Differences

The results revealed no evidence of differences in the magnitude of change for any secondary outcome measure between the 2 treatment conditions. Significant differences in the magnitude of change between the exposure treatment and waitlist control group and between the CR treatment and waitlist control group were found with regard to posttraumatic maladaptive beliefs (d=0.77 and d=0.86, respectively), anxiety symptom severity (d=0.56 and d=0.60, respectively), depressive symptom severity (d=0.67 and d=0.85, respectively), and quality of life (d=−0.51 and d=−0.72, respectively). The differences in the magnitude of change regarding somatoform symptom severity (d=0.52) and trauma-related guilt cognitions (d=0.49) were significant only between the CR treatment and waitlist control group. The estimated between-group differences between baseline and the postassessment time point for all secondary outcomes in the ITT sample are shown in Table 6.

Sensitivity Analysis (ITT Sample)

The results of the sensitivity analysis for the ITT sample are reported in Tables S6-S8 in Multimedia Appendix 1. For the CR treatment and exposure treatment, changes in overall posttraumatic stress symptom severity between baseline and the postassessment time point were statistically significant when individual values increased by 25% or 50%. Effect sizes decreased to d=−0.79 and d=−0.57 in the CR treatment and to d=−0.75 and d=−0.51 in the exposure treatment. Under the most conservative *missing not at random* (MNAR) assumption, with a 75% increase in individual imputed values, only changes within the CR treatment remained statistically significant (d=−0.38). For the waitlist control condition, comparisons were either nonsignificant or marked by a significant increase in posttraumatic stress symptom severity.

The results of the between-group comparisons showed no significant differences between the 2 treatment conditions and the waitlist control condition regarding the magnitude of change in posttraumatic stress symptom severity between baseline and the second, third, and fourth measurement time points. In addition, no significant differences were found between the 2 treatment conditions at any measurement time point.

Differences between the exposure treatment and waitlist control group in the magnitude of change regarding overall posttraumatic stress symptom severity between baseline and the postassessment time point were significant under the 25% and 50% MNAR assumptions. Differences between the CR treatment and waitlist control group regarding the magnitude of change in overall posttraumatic stress symptom severity between baseline and the postassessment time points were significant under all 3 MNAR assumptions.

Discussion

Principal Findings

This study aimed to evaluate an internet-based CR treatment and an exposure treatment for Arabic-speaking participants with PTSD in terms of efficacy and course of PTSD symptom severity during treatment. First, we sought to examine the association between the 2 treatment conditions and treatment use. For this purpose, we investigated the proportion of participants who started treatment and the proportion of those who dropped out during treatment in both treatment conditions. Overall, 16.7% (40/240) of participants did not start treatment, which is similar to previous studies conducted in internet-based and face-to-face settings in different populations [11,62,63]. This finding underlines that, regardless of the treatment format and the characteristics of the specific sample, a substantial number of people drop out before attending the first session. In this sample, female participants seemed to be more likely to start treatment. The proportion of participants who dropped out after starting treatment was 38.5% (77/200), which is similar to the dropout rates reported in a study investigating different trauma-focused cognitive behavioral interventions for PTSD in a face-to-face setting [63]. The dropout rate in our study was indeed higher than the mean dropout rate of 20.9% reported in a recent meta-analysis of guideline-recommended PTSD treatments [64]. The highest mean dropout rates in the meta-analysis were reported for cognitive processing therapy (34%) and prolonged exposure (28.7%). However, heterogeneity was high across the studies in this meta-analysis. The dropout rate in our study was also higher than that reported in a previous study using the same short internet-based exposure treatment protocol in an Arabic-speaking population with PTSD as in this study [11]. A possible explanation for this discrepancy might lie in the changing circumstances of living conditions during the COVID-19 pandemic. This study was conducted during the pandemic, which was marked by multiple lockdowns. This might have led to a lower degree of privacy or even endangerment at home (ie, an increase in domestic violence [65]), making it difficult to continue with the treatments. Although internet-based interventions are especially important when physical contact is limited, a minimum of privacy is required to engage in treatment. During the pandemic, it may have been especially difficult for the participants to maintain their privacy as many young unmarried adults in Arabic-speaking countries tend to co-reside with other family members (eg, parents) because of the high value placed on family ties [66]. Furthermore, it is possible that participants could not continue treatment because of a SARS-CoV-2 infection.

There was no significant association between treatment condition and the proportion of participants who did not start treatment versus those who wrote at least one letter. Further, no significant association between treatment condition and the proportion of dropouts and completers was found. The duration of treatment was also similar in both treatment conditions. These findings demonstrate that the 2 treatment conditions were comparable with regard to the use of the treatments. Thus, exposure treatment does not generally seem to be more “aversive” than CR treatment (without detailed confrontation of the trauma) in Arabic-speaking participants even though many participants in this sample had experienced sexual violence, which might be associated with a strong fear of disclosing details of traumatic events [67]. In line with this lack of difference between the 2 treatment conditions regarding treatment use, Gutner et al [63] reported similar proportions of nonstarters versus starters and dropouts versus completers across multiple face-to-face treatment conditions with female

participants with PTSD after experiences of interpersonal violence. Thus, the use of different treatment methods seems to be independent of the specifics of different trauma samples. Notably, some differences between completers and dropouts emerged in our study, mainly in the CR treatment. Participants who completed CR treatment were less impaired by general posttraumatic stress and depressive symptoms and reported lower levels of negative alterations in cognitions and mood and posttraumatic maladaptive beliefs. In a study investigating female veterans with military sexual trauma who received cognitive processing therapy, lower trauma-related negative cognitions about the self emerged as a protective factor against attrition, and lower negative cognitions about others or the world were associated with a higher number of sessions (albeit not statistically significantly) [68]. However, research findings on predictors of dropout are inconsistent, with a previous meta-analysis finding no systematic differences across different interventions [69]. Furthermore, the treatment satisfaction of completers was examined and compared between the 2 treatments. In this study, the participants who completed any of the treatments seemed to be satisfied with the treatment and experienced it as helpful, highlighting that a good working alliance, which forms the basis for therapeutic processes, can be established in highly standardized internet-based treatments in Arabic-speaking populations [70]. However, in both conditions, nearly half of completers who answered the evaluation questions (45/118, 38.1%) reported that they experienced the treatment as too short. Thus, greater flexibility regarding treatment duration may be required for our sample of Arabic-speaking people with PTSD, who in addition may be struggling with the instability of living conditions in many countries in the region. For instance, it may be necessary to provide a variable treatment duration, with more sessions offered to those who still feel in need of treatment or may not have benefited sufficiently at the symptom level. In summary, a more personalized and tailored treatment approach (ie, concerning the number of sessions) to match the needs and characteristics of the participants might be necessary for internet-based settings as well.

Finally, we examined changes in posttraumatic stress symptom severity, posttraumatic cognitions, anxiety, depressive and somatoform symptom severity, and quality of life between baseline and the postassessment time point and compared them among all 3 conditions. Consistent with previous research on internet-based trauma-focused interventions in Arabic-speaking people [14,15], both treatments resulted in significant changes in PTSD symptom severity between baseline and the postassessment time point and performed better than the waitlist control group. Between baseline and the postassessment time point, we found large effects for overall PTSD symptom severity and for most symptom clusters within both treatment conditions. Most participants reliably improved during any of the treatment conditions (65.6/118, 55.6% in CR treatment and 68.4/122, 56.1% in exposure treatment). This suggests that internet-based interventions with a focus on processing the traumatic experience are applicable in Arabic-speaking populations independent of the specific treatment method being applied. Although the mean duration of both treatment conditions was slightly longer than intended (mean duration of approximately 35, SD 15 d for CR treatment and 32, SD 13 d for exposure treatment for completers), the treatments were still much shorter than most cognitive behavioral treatments conducted in a face-to-face setting. These findings highlight the potential of brief trauma-focused interventions delivered via the internet to close the treatment gap in the region. Interestingly, significant changes in PTSD symptom severity within both conditions mainly emerged after 4 letters, with no significant differences between the 2 conditions. This corresponds to the findings of a previous study in veterans with PTSD, which reported that a median of 4 sessions of prolonged exposure treatment was necessary to achieve symptomatic changes [71]. Differences in the magnitude of change between the exposure treatment and waitlist control groups, as well as between the CR treatment and waitlist control groups, were significant for the comparison between baseline and the postassessment time point. Therefore, despite their differing theoretical foundations, exposure and cognitive treatment techniques did not appear to target different PTSD symptom clusters at different time points as the 2 treatment conditions were comparable regarding changes in the 4 PTSD symptom clusters. These findings are in line with those of studies comparing exposure and cognitive techniques in face-to-face settings [18-20], which similarly revealed no differences in efficacy between the 2 techniques. Notably, some significant symptom changes between baseline and the postassessment time point were also observed in the waitlist control group, with re-experiencing symptoms declining significantly leading to a significant decline in overall PTSD symptom severity, although the effect sizes were small. A meta-analysis by Bradley et al [21] encompassing 15 randomized controlled trials also found a small “no-treatment” effect on PTSD symptoms. Nevertheless, a significant change in overall posttraumatic stress symptom severity between baseline and the postassessment time point could not be detected in the waitlist control group under MNAR conditions.

In line with previous research with Arabic-speaking people [11], changes in secondary outcomes between baseline and the postassessment time point were significant in both treatment conditions. In our sample, the largest effects for within-group changes were found for depressive and anxiety symptom severity, and the lowest effects were found for somatoform symptom severity and trauma-related guilt cognitions. As the 2 treatment conditions did not differ significantly in any of the secondary outcomes, they appear to be equally effective for the treatment of other comorbid mental health symptoms, posttraumatic cognitions, and quality of life aspects in Arabic-speaking populations. The treatment conditions outperformed the waitlist control group in terms of most secondary outcomes, with mostly medium-sized effects.

The sensitivity analysis did not alter the conclusions regarding changes within both treatment conditions as well as differences among all conditions under the 25% and 50% MNAR assumptions for overall posttraumatic stress symptom severity between baseline and the postassessment time point. Under the most conservative 75% MNAR assumption, only changes in overall posttraumatic symptom severity between baseline and the postassessment time point in the CR treatment remained significant. With regard to within-group changes during

treatment, the results of the sensitivity analysis provided some indication that, during the first sessions, a positive change seems unlikely.

Limitations

Some limitations need to be considered when interpreting our findings. The sample represented a rather specific group of the Arabic-speaking population as it mainly comprised female, young, and well-educated participants living in metropolitan areas. It seems that young and well-educated people with a stable internet connection are more familiar with new media and are more comfortable using the internet as a medium to receive treatment. In line with this, research has shown that women are more likely to use the internet for health purposes and that youth and higher education seem to positively affect health-related internet use [72]. Moreover, because of limited personnel resources, we were unable to assess the participants' clinical status at the end of treatment by means of a clinical interview. Thus, the results on the efficacy of the intervention are based on self-report measures. Finally, because of the limited follow-up data, we did not study the long-term effects of the interventions.

Conclusions and Outlook

In summary, the results of this study suggest that both exposure and CR treatment delivered via the internet are effective for Arabic-speaking people with PTSD. Even though both treatments were very brief, they performed significantly better than a waitlist control group, showing the high potential of brief trauma-focused cognitive behavioral treatment delivered via the internet to serve more people in need of treatment for symptoms of PTSD. Both treatments seem to lead to similar changes in PTSD and other comorbid mental health symptoms. It is worth mentioning that we observed higher attrition rates than in previous studies, which might be due to the COVID-19 pandemic and the associated potential lack of privacy at home. In view of the numerous challenges faced by Arabic-speaking people and the limited resources for mental health problems, internet-based interventions are a suitable option as they can be delivered independently of location, can be received at home, and have a short duration. In areas where more psychological help for mental health problems is needed, the interventions can be scaled up accordingly. Nevertheless, it is essential to gain more knowledge about those individual participants who did not benefit from any of the treatments (ie, reasons for dropout or nonimprovement). Therefore, future research needs to differentially examine predictors of treatment outcomes for both treatment types to obtain greater insights into whether specific groups of Arabic-speaking participants (ie, those with particular symptom profiles or higher posttraumatic emotions, cognitions, and behavior at baseline) benefit more from one treatment or the other and whether further adaptations are needed. In addition, it would be fruitful to examine participants' individual trajectories of change to identify specific groups of individuals and examine the characteristics of those who dropped out of any of the treatments (ie, through qualitative interviews). Future research should also examine the long-term effects of both treatment types.

Acknowledgments

The authors gratefully acknowledge the support, advice, and expertise of all members of the Ilajnafsy team in Germany and Egypt. They thank Freya Specht for assisting with the data collection process. The Ilajnafsy project was funded by Misereor, Germany. The funders had no role in the study design, data collection, data analysis, data interpretation, or reporting of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of both treatment conditions with writing examples, results of completer analyses, and results of sensitivity analysis.
[DOCX File , 107 KB - mental_v10i1e48689_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1239 KB - mental_v10i1e48689_app2.pdf]

References

1. A decade of displacement in the Middle East and North Africa. Internal Displacement Monitoring Centre. 2021 Feb. URL: https://www.internal-displacement.org/publications/a-decade-of-displacement-in-the-middle-east-and-north-africa [accessed 2023-01-17]
2. Awdeh A, Hamadi H. Factors hindering economic development: evidence from the MENA countries. Int J Emerg Mark 2019 Apr 01;14(2):281-299. [doi: 10.1108/ijoem-12-2017-0555]
3. Carta MG, Moro MF, Bass J. War traumas in the Mediterranean area. Int J Soc Psychiatry 2015 Feb 27;61(1):33-38. [doi: 10.1177/0020764014535754] [Medline: 24869846]

4. Hassan G, Ventevogel P, Jefee-Bahloul H, Barkil-Oteo A, Kirmayer LJ. Mental health and psychosocial wellbeing of Syrians affected by armed conflict. Epidemiol Psychiatr Sci 2016 Apr;25(2):129-141 [FREE Full text] [doi: 10.1017/S2045796016000044] [Medline: 26829998]
5. Zuberi A, Waqas A, Naveed S, Hossain MM, Rahman A, Saeed K, et al. Prevalence of mental disorders in the WHO Eastern Mediterranean region: a systematic review and meta-analysis. Front Psychiatry 2021 Jul 14;12:665019 [FREE Full text] [doi: 10.3389/fpsyt.2021.665019] [Medline: 34335323]
6. Okasha A, Karam E, Okasha T. Mental health services in the Arab world. World Psychiatry 2012 Feb;11(1):52-54 [FREE Full text] [doi: 10.1016/j.wpsyc.2012.01.008] [Medline: 22295010]
7. Stein J, Niemeyer H, Meyer C, Wirz C, Eiling A, Gruzman R, et al. Posttraumatic stress in adult civilians exposed to violent conflict, war and associated human rights abuses in the Eastern Mediterranean region: a systematic review and meta-analysis. J Affect Disord 2021 Nov 01;294:605-627. [doi: 10.1016/j.jad.2021.06.042] [Medline: 34332362]
8. WHO mental health atlas 2017. World Health Organization. 2018 Jan 1. URL: https://www.who.int/publications/i/item/9789241514019 [accessed 2023-01-23]
9. Stein J, Knaevelsrud C. New technologies in the treatment of psychological disorders – overview, potentials and barriers for conflict-affected populations: theory, research and clinical practice. In: Mental Health of Refugee and Conflict-Affected Populations. Cham, Switzerland: Springer International Publishing; Dec 2018:373-391.
10. Kayrouz R, Dear BF, Karin E, Fogliati VJ, Gandy M, Keyrouz L, et al. Acceptability of mental health services for anxiety and depression in an Arab sample. Community Ment Health J 2018 Aug 24;54(6):875-883. [doi: 10.1007/s10597-018-0235-y] [Medline: 29368132]
11. Böttche M, Wagner B, Vöhringer M, Heinrich M, Stein J, Selmo P, et al. Is only one cognitive technique also effective? Results from a randomized controlled trial of two different versions of an internet-based cognitive behavioural intervention for post-traumatic stress disorder in Arabic-speaking countries. Eur J Psychotraumatol 2021 Jul 15;12(1):1943870 [FREE Full text] [doi: 10.1080/20008198.2021.1943870] [Medline: 34345377]
12. Simon N, Robertson L, Lewis C, Roberts NP, Bethell A, Dawson S, et al. Internet-based cognitive and behavioural therapies for post-traumatic stress disorder (PTSD) in adults. Cochrane Database Syst Rev 2021 May 20;5(5):CD011710 [FREE Full text] [doi: 10.1002/14651858.CD011710.pub3] [Medline: 34015141]
13. Bisson JI, Ariti C, Cullen K, Kitchiner N, Lewis C, Roberts NP, et al. Guided, internet based, cognitive behavioural therapy for post-traumatic stress disorder: pragmatic, multicentre, randomised controlled non-inferiority trial (RAPID). BMJ 2022 Jun 16;377:e069405 [FREE Full text] [doi: 10.1136/bmj-2021-069405] [Medline: 35710124]
14. Kayrouz R, Dear BF, Kayrouz B, Karin E, Gandy M, Titov N. Meta-analysis of the efficacy and acceptability of cognitive-behavioural therapy for Arab adult populations experiencing anxiety, depression or post-traumatic stress disorder. Cogn Behav Ther 2018 Sep;47(5):412-430. [doi: 10.1080/16506073.2018.1445124] [Medline: 29714106]
15. Knaevelsrud C, Brand J, Lange A, Ruwaard J, Wagner B. Web-based psychotherapy for posttraumatic stress disorder in war-traumatized Arab patients: randomized controlled trial. J Med Internet Res 2015 Mar 20;17(3):e71 [FREE Full text] [doi: 10.2196/jmir.3582] [Medline: 25799024]
16. Ehlers A, Wild J, Warnock-Parkes E, Grey N, Murray H, Kerr A, et al. Therapist-assisted online psychological therapies differing in trauma focus for post-traumatic stress disorder (STOP-PTSD): a UK-based, single-blind, randomised controlled trial. Lancet Psychiatry 2023 Aug;10(8):608-622. [doi: 10.1016/s2215-0366(23)00181-5]
17. Lewis C, Roberts NP, Andrew M, Starling E, Bisson JI. Psychological therapies for post-traumatic stress disorder in adults: systematic review and meta-analysis. Eur J Psychotraumatol 2020 Mar 10;11(1):1729633 [FREE Full text] [doi: 10.1080/20008198.2020.1729633] [Medline: 32284821]
18. Resick PA, Nishith P, Weaver TL, Astin MC, Feuer CA. A comparison of cognitive-processing therapy with prolonged exposure and a waiting condition for the treatment of chronic posttraumatic stress disorder in female rape victims. J Consult Clin Psychol 2002 Aug;70(4):867-879. [doi: 10.1037/0022-006x.70.4.867]
19. Tarrier N, Pilgrim H, Sommerfield C, Faragher B, Reynolds M, Graham E, et al. A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic posttraumatic stress disorder. J Consult Clin Psychol 1999 Feb;67(1):13-18. [doi: 10.1037/0022-006x.67.1.13]
20. Horesh D, Qian M, Freedman S, Shalev A. Differential effect of exposure-based therapy and cognitive therapy on post-traumatic stress disorder symptom clusters: a randomized controlled trial. Psychol Psychother 2017 Jun;90(2):235-243 [FREE Full text] [doi: 10.1111/papt.12103] [Medline: 27561944]
21. Bradley R, Greene J, Russ E, Dutra L, Westen D. A multidimensional meta-analysis of psychotherapy for PTSD. Am J Psychiatry 2005 Feb;162(2):214-227. [doi: 10.1176/appi.ajp.162.2.214] [Medline: 15677582]
22. Sloan DM, Marx BP, Lee DJ, Resick PA. A brief exposure-based treatment vs cognitive processing therapy for posttraumatic stress disorder: a randomized noninferiority clinical trial. JAMA Psychiatry 2018 Mar 01;75(3):233-239 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.4249] [Medline: 29344631]
23. Wild J, Warnock-Parkes E, Grey N, Stott R, Wiedemann M, Canvin L, et al. Internet-delivered cognitive therapy for PTSD: a development pilot series. Eur J Psychotraumatol 2016 Nov 8;7:31019 [FREE Full text] [doi: 10.3402/ejpt.v7.31019] [Medline: 27837579]

24. Spence J, Titov N, Johnston L, Jones MP, Dear BF, Solley K. Internet-based trauma-focused cognitive behavioural therapy for PTSD with and without exposure components: a randomised controlled trial. J Affect Disord 2014 Jun;162:73-80. [doi: 10.1016/j.jad.2014.03.009] [Medline: 24767009]
25. Nossier SA. Violence against women in the Arab world. J Egypt Public Health Assoc 2015;90(3):85-86. [doi: 10.1097/01.epx.0000471671.51800.34]
26. Vöhringer M, Knaevelsrud C, Wagner B, Slotta M, Schmidt A, Stammel N, et al. Should I stay or must I go? Predictors of dropout in an internet-based psychotherapy programme for posttraumatic stress disorder in Arabic. Eur J Psychotraumatol 2020 Jan 23;11(1):1706297 [FREE Full text] [doi: 10.1080/20008198.2019.1706297] [Medline: 32082510]
27. First MB, Williams JB, Karg RS, Spitzer RL. SCID-5-CV Structured Clinical Interview for DSM-5 Disorders, Clinician Version. Arlington, VA: American Psychiatric Association; 2016.
28. Center Überleben. URL: https://ilajnafsy.bzfo.de/portal/ [accessed 2023-12-04]
29. Snow G. R package: Randomization for block random clinical trials. The Comprehensive R Archive Network. 2020 Apr 06. URL: https://cran.r-project.org/web/packages/blockrand/blockrand.pdf [accessed 2023-03-05]
30. Lange A, van de Ven JP, Schrieken B. Interapy: treatment of post-traumatic stress via the internet. Cogn Behav Ther 2003;32(3):110-124. [doi: 10.1080/16506070302317] [Medline: 16291543]
31. Lecrubier Y, Sheehan DV, Weiller E, Amorim P, Bonora I, Sheehan KH, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur Psychiatr 1997;12(5):224-231. [doi: 10.1016/s0924-9338(97)83296-8]
32. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
33. Skinner HA. The drug abuse screening test. Addict Behav 1982 Jan;7(4):363-371. [doi: 10.1016/0306-4603(82)90005-3] [Medline: 7183189]
34. Mollica RF, Caspi-Yavin Y, Bollini P, Truong T, Tor S, Lavelle J. The Harvard Trauma Questionnaire: validating a cross-cultural instrument for measuring torture, trauma, and posttraumatic stress disorder in Indochinese refugees. J Nervous Ment Dis 1992;180(2):111-116. [doi: 10.1097/00005053-199202000-00008]
35. Foa EB, Cashman L, Jaycox L, Perry K. The validation of a self-report measure of posttraumatic stress disorder: the Posttraumatic Diagnostic Scale. Psychol Assess 1997 Dec;9(4):445-451. [doi: 10.1037/1040-3590.9.4.445]
36. Weathers FW, Blake DD, Schnurr PP, Kaloupek DG, Marx BP, Keane TM. The Life Events Checklist for DSM-5 (LEC-5) – standard. National Center for Posttraumatic Stress Disorder. 2013. URL: https://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp [accessed 2022-11-23]
37. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD checklist for DSM-5 (PCL-5). National Center for PTSD. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2022-11-23]
38. Ibrahim H, Ertl V, Catani C, Ismail AA, Neuner F. The validity of posttraumatic stress disorder checklist for DSM-5 (PCL-5) as screening instrument with Kurdish and Arab displaced populations living in the Kurdistan region of Iraq. BMC Psychiatry 2018 Aug 16;18(1):259 [FREE Full text] [doi: 10.1186/s12888-018-1839-z] [Medline: 30115040]
39. Vogt DS, Shipherd JC, Resick PA. Posttraumatic maladaptive beliefs scale: evolution of the personal beliefs and reactions scale. Assessment 2012 Sep 21;19(3):308-317. [doi: 10.1177/1073191110376161] [Medline: 20660470]
40. Kubany ES, Haynes SN, Abueg FR, Manke FP, Brennan JM, Stahura C. Development and validation of the trauma-related guilt inventory (TRGI). Psychol Assess 1996 Dec;8(4):428-444. [doi: 10.1037/1040-3590.8.4.428]
41. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
42. Sawaya H, Atoui M, Hamadeh A, Zeinoun P, Nahas Z. Adaptation and initial validation of the patient health questionnaire - 9 (PHQ-9) and the generalized anxiety disorder - 7 questionnaire (GAD-7) in an Arabic speaking Lebanese psychiatric outpatient sample. Psychiatry Res 2016 May 30;239:245-252. [doi: 10.1016/j.psychres.2016.03.030] [Medline: 27031595]
43. Becker S, Al Zaid K, Al Faris E. Screening for somatization and depression in Saudi Arabia: a validation study of the PHQ in primary care. Int J Psychiatry Med 2002 Sep 22;32(3):271-283. [doi: 10.2190/xtdd-8l18-p9e0-jyrv]
44. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002 Sep 01;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
45. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
46. AlHadi AN, AlAteeq DA, Al-Sharif E, Bawazeer HM, Alanazi H, AlShomrani AT, et al. An arabic translation, reliability, and validation of patient health questionnaire in a Saudi sample. Ann Gen Psychiatry 2017 Sep 6;16(1):32 [FREE Full text] [doi: 10.1186/s12991-017-0155-1] [Medline: 28878812]
47. Kayrouz R, Dear BF, Johnston L, Gandy M, Fogliati VJ, Sheehan J, et al. A feasibility open trial of guided internet-delivered cognitive behavioural therapy for anxiety and depression amongst Arab Australians. Internet Interv 2015 Mar;2(1):32-38. [doi: 10.1016/j.invent.2014.12.001]

48. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [doi: 10.1097/00006842-200203000-00008] [Medline: 11914441]
49. Schmidt S, Mühlan H, Power M. The EUROHIS-QOL 8-item index: psychometric results of a cross-cultural field study. Eur J Public Health 2006 Aug;16(4):420-428. [doi: 10.1093/eurpub/cki155] [Medline: 16141303]
50. Ohaeri JU, Awadalla AW. The reliability and validity of the short version of the WHO quality of life instrument in an Arab general population. Ann Saudi Med 2009;29(2):98-104 [FREE Full text] [doi: 10.4103/0256-4947.51790] [Medline: 19318760]
51. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.R-project.org/ [accessed 2023-11-27]
52. Muthén LK, Muthén BO. Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén; 2017.
53. McArdle JJ. Latent variable modeling of differences and changes with longitudinal data. Annu Rev Psychol 2009 Jan 01;60(1):577-605. [doi: 10.1146/annurev.psych.60.110707.163612] [Medline: 18817479]
54. Klopack ET, Wickrama KK. Modeling latent change score analysis and extensions in Mplus: a practical guide for researchers. Struct Equ Modeling 2020;27(1):97-110 [FREE Full text] [doi: 10.1080/10705511.2018.1562929] [Medline: 33013155]
55. Geiser C. Data Analysis with Mplus. New York, NY: Guilford Publications; 2013.
56. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037/0022-006x.59.1.12]
57. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
58. Enders C. Applied Missing Data Analysis. New York, NY: Guilford Publications; 2010.
59. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
60. van Buuren S. Flexible Imputation of Missing Data, Second Edition. Boca Raton, FL: CRC Press; 2018.
61. O'Kelly M, Ratitch B. Clinical Trials with Missing Data: A Guide for Practitioners. Hoboken, NJ: John Wiley & Sons; 2014.
62. Simon N, McGillivray L, Roberts NP, Barawi K, Lewis CE, Bisson JI. Acceptability of internet-based cognitive behavioural therapy (i-CBT) for post-traumatic stress disorder (PTSD): a systematic review. Eur J Psychotraumatol 2019 Aug 27;10(1):1646092 [FREE Full text] [doi: 10.1080/20008198.2019.1646092] [Medline: 31497259]
63. Gutner CA, Gallagher MW, Baker AS, Sloan DM, Resick PA. Time course of treatment dropout in cognitive-behavioral therapies for posttraumatic stress disorder. Psychol Trauma 2016 Jan;8(1):115-121 [FREE Full text] [doi: 10.1037/tra0000062] [Medline: 26098737]
64. Varker T, Jones KA, Arjmand HA, Hinton M, Hiles SA, Freijah I, et al. Dropout from guideline-recommended psychological treatments for posttraumatic stress disorder: a systematic review and meta-analysis. J Affect Disord Report 2021 Apr;4:100093. [doi: 10.1016/j.jadr.2021.100093]
65. Piquero AR, Jennings WG, Jemison E, Kaukinen C, Knaul FM. Domestic violence during the COVID-19 pandemic - evidence from a systematic review and meta-analysis. J Crim Justice 2021 May;74:101806 [FREE Full text] [doi: 10.1016/j.jcrimjus.2021.101806] [Medline: 36281275]
66. Mehio-Sibai A, Beydoun MA, Tohme RA. Living arrangements of ever-married older Lebanese women: is living with married children advantageous? J Cross Cult Gerontol 2009 Mar;24(1):5-17 [FREE Full text] [doi: 10.1007/s10823-008-9057-7] [Medline: 18283537]
67. Ullman SE, O'Callaghan E, Shepp V, Harris C. Reasons for and experiences of sexual assault nondisclosure in a diverse community sample. J Fam Violence 2020 Nov;35(8):839-851 [FREE Full text] [doi: 10.1007/s10896-020-00141-9] [Medline: 33746358]
68. Holder N, Holliday R, Wiblin J, LePage JP, Surís A. Predictors of dropout from a randomized clinical trial of cognitive processing therapy for female veterans with military sexual trauma-related PTSD. Psychiatry Res 2019 Jun;276:87-93. [doi: 10.1016/j.psychres.2019.04.022] [Medline: 31030005]
69. Imel ZE, Laska K, Jakupcak M, Simpson TL. Meta-analysis of dropout in treatments for posttraumatic stress disorder. J Consult Clin Psychol 2013 Jun;81(3):394-404 [FREE Full text] [doi: 10.1037/a0031474] [Medline: 23339535]
70. Wagner B, Brand J, Schulz W, Knaevelsrud C. Online working alliance predicts treatment outcome for posttraumatic stress symptoms in Arab war-traumatized patients. Depress Anxiety 2012 Jul 07;29(7):646-651. [doi: 10.1002/da.21962] [Medline: 22678971]
71. Holder N, Shiner B, Li Y, Madden E, Neylan TC, Seal KH, et al. Determining the median effective dose of prolonged exposure therapy for veterans with posttraumatic stress disorder. Behav Res Ther 2020 Dec;135:103756. [doi: 10.1016/j.brat.2020.103756] [Medline: 33157475]
72. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of e-health services: a study of seven countries. BMC Public Health 2007 Apr 10;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]

Abbreviations

CR: cognitive restructuring
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
GAD-7: Generalized Anxiety Disorder–7
ITT: intention-to-treat
MNAR: missing not at random
PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
PHQ-9: Patient Health Questionnaire–9
PMBS: Posttraumatic Maladaptive Beliefs Scale
PTSD: posttraumatic stress disorder
RCSI: reliable and clinically significant improvement
SCID-5: Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
TRGI: Trauma-Related Guilt Inventory

Edited by J Torous; submitted 27.07.23; peer-reviewed by M Bragesjö, N Titov; comments to author 18.08.23; revised version received 15.09.23; accepted 16.09.23; published 13.12.23.

Please cite as:
Stein J, Vöhringer M, Wagner B, Stammel N, Nesterko Y, Böttche M, Knaevelsrud C
Exposure Versus Cognitive Restructuring Techniques in Brief Internet-Based Cognitive Behavioral Treatment for Arabic-Speaking People With Posttraumatic Stress Disorder: Randomized Clinical Trial
JMIR Ment Health 2023;10:e48689
URL: https://mental.jmir.org/2023/1/e48689
doi:10.2196/48689
PMID:38090792

Original Paper

Microtemporal Dynamics of Dietary Intake, Physical Activity, and Impulsivity in Adult Attention-Deficit/Hyperactivity Disorder: Ecological Momentary Assessment Study Within Nutritional Psychiatry

Alea Ruf[1], PhD; Andreas B Neubauer[2,3], PhD; Elena D Koch[4], PhD; Ulrich Ebner-Priemer[4,5], Prof Dr; Andreas Reif[1], Prof Dr; Silke Matura[1], Dr rer med, PD

[1]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital, Goethe University, Frankfurt, Germany

[2]DIPF | Leibniz Institute for Research and Information in Education, Frankfurt, Germany

[3]Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt, Germany

[4]Mental mHealth Lab, Institute of Sports and Sports Science, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[5]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

Corresponding Author:
Alea Ruf, PhD
Department of Psychiatry, Psychosomatic Medicine and Psychotherapy
University Hospital
Goethe University
Heinrich-Hoffmann-Straße 10
Frankfurt, 60528
Germany
Phone: 49 69 6301 83348
Email: alea.ruf@kgu.de

Abstract

Background: Increasing attention is being paid to lifestyle factors, such as nutrition and physical activity (PA), as potential complementary treatment options in attention-deficit/hyperactivity disorder (ADHD). Previous research indicates that sugar and saturated fat intake may be linked to increased impulsivity, a core symptom of ADHD, whereas protein intake and PA may be related to reduced impulsivity. However, most studies rely on cross-sectional data that lack microtemporal resolution and ecological validity, wherefore questions of microtemporal dynamics (eg, is the consumption of foods high in sugar associated with increased impulsivity within minutes or hours?) remain largely unanswered. Ecological momentary assessment (EMA) has the potential to bridge this gap.

Objective: This study is the first to apply EMA to assess microtemporal associations among macronutrient intake, PA, and state impulsivity in the daily life of adults with and without ADHD.

Methods: Over a 3-day period, participants reported state impulsivity 8 times per day (signal-contingent), recorded food and drink intake (event-contingent), and wore an accelerometer. Multilevel 2-part models were used to study the association among macronutrient intake, PA, and the probability to be impulsive as well as the intensity of impulsivity (ADHD: n=36; control: n=137).

Results: No association between macronutrient intake and state impulsivity was found. PA was not related to the intensity of impulsivity but to a higher probability to be impulsive (ADHD: β=−.09, 95% CI −0.14 to −0.04; control: β=−.03, 95% CI −0.05 to −0.01). No evidence was found that the combined intake of saturated fat and sugar amplified the increase in state impulsivity and that PA alleviated the positive association between sugar or fat intake and state impulsivity.

Conclusions: Important methodological considerations are discussed that can contribute to the optimization of future EMA protocols. EMA research in the emerging field of nutritional psychiatry is still in its infancy; however, EMA is a highly promising and innovative approach as it offers insights into the microtemporal dynamics of psychiatric symptomology, dietary intake, and PA in daily life.

(JMIR Ment Health 2023;10:e46550) doi:10.2196/46550

KEYWORDS

impulsivity; nutrition; macronutrient intake; physical activity; ecological momentary assessment; EMA; attention-deficit/hyperactivity disorder; ADHD; diet; neurodevelopmental; hyperactivity; macronutrient; psychiatry; symptoms; mobile; impulsivity; mobile phone

Introduction

Background

Attention-deficit/hyperactivity disorder (ADHD), initially conceptualized as a neurodevelopmental disorder restricted to childhood, is now recognized as a condition persisting into adulthood, with approximately 2.5% to 3% of adults showing clinically relevant symptoms [1,2]. ADHD is characterized by impaired symptoms of hyperactivity, inattention, and impulsivity. Individuals with ADHD are at risk for poor academic performance [3], accidents [4], financial problems [5], and numerous adverse outcomes (the review by Kooij et al [6] provides an overview). A multimodal and multidisciplinary approach, including pharmacotherapy, cognitive behavioral therapy, psychoeducation, and coaching, should be applied for the treatment of adult ADHD [6]. However, adult ADHD is underdiagnosed and undertreated [6], likely because of the unavailability of diagnostic services or their limited availability in very few specialized facilities [7]. Despite the high efficacy of pharmacotherapy for the short-term treatment of ADHD [8], not every patient responds to medication (eg, 24% of nonresponders [9]), long-term effects of pharmacotherapy are understudied [8], and some studies suggest that pharmacological treatment is positively associated with symptom severity in the long term [10]. Furthermore, pharmacotherapy is associated with several side effects such as increased heart rate and blood pressure [11], reduced appetite [12], and sleep problems (for an overview, refer to the review by Stein et al [13]), and adherence to pharmacological treatment is often low [14]. These disadvantages and challenges of pharmacotherapy highlight the need for easily accessible complementary treatment options for adults with ADHD. Lifestyle factors, such as nutrition and physical activity (PA), might be promising targets for the development of complementary treatments [15-17].

Nutrition, Impulsivity, and ADHD

Although it is well known that nutrition has a significant impact on physical health, evidence is growing that nutrition also plays an important role in mental health and mental functioning, wherefore the emerging field of nutritional psychiatry is attracting growing attention [18]. For instance, the Mediterranean diet seems to have protective effects against depressive symptoms [19] and may even be an effective treatment strategy for depression [20]. Increasing attention is also being paid to the role of nutrition in ADHD (for an overview, refer to the reviews by Breda et al [15] and Pinto et al [17]). Del-Ponte et al [21] conducted a meta-analysis and found that healthy dietary patterns were associated with a decreased risk for ADHD (odds ratio [OR] 0.65, 95% CI 0.44-0.97), whereas unhealthy dietary patterns were associated with an increased risk for ADHD (OR 1.41, 95% CI 1.15-1.74). The authors concluded that the findings suggest that healthy diets rich in fruits and vegetables can protect against ADHD, whereas diets high in refined sugar and saturated fat can increase the risk [21]. These findings were confirmed by a recent meta-analysis that found that healthy dietary patterns characterized by fruits, vegetables, and fish were associated with a reduced risk for ADHD (OR 0.63, 95% CI 0.41-0.96), whereas Western dietary patterns, consisting of red meat, processed meat, animal fat, and salt (OR 1.92, 95% CI 1.13-3.26), and junk food dietary patterns, including sweets, sweetened beverages, snacks, ice creams, and fast foods (OR 1.51, 95% CI 1.06-2.16), were found to be associated with an increased risk for ADHD [22]. A positive association between total sugar intake (ie, from sugar-sweetened beverages and dietary sources) and the risk of ADHD was found in another meta-analysis [23]. Although these meta-analytical findings support the presence of a link between nutrition and ADHD, most of the included studies do not allow causal conclusions. Cross-sectional studies do not provide information on the directionality of the association and lack microtemporal resolution and ecological validity. More high-quality studies (eg, randomized and microrandomized controlled trials) are needed to gain insights into causality and underlying mechanisms. As ADHD symptomology is dynamic in nature [24], studies of high temporal resolution and ecological validity, such as ecological momentary assessment (EMA) studies, are required to obtain evidence on short-term, microtemporal associations in daily life. EMA comprises repeated assessments of behaviors (eg, food intake), experiences (eg, impulsivity), and physiological parameters throughout a day in the moment and in the natural environment, which enables studying complex psychological, behavioral, and physiological processes [25].

Evidence indicates that nutrition is not only associated with ADHD per se but also with core symptoms of ADHD, such as impulsivity and executive dysfunction. Impulsivity is a multidimensional construct [26] that manifests as “impatience, acting without thinking, spending impulsively, starting new jobs and relationships on impulse, and sensation seeking behaviours” [27]. Impulsivity is closely linked to impaired executive function [28]. Accordingly, the concepts of impulsivity [28] and executive dysfunction [29] have been used to describe the same impairments in ADHD. A cross-sectional study among a nonclinical adult sample found an association between fast food consumption and greater impulsivity (ie, delay discounting, a behavioral measure of impulsivity) [30]. However, directionality remains unclear. Further studies in nonclinical samples of young adults found an association between a Western-style diet high in saturated fat and added sugar and greater trait impulsivity, with hypothesized bidirectional causation [31], and a relationship between stronger inhibitory control and lower consumption of foods high in saturated fat [32]. Research in rodents provides the first evidence on the causality of the relationship between nutrition and

impulsivity by showing that a high-fat and high-sugar diet increases impulsivity [33,34].

In contrast to fat and sugar intake, protein intake appears to have beneficial effects on executive function. A randomized controlled dietary intervention trial in healthy men found improved reaction times on the "Go/No-Go task," a measure of response inhibition and a proxy for impulsivity, in the group that followed a high-protein diet for 3 weeks [35]. Tryptophan, an essential amino acid and precursor to serotonin found in almost all proteins, has been discussed as a potential underlying mechanism. For instance, it was found that dietary intake of tryptophan was associated with lower levels of one facet of emotion-related impulsivity (ie, Pervasive Influence of Feelings) [36].

Although many studies have assessed the relationship between ADHD symptomology or impulsivity and long-term dietary exposure (eg, habitual food intake captured through food frequency questionnaires [FFQs]), there is also some evidence that the intake of certain macronutrients has short-term effects on executive function. For instance, Brandley and Holton [37] showed that a nutritionally balanced breakfast with an optimal macronutrients ratio (ie, 25% fat, 45% carbohydrate, and 29% protein) improved executive function of college students with and without ADHD an hour after consumption.

PA, Impulsivity, and ADHD

Not only nutrition but also PA seems to influence impulsivity and executive function as PA is associated with reduced impulsivity. A meta-analysis found a significant overall effect of acute, but not chronic physical exercise on executive functions in children, adolescents, and young adults [38]. Several meta-analyses have studied the effect of exercise on cognition and behavior in children with ADHD showing that exercise has positive effects on executive functions [39-41]. Research assessing the effect of PA in adults with ADHD remains limited [39,42]. However, the initial findings are promising. For instance, a cross-sectional pilot study found that adults with ADHD who engaged in frequent aerobic PA reported significantly lower levels of behavioral impulsivity [43]. In addition, the first evidence indicates that noncardio PA (ie, whole body vibration) has positive effects on cognitive functioning in adults with ADHD [44,45]. A study in college students with and without ADHD showed that in those with ADHD inhibitory performance, one aspect of executive function improved after acute exercise; however, all aspects of executive functions improved in those without ADHD [46]. In a study following a counterbalanced repeated measures design with a control condition, adults with ADHD improved reaction times in congruent and incongruent trials of the flanker task after 30 minutes of continuous stationary cycling, indicating benefits of acute exercise [47].

Objectives

ADHD is characterized by dynamic symptoms that manifest as states of hyperactivity, inattention, and impulsivity [24]. However, most studies rely on cross-sectional data that lack microtemporal resolution and do not provide evidence with regard to true temporal associations. Retrospective reports (eg, FFQs and trait questionnaires) are prone to cognitive biases—potentially, in particular, in individuals with ADHD—and overlook fluctuations of ADHD symptoms [48]. It remains unstudied whether the intake of certain macronutrients or the engagement in PA alters impulsivity within minutes or hours in everyday life. EMA can bridge this gap, as it offers great potential to provide novel insights into ADHD symptomatology in daily life to better understand behavior and functioning at the intraindividual level [24,48]. Despite the need for studies of high temporal resolution and ecological validity and the potential of EMA, so far, no study has applied EMA to assess whether the intake of certain macronutrients and engagement in PA are associated with changes in state impulsivity in daily life. Therefore, the aim of this study was to apply EMA to assess short-term, microtemporal associations between macronutrient intake, PA, and state impulsivity in daily life of adults with and without ADHD. On the basis of previous research, six research questions were elaborated. We examined whether (1) the intake of sugar and (2) the consumption of saturated fats are associated with an increase in state impulsivity, and whether (3) the intake of proteins and (4) the engagement in PA are associated with decreased state impulsivity. As there is evidence that the combination of a high-fat and high-sugar diet was associated with ADHD and impulsivity [21,31,33], we assessed whether (5) the combined intake (ie, the interaction) of saturated fat and sugar amplifies the increase in state impulsivity. Intriguingly, research indicates that PA may buffer against the adverse effects of fat intake on cognitive functioning [49,50]. Therefore, we investigated whether (6) PA alleviates the positive association between sugar or fat intake and state impulsivity.

Methods

Procedure

Data were collected within the Eat2beNICE-APPetite study, which comprises 2 in-person sessions as well as an EMA period (parts of the data of this study have been used for different research questions [51-53]). In the first in-person session, participants completed questionnaires and received comprehensive training to familiarize them with the APPetite mobile app, which was used for the EMA period (refer to the study by Ruf et al [51] for further details). Body weight and height were measured to calculate BMI.

Ethics Approval

The local ethics committee of the faculty of medicine of the Goethe University Frankfurt (Ethikkommission des Fachbereichs Medizin der Goethe-Universität) approved the study (reference number: 192/18).

Informed Consent

All participants declared that they understood the study procedure and signed a written informed consent. This study was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki, 1975).

EMA Protocol

The participants used a study smartphone to complete the EMA protocol of the APPetite mobile app for 3 consecutive days (2 weekdays and 1 weekend day). Participants received 8 semirandom signal-contingent prompts per day (between 8 AM and 10 PM, with at least 1 hour in-between prompts). Each prompt assessed state impulsivity. The prompts could be postponed for up to 25 minutes. Participants were able to record food intake at any time (ie, event-contingent) through the incorporated APPetite-food record. In addition, a time-contingent prompt at 9 PM asked participants whether all foods and drinks consumed during the day were recorded. Further details on the APPetite mobile app can be found in the study by Ruf et al [51].

Sample

The Eat2beNICE-APPetite study recruited participants from 4 existing study cohorts. Adults with ADHD were invited from (1) the PROUD (Prevention of Comorbid Depression and Obesity in Attention-Deficit/Hyperactivity Disorder) study [54]; (2) the BipoLife-A1 study that follows up individuals with an increased risk for bipolar disorders, including patients affected by ADHD or depression (or both) [55,56]; and (3) the PROBIA (Treating Impulsivity with Probiotics in Adults) study, which recruited patients with ADHD or borderline personality disorder (or both) [57]. Healthy controls were recruited from the Longitudinal Resilience Assessment study, which enrolled individuals not affected by psychiatric conditions and followed them up since 2016 [58].

In total, 43 adults with ADHD and 185 adults without ADHD participated in the study. After the first in-person session, 4 participants without ADHD dropped out because of personal reasons (eg, spontaneous vacation) or the inability to respond to prompts (eg, because of work commitments). Data of 1 participant without ADHD were excluded as they proved to be untrue. Data of 26 participants without ADHD and 6 participants with ADHD were excluded because of incomplete records of food intake (eg, only 1 meal recorded). One participant in the ADHD sample had to be excluded because no PA data could be retrieved because of technical problems. One participant without ADHD was excluded because BMI was unavailable owing to scale malfunction. Furthermore, 16 participants without ADHD were excluded from the analyses because they showed no variation in state impulsivity across prompts (ie, all Momentary Impulsivity Scale [MIS] items consistently answered with *1—not applicable*). The final sample included 36 participants with ADHD and 137 participants without ADHD. Demographics of the ADHD and control groups are presented in Table 1. The samples differed in terms of gender (χ^2_1=4.5, *P*=.03), age (Mann-Whitney *U*=1667; *P*=.002), and BMI (kg/m^2; Mann-Whitney *U*=1433; *P*<.001).

Table 1. Demographics of the sample with and without attention-deficit/hyperactivity disorder (ADHD).

	ADHD sample (n=36)	Control sample (n=137)
Gender, n (%)		
Woman	19 (53)	100 (73)
Man	17 (47)	37 (27)
Age (years), mean (SD)	35.25 (12.04)	28.8 (7.72)
BMI (kg/m^2), mean (SD)	29.06 (7.87)	24.08 (4.14)

Measures

Macronutrient Intake

Macronutrient intake was captured using the APPetite mobile app, which comprises a food record [51]. The food recording follows a six-step process: (1) selection of meal type, (2) entry of time of intake, (3) selection of consumed foods and drinks, (4) specification of consumed amounts, (5) presentation of reminders for commonly forgotten foods, and (6) indication of the predominant reason for eating or drinking. Participants were instructed to record their food and drink as soon as possible after consuming them. To generate nutritional values (ie, sugar, saturated fat, and protein intake), the collected food entries were transferred to myfood24-Germany [59] by trained staff. A feasibility, usability, and validation study was conducted to evaluate the APPetite mobile app. The findings indicated that the APPetite mobile app is a feasible and valid dietary assessment tool that is more accurate compared with 24-hour recalls [51].

State Impulsivity

State impulsivity was assessed using the MIS [60]. The MIS captures state impulsivity on the basis of 4 items, each of which comprises a statement (eg, “I said things without thinking”). Participants rated how well each statement described their behavior, cognition, and experiences since the last prompt or since waking up in the first daily prompt on a 5-point scale. A sum score of the items was calculated. Higher values indicate greater state impulsivity. Note that the original response scale (1=very slightly or not at all, 2=a little, 3=moderately, 4=quite a bit, and 5=extremely) was slightly altered during translation (translate–back-translate procedure with a native bilingual speaker in English and German) because the literal translation lacked differentiability. One main difference between the English and German version is that *1* on the response scale stands for *not applicable* in the German version (1=nicht zutreffend ["not applicable" in German], 2=eher nicht zutreffend ["partially not applicable" in German], 3=teils-teils ["half and half" in German], 4=eher zutreffend ["partially applicable" in German], and 5=zutreffend ["applicable" in German]). In the final data sets, the McDonald ω of the MIS was 0.576 (within)

and 0.832 (between) in the ADHD sample and 0.505 (within) and 0.768 (between) in the control sample.

Trait Impulsivity

The UPPS-P Impulsive Behavior Scale [61,62] was used to assess trait impulsivity based on 59 items. Each item described a statement (eg, "I have trouble controlling my impulses"). Participants reported how well each statement described them on a 4-point scale from *agree strongly* to *disagree strongly*. The UPPS-P Impulsive Behavior Scale assesses impulsivity as a multifaceted construct that includes the following subscales: negative urgency (12 items), positive urgency (14 items), lack of premeditation (11 items), lack of perseverance (10 items), and sensation seeking (12 items). The German translation of the items for the scales negative urgency, lack of premeditation, lack of perseverance, and sensation seeking were taken from Schmidt et al [63]. The items of the subscale positive urgency were translated through the translate-back-translate procedure. In this study's samples, internal consistency was ADHD, α=.92 and control, α=.89 for negative urgency; ADHD, α=.81 and control, α=.75 for premeditation; ADHD, α=.83 and control, α=.83 for perseverance; ADHD, α=.89 and control, α=.85 for sensation seeking; and ADHD, α=.93 and control, α=.92 for positive urgency.

Physical Activity

PA was captured objectively using Move 3 sensors (movisens GmbH). Participants wore the sensor on their nondominant wrist during the EMA period (day and night). The software DataAnalyzer (version 1.13.7; movisens GmbH) was used to calculate the movement acceleration intensity per minute (mg/min) from raw accelerometry. Nonwear time was excluded from the analysis.

Data Preprocessing

A total of 13 single days of the control sample had to be excluded owing to incomplete dietary data. Data preprocessing was completed based on the time intervals for which state impulsivity was assessed (ie, time between current prompt and previous prompt or waking up). To study the association between macronutrient intake, PA, and state impulsivity, each of these time interval was matched to concurrent sugar, saturated fat, and protein intake and mean PA (ie, mean movement acceleration). Concurrent intake was defined as the sum of any intake of sugar, saturated fat, or protein within the respective time interval. On the basis of the movement acceleration intensity per minute (mg/min), the mean movement acceleration was calculated for each time interval in which the sensor was worn for at least two-thirds of the time.

The level-1 predictors sugar, fat, and protein intake and PA were person mean centered to produce unbiased estimates of the within-person effect [64]. To avoid estimation problems owing to substantial differences in the variance of the predictors and the outcome, the level-1 predictors were divided by 10. The level-2 covariates age and BMI were centered at 30 and 25, respectively, to make the model intercept more interpretable as recommended by Viechtbauer [65]. Grand mean centering was used for level-2 covariate trait impulsivity. The level-2 covariate gender was coded as 0 (male) and 1 (female).

The MIS items were not completed at 401 time intervals (control: n=297; ADHD: n=104) and were therefore excluded. Owing to the semirandom sampling protocol, the time intervals varied in length. Beyond that, the option to postpone prompts and the assessment of state impulsivity "since waking up" in the first prompt produced rather short or long time intervals. As we did not expect an effect of macronutrient intake and PA on state impulsivity within <15 minutes [66] and were interested in short-term associations, time intervals shorter than 15 minutes (control: n=21; ADHD: n=8) and longer than 3 hours (control: n=118; ADHD: n=27) were excluded. In addition, time intervals in which the level-1 predictor PA was not available (eg, because of the exclusion criteria for time intervals in which the sensor was worn <two-thirds of the time), were excluded (control: n=166; ADHD: n=73). The final data set included 629 time intervals in the ADHD and 2464 in the control sample and is provided in Multimedia Appendix 1.

Data Analysis

Owing to the nested data structure (time intervals [level 1] nested within individuals [level 2]), multilevel models were needed for the analyses. The MIS score, the outcome of this study, showed a strongly right-skewed distribution that did not meet the assumptions of linear multilevel modeling. Furthermore, owing to a significant proportion of the lowest MIS score (all items answered with 1 *not applicable*, resulting in an MIS score of 4), a gamma multilevel model was unable to represent the right skew in the data adequately. To account for the inflation of the lowest MIS score (ie, 4), which represents the absence of impulsivity, we used a multilevel 2-part model that allows to account for zero-inflated, continuous data (ie, semicontinuous data). This type of model allows studying whether the intake of certain macronutrients and PA is associated with the occurrence of state impulsivity (ie, is an individual impulsive at all?) and the intensity of impulsivity (ie, if an individual is impulsive, how impulsive are they?). To move the inflation from 4 to 0, the 5-point scale of the MIS was recoded (1 to 0, 2 to 1, 3 to 3, 4 to 3, and 5 to 4). The model we applied combines a multilevel logistic regression in the zero part to study the occurrence of state impulsivity and a multilevel gamma regression (to account for the right skew in the positive values) in the continuous part of the model to assess the intensity of state impulsivity. The model does not only allow to study the occurrence and intensity of state impulsivity but also accounts for the potential dependency between the 2 outcome components by modeling a cross-part correlation. Although logistic regressions typically predict the outcome to be 1, the multilevel logistic regression in the zero part of the model used in this study predicts no impulsivity (outcome=0), that is, the probability not to be impulsive in a given individual in a given time interval.

To examine the associations between macronutrient intake, PA, and state impulsivity (research question 1 to 4), a model with the level-1 predictors sugar, saturated fat, and protein intake and PA in both model parts (ie, the logistic regression as well as the gamma regression) was run. A joint model was chosen to control for the other predictors, as the intake of different macronutrients and PA naturally does not occur in isolation, but in combination. Next, the interaction between the level-1

predictors sugar and fat intake was added to both model parts (research question 5). Finally, a model, including the interaction between sugar intake, fat intake, and PA (ie, 4 interactions modeled: 2-way interactions between sugar intake and PA, between fat intake and PA, between sugar and fat intake, and 3-way interaction among sugar intake, fat intake, and PA) in both model parts was run (research question 6). In all models, the level-2 covariates gender, age, BMI, and trait impulsivity were included. All models included random intercepts in both model parts (ie, we expect individuals to differ in their average probability not to be impulsive and the average intensity of state impulsivity) and random slopes for all level-1 predictors (and their interaction) to examine whether the effects differ between individuals. The 3 models were run separately for the ADHD and the control group.

All models were estimated using the R-package *brms* [67,68], which supports Bayesian multilevel modeling. Credible intervals (95% CI) of fixed effects that do not include 0 were interpreted as significant effects. As nonpositive estimates for SDs are not allowed, the lower limit of the CI of random effects that are equal to 0 suggest that the random effect is not significant (ie, that individual differences in the intercept, the effects of the level-1 predictors, or the interactions between level-1 predictors are small and possibly not statistically meaningful). Details on the model used in this study (eg, implementation and interpretation) can be found in the study by Ruf et al [52].

The estimation of model parameters was based on 10,000 iterations. The initial values for the sampler were set to 0 (*init=0*) and the maximum tree depth was set to 11 to reach convergence in 2 models (see the open R code provided in Multimedia Appendix 2). The default settings of all other sampling and prior parameters were maintained. R (version 4.2.2; R Foundation for Statistical Computing [69]), RStudio (version 2022.7.2.576; Posit [70]), *brms* (version 2.18.0), and *rstan* (version 2.26.13) [71] were used to perform the analyses.

Results

Descriptive Findings

Descriptive statistics of the level-1 predictors sugar, saturated fat, and protein intake and PA and the level-2 covariate trait impulsivity for the ADHD sample and the control sample are shown in Table 2. Participants reported not to be impulsive (MIS=0) in 28.6% (180/629) of the time intervals in the ADHD sample and in 58.2% (1434/2464) of the time intervals in the control sample. Within the time intervals in which participants reported to be impulsive (ADHD: n=449; control: n=1030), state impulsivity was rated on average 3.9 (SD 2.6) in the ADHD sample and 2.8 (SD 1.9) in the control sample on the shifted response scale (ranging from 0 to 16). To test whether individuals with and without ADHD differed in the occurrence and the intensity of impulsivity, a multilevel 2-part model with the level-2 predictor ADHD diagnosis (0=no ADHD, 1=ADHD) including all participants was calculated (see model 0 in the open R code provided in Multimedia Appendix 2). Results showed that individuals with ADHD were less likely not to be impulsive (ie, significant fixed effect of ADHD diagnosis in the zero part: −1.69, SE 0.32, 95% CI −2.34 to −1.06) and reported significantly higher levels of impulsivity intensity (ie, significant fixed effect of ADHD diagnosis in the continuous part: 0.36, SE 0.08, 95% CI 0.20-0.51) compared with individuals without ADHD.

Average compliance with the signal-contingent prompts (ie, percentage of complete prompts within received prompts) was 89.6 (SD 12.4) in the ADHD and 90.3 (SD 11.5) in the control sample (not including participants and days that were excluded as a whole, but including time intervals that were excluded from the final analyses based on interval length and missing PA or MIS).

Table 2. Descriptive statistics of the level-1 predictors (control: n=2464; ADHD: n=629) and level-2 covariate trait impulsivity (control: n=137; ADHD: n=36).

	ADHD[a] sample		Control sample	
	Values, mean (SD)	Values, median[b] (range)	Values, mean (SD)	Values, median (range)
Level 1				
Sugar intake in g	10.19 (17.50 [overall]; 5.71 [between])	0 (0-101.05)	9.87 (15.94 [overall]; 4.41 [between])	1.04 (0-186.67)
Saturated fat intake in g	4.46 (8.97 [overall]; 2.39 [between])	0 (0-76.15)	4.47 (7.92 [overall]; 2.11 [between])	0.03 (0-65.23)
Protein intake in g	10.42 (20.34 [overall]; 4.95 [between])	0 (0-145.2)	9.93 (16.73 [overall]; 4.13 [between])	0.64 (0-137.09)
PA[c] (acceleration) in mg	130.85 (77.07 [overall]; 38.04 [between])	120.63 (8.64-815.3)	144.68 (78.77 [overall]; 31.68 [between])	134.13 (10.23-1445.94)
Level 2				
Trait impulsivity	30.24 (5.08)	30.1 (18.2-40.0)	23.78 (3.68)	23.4 (16.4-35.8)

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]Median values were included to highlight that the level-1 predictors followed skewed distributions and were zero inflated (ie, food intake did not occur within each time interval, where macronutrient intake is equal to 0).

[c]PA: physical activity.

Findings From the Multilevel 2-Part Models

Interpretation Overview

As the estimates of the continuous part (ie, the gamma regression) of the multilevel 2-part model are modeled on the log scale, the exponential is used to obtain estimates in the original metric. In the zero part (ie, the logistic regression), estimates were modeled on the logit scale. The intercept of the zero part represents the average log-odds of no impulsivity across all participants when all the predictors were 0. The inverse logit function (eg, the *plogis* function in R) can be used to transform the log-odds to the probability not to be impulsive. The predictor estimates in the zero part represent the expected change in the log-odds of no impulsivity for a 1-unit increase in each predictor, respectively. To obtain the expected change in the probability not to be impulsive, the probability of the intercept (ie, *plogis*[intercept]) can be compared with the predicted probability when the respective predictor takes on a certain value (eg, if the chosen value of the predictor is 1, the predicted probability is *plogis*[intercept + fixed effect of the predictor]).

Sugar, Saturated Fat, and Protein Intake; PA; and State Impulsivity

Results of the model, including fixed and random effects for sugar, saturated fat, and protein intake and PA in both model parts are shown in Table 3 for the ADHD sample. The intercept of the zero part indicates that the mean probability not to be impulsive is 11.7% (*plogis*[−2.02]) when all predictors and covariates are equal to 0. Sugar, saturated fat, and protein intake had no significant fixed effect on the probability not to be impulsive. However, the effect of saturated fat and protein intake on the probability of no impulsivity differs across individuals with an *SD* of 0.29 and 0.22, respectively. In time intervals in which PA is 1 unit (ie, 10 mg) above 0 (ie, above the person mean) and all other predictors are 0, the probability not to be impulsive is 10.8% (*plogis*[−2.02−0.09]), that is, a 1-unit increase in PA is associated with a decrease in the probability not to be impulsive of 0.9% (11.7%−10.8%=0.9%) when all other predictors are 0. This indicates that higher levels of PA are associated with a higher probability to be impulsive. The intercept of the continuous part of the model demonstrates that when all predictors and covariates are equal to 0, participants with ADHD report an average impulsivity intensity of 3.32 ($e^{1.20}$). There was no significant fixed effect of sugar, saturated fat, and protein intake and PA. Accordingly, the intake of sugar, saturated fat, and protein intake and the level of PA were not associated with the intensity of state impulsivity. There was a negative cross-part correlation (−0.38, SE 0.17, 95% CI −0.68 to −0.03) indicating that individuals who are impulsive more often are more impulsive when they are impulsive suggesting that the frequency and intensity of impulsivity correlate.

Table 3. Model estimates of the multilevel 2-part model including fixed and random effects for sugar, saturated fat, and protein intake as well as physical activity (PA) in both model parts in the attention-deficit/hyperactivity disorder sample.

Model 1	Zero part, estimate (SE; 95% CI)	Continuous part, estimate (SE; 95% CI)
Fixed effects		
Intercept	−2.02 (0.55; −3.14 to −0.97)	1.20 (0.12; 0.96 to 1.44)
Sugar intake	−0.01 (0.10; −0.22 to 0.18)	0.01 (0.02; −0.03 to 0.04)
Saturated fat intake	−0.10 (0.29; −0.70 to 0.45)	0.01 (0.04; −0.07 to 0.10)
Protein intake	−0.09 (0.14; −0.38 to 0.16)	−0.02 (0.02; −0.05 to 0.01)
PA	−0.09 (0.03; −0.14 to −0.04)	0.00 (0.00; −0.00 to 0.01)
Gender	0.40 (0.74; −1.04 to 1.89)	−0.15 (0.17; −0.48 to 0.20)
Age (years)	−0.02 (0.04; −0.09 to 0.05)	0.00 (0.01; −0.01 to 0.02)
BMI (kg/m^2)	0.11 (0.05; 0.01 to 0.20)	0.00 (0.01; −0.02 to 0.02)
Trait impulsivity	−0.11 (0.07; −0.24 to 0.03)	0.03 (0.02; 0.00 to 0.07)
Random effects		
SD (intercept)	1.83 (0.32; 1.29 to 2.55)	0.44 (0.06; 0.33 to 0.58)
SD (sugar intake)	0.12 (0.09; 0.00 to 0.34)	0.03 (0.02; 0.00 to 0.07)
SD (saturated fat intake)	0.29 (0.24; 0.01 to 0.89)	0.04 (0.03; 0.00 to 0.11)
SD (protein intake)	0.22 (0.14; 0.01 to 0.55)	0.02 (0.01; 0.00 to 0.04)
SD (PA)	0.06 (0.04; 0.00 to 0.14)	0.01 (0.01; 0.00 to 0.03)

Table 4 shows the results of the same model (ie, fixed and random effects for sugar, saturated fat, and protein intake and PA in both model parts) for the control sample. Individuals without ADHD had a mean probability not to be impulsive of 69.2% (*plogis*[0.81]) when all predictors and covariates were equal to 0. Similar to the ADHD sample, sugar, saturated fat, and protein intake had no significant fixed effect on the probability not to be impulsive. However, in contrast to the ADHD sample, the effects of saturated fat and protein intake on the probability not to be impulsive did not differ among individuals without ADHD. However, higher levels of PA were also associated with a higher probability to be impulsive in

adults without ADHD. Accordingly, the probability not to be impulsive is 68.6% (*plogis*[0.81−0.03]) in time intervals in which PA is 1 unit (ie, 10 mg) above 0 (ie, above the person mean) and all other predictors are 0. Consequently, a 1-unit increase in PA is associated with a decrease in the probability not to be impulsive by 0.6% (69.2%−68.6%=0.6%) when all other predictors in the model are 0. Again, the cross-part correlation (−0.53, SE 0.09, 95% CI −0.69 to −0.33) was negative, indicating that individuals who were impulsive more often were more impulsive when they were impulsive.

Table 4. Model estimates of the multilevel 2-part model, including fixed and random effects for sugar, saturated fat, and protein intake as well as physical activity (PA) in both model parts in the control sample.

Model 2	Zero part, estimate (SE; 95% CI)	Continuous part, estimate (SE; 95% CI)
Fixed effects		
Intercept	0.81 (0.27; 0.28 to 1.35)	0.68 (0.07; 0.54 to 0.83)
Sugar intake	−0.02 (0.04; −0.10 to 0.06)	−0.01 (0.01; −0.04 to 0.02)
Saturated fat intake	−0.04 (0.11; −0.25 to 0.18)	0.00 (0.04; −0.07 to 0.08)
Protein intake	−0.04 (0.05; −0.13 to 0.06)	−0.02 (0.02; −0.05 to 0.02)
PA	−0.03 (0.01; −0.05 to −0.01)	0.01 (0.00; 0.00 to 0.02)
Gender	−0.52 (0.33; −1.17 to 0.12)	0.17 (0.09; −0.00 to 0.34)
Age (years)	−0.02 (0.02; −0.05 to 0.02)	−0.00 (0.00; −0.01 to 0.01)
BMI (kg/m^2)	0.03 (0.04; −0.04 to 0.11)	0.00 (0.01; −0.02 to 0.02)
Trait impulsivity	−0.20 (0.04; −0.27 to −0.13)	0.05 (0.01; 0.03 to 0.07)
Random effects		
SD (intercept)	1.44 (0.12; 1.22 to 1.69)	0.33 (0.03; 0.27 to 0.39)
SD (sugar intake)	0.06 (0.04; 0.00 to 0.17)	0.03 (0.02; 0.00 to 0.06)
SD (saturated fat intake)	0.14 (0.10; 0.00 to 0.38)	0.05 (0.04; 0.00 to 0.14)
SD (protein intake)	0.05 (0.04; 0.00 to 0.15)	0.02 (0.02; 0.00 to 0.06)
SD (PA)	0.07 (0.02; 0.04 to 0.10)	0.01 (0.01; 0.00 to 0.02)

Interaction Between Sugar and Saturated Fat Intake

To study whether the combined intake of saturated fat and sugar amplifies the effect of sugar and saturated fat intake on state impulsivity, the interaction between the level-1 predictors sugar and fat intake was added to both models. Results of the ADHD sample are shown in Table S1 in Multimedia Appendix 3 and results of the control sample are shown in Table S2 in Multimedia Appendix 3. In both samples and both model parts, the interaction between sugar and saturated fat intake was not significant. However, in the ADHD sample, the interaction effect between sugar and saturated fat intake on the probability not to be impulsive varied across participants with an SD of 0.25 in the zero part.

Buffering Effect of PA on the Association Between Sugar and Fat Intake and State Impulsivity

To test whether PA alleviates the positive association between sugar or fat intake and state impulsivity, a model that included the interaction between sugar intake, fat intake, and PA (ie, 4 interactions: 2-way interactions between sugar intake and PA, between fat intake and PA, between sugar and fat intake, and 3-way interaction among sugar intake, fat intake, and PA) in both model parts was run for each sample. Table S3 in Multimedia Appendix 3 shows the results of the ADHD sample and Table S4 in Multimedia Appendix 3 shows the results of the control sample. In both samples, the 4 interactions in the zero and the continuous part of the model were not significant. Only in the ADHD sample, the interaction effect of sugar and saturated fat intake (as in the previous model) and the 3-way interaction among sugar and saturated fat intake and PA varied significantly between participants with ADHD.

Discussion

Principal Findings

Although impulsivity, a core symptom of ADHD, can contribute to the disruption of daily functioning, the first evidence indicates that the intake of certain macronutrients and the engagement in PA might alter impulsivity and executive function. However, despite the potential of digital and mobile technologies, studies of high temporal resolution and ecological validity are lacking, and it remains unanswered whether the intake of certain macronutrients and the engagement in PA are associated with short-term changes in state impulsivity in everyday life. Therefore, this study applied EMA to assess short-term, microtemporal dynamics of macronutrient intake, PA, and state impulsivity in daily life of adults with and without ADHD. Contrary to previous findings that suggest that the intake of sugar and saturated fat is associated with greater impulsivity [30-34], whereas the intake of proteins is linked to decreased impulsivity and improved executive function [35,36], no association between macronutrient intake and state impulsivity (ie, the probability to be impulsive and the intensity of

impulsivity) was found in this study. However, some between-person variability was observed. Furthermore, in contrast to prior research indicating that PA is associated with reduced impulsivity [38-40], no relationship between PA and the intensity of impulsivity was found and PA was associated with an increased probability to be impulsive in both samples. No evidence was found that the combined intake of saturated fat and sugar amplified the increase in state impulsivity and that PA alleviated the positive association between sugar or fat intake and state impulsivity.

One reason for not finding an association between macronutrient intake and state impulsivity in this study could be the (varying) length of the time intervals in which the association was studied. In this context, several methodological considerations need to be discussed. (1) Assessment of state impulsivity: as 3 items of the MIS describe specific actions ("I said things without thinking," "I spent more money than I meant to," "I made a 'spur of the moment' decision"), the assessment has to be based on time intervals (here "since the last prompt"). Only the item "I have felt impatient" could be adapted to allow an assessment on the momentary level (ie, "Right now I feel impatient"). A momentary assessment would allow to specify time intervals before the impulsivity assessment more flexibly to study the temporal sequence of the association. (2) Lengths of time intervals: a semirandom signal-contingent EMA protocol was used to assess state impulsivity, so that participants could not predict the exact time of the next prompt. This allowed capturing a better reflection of the participants' daily lives [72]. However, it results in time intervals of different lengths (with at least 1 hour in-between 2 prompts in this study). However, the first daily assessment of state impulsivity "since waking up" and the postponement of prompts led to even shorter or rather long time intervals. Consequently, the length of time intervals varied considerably. Therefore, time intervals shorter than 15 minutes and longer than 3 hours were excluded from analysis. (3) Temporal manifestation of the effect of macronutrients: research providing evidence regarding the time frame in which macronutrients affect cognition and behavior is still lacking. Beyond that, the time frame in which the effects occur might differ across macronutrients. As the assessment of impulsivity was based on time intervals, this study was restricted to assessing the association between state impulsivity and macronutrient intake within each time interval for which impulsivity was assessed (ie, impulsivity and concurrent macronutrient intake, not intake and subsequent impulsivity). Considering time intervals were of rather different length (15 min to 3 h), the approach of this study might overlook the effect of macronutrient intake. As it is not yet clear within which time frame specific macronutrients affect cognition and behavior, controlled studies are needed to establish the temporal manifestation of the effects of macronutrients on impulsivity. This knowledge is required to adjust EMA protocols to study the relationship between macronutrient, PA, and state impulsivity more systematically. For instance, a time or interval-contingent assessment of state impulsivity could be considered to obtain time intervals of similar length (eg, prompts every hour) [72]. The intervals between prompts should be determined on basis of the novel input and evidence that controlled studies can offer in terms of the temporal manifestation of short-term macronutrients' effects on impulsivity.

The operationalization of state impulsivity is also important in this context. This study used a self-report measure to capture state impulsivity [60]. However, it remains unclear whether macronutrient intake alters impulsivity, or whether the effect of macronutrients might only affect subdomains (eg, response inhibition). Future studies should include impulsivity-related behavioral measures (eg, stop-signal task, the Digital Marshmallow Test [73], and mobile Balloon Analogue Risk Task [74]), which also open up the opportunity to assess impulsivity momentarily. The momentary assessment of impulsivity, in turn, enables to assess the temporal sequence of the association. In addition, it can contribute to establishing time windows within which macronutrients might affect state impulsivity, as the length of time intervals, within which the association is studied, can be set independently.

Although no overall association between macronutrient intake and state impulsivity was found, some variation (ie, individual differences) in the effect of certain macronutrients was observed, particularly in the ADHD sample. Muth and Park [75] pointed out that age, PA, and glucose metabolism are potential factors contributing to individual differences in the macronutrient-cognition relationship. In this study, PA did not moderate the relationship between sugar and fat intake and state impulsivity. However, future studies should aim at identifying factors that moderate the macronutrient-impulsivity relationship. In this context, continuous glucose monitoring could be a feasible and promising addition to future EMA studies.

In contrast to the expectation, PA was not associated with reduced impulsivity (ie, lower probability to be impulsive and decreased intensity of impulsivity), but with an increase in the probability to be impulsive in adults with and without ADHD. However, it is important to note that temporal associations do not reflect causality. For instance, an individual might experience a state of impulsivity and, consequently, actively decide to engage in PA as a counteracting measure, resulting in a positive association between PA and the probability to be impulsive. Hence, the observed association does not imply that engaging in PA leads to a higher probability to be impulsive. Furthermore, these findings might be a result of the operationalization of PA. Using the mean acceleration within the time intervals might average out relevant fluctuations in PA intensity. Hence, it cannot be differentiated between time intervals in which an individual shows low levels of PA continuously and time intervals in which an individual engages in some moderate-to-vigorous PA but is inactive for the remainder of the time. However, acute PA has been shown to have positive effects on executive function [76-80], whereas taking the mean PA might overlook the association between PA and state impulsivity. Because the time intervals differed considerably in length, the mean acceleration was the most straightforward operationalization of PA in this study. Future EMA studies should consider implementing a sampling approach that yields time intervals of similar length (see *Discussion* section) to allow the operationalization of PA as minutes engaged in different PA intensities (eg, minutes of moderate-to-vigorous PA). In addition, the differentiation

between exercise and nonexercise PA could provide further insights, as found for the effect of PA on mood [81]. A further explanation for the positive association between PA and the probability to be impulsive could be the nature of MIS items. For instance, being impatient could manifest as walking up and down or twiddling with something. Spending more money than intended might be more likely to occur when being out and about (eg, walking around the city). This further highlights the importance of differentiating between different PA intensities and between exercise and nonexercise PA in future studies.

Strengths and Limitations

This study is, to the best of our knowledge, the first to assess the relationship between macronutrient intake, PA, and state impulsivity in daily life using EMA. This kind of research is innovative and highly important to better understand fluctuations of ADHD symptomology in daily life and provides novel evidence of high temporal resolution and ecological validity, which is highly relevant to the growing field of Nutritional Psychiatry. However, the findings of this study should be interpreted in the light of some limitations. First, the assessment of impulsivity as well as food intake is based on self-reports. However, participants might be less likely to report foods and drinks and respond to prompts when being (more) impulsive, which might have caused some bias (ie, systematic noncompliance). Therefore, more objective assessments of dietary intake (eg, passive detection of eating events [82] and automatized photo-based dietary assessment) and impulsivity (eg, passive detection of impulsive behavior [83]) are desirable. However, although self-reports are generally prone to bias, particularly self-reports of food intake [84], assessing food intake in real time or near real time, as done in this study, minimizes recall biases compared with typically used retrospective dietary assessments (eg, FFQs). In addition, the smartphone app used for the dietary assessment in this study, the APPetite mobile app, was subject to a validation study, which showed that the app assessed food intake more accurately compared with widely used 24-hour recalls in healthy participants [51]. Second, this study did not allow the establishment of temporal sequences or causal relationships. Although EMA studies are essential to shed light on ecologically valid microtemporal associations between macronutrient intake, PA, and state impulsivity, complementary controlled studies are needed to gain insights into the directionality of the short-term effects of macronutrients and PA on state impulsivity. Third, the EMA period comprised only 3 days, which might not be sufficient to capture the complete spectrum of the associations of interest. However, given the high burden and time investment of the food recording, longer assessment periods would not have been feasible. Advanced dietary assessment methods for the naturalistic setting, such as wearable sensors passively detecting eating behavior and accurate automatized photo-based assessment of macronutrient composition, which require minimal user interaction, are required to allow prolonged EMA periods.

The strengths of this study were (1) the objective assessment of PA, (2) the application of sophisticated statistical models (ie, multilevel 2-part models), and (3) the inclusion of adults with and without ADHD. For instance, Abramovitch et al [43] used a self-report questionnaire to capture PA, although 2 systematic reviews indicated that indirect measures of PA (ie, self-reports) are unsatisfactory given that they differ substantially from direct, objective measures, such as accelerometers [85,86]. As impulsivity is widely understood as a characteristic that everyone shows some degree of—with clinical samples such as individuals with ADHD showing particularly high levels—it is especially interesting to study the association between macronutrient intake, PA, and state impulsivity in a clinical and control samples.

Conclusions

This study is the first to apply EMA to assess short-term, microtemporal associations between macronutrient intake, PA, and state impulsivity in everyday life of adults with and without ADHD. Although EMA research in the context of Nutritional Psychiatry is still in its infancy, it is undeniably a highly promising and innovative approach to gain insights into microtemporal dynamics of psychiatric symptomology and lifestyle behaviors in daily life. This study provides and discusses important methodological considerations that can help advance the field and contribute to the optimization and tailoring of future EMA protocols. In addition, the findings from EMA studies can help build the foundation for the development of just-in-time adaptive interventions. This type of intervention represents a key element of digital psychiatry as it provides personalized support in daily life of patients right at the time it is needed most.

Acknowledgments

The authors thank Simone Demmel, Hiba El Jomaa, Jeanne Julia Flemming, Hendrik Friedrichsen, Maryam Golzarnia, Krystyna Kotthaus, Emina Ricciardi, Annebirth Steinmann, Söri Wenz, Maike Wisker, and Luise Wortmann for their contribution to data collection of the APPetite study.

This study was supported by the European Union's Horizon 2020 Research and Innovation Program under grant 728018. The funding source has had no involvement in the study design, data collection, interpretation of the findings, or writing of this manuscript.

Data Availability

The data and R code that support the findings of this study are available in the Multimedia Appendices 1 and 2 of this paper.

Authors' Contributions

A Ruf, EDK, UE-P, A Reif, and SM designed and planned the study. A Ruf contributed to data collection, conducted the data preprocessing and data analyses, and wrote the first draft of the manuscript. A Ruf and ABN contributed to data interpretation. A Ruf, ABN, EDK, UE-P, A Reif, and SM critically reviewed, edited, and approved the final manuscript.

Conflicts of Interest

UE-P is a consultant for Boehringer-Ingelheim and speaker honorarium from Angelini Pharma, both of which had no influence over the content of this paper. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Open data.
[ZIP File (Zip Archive), 67 KB - mental_v10i1e46550_app1.zip]

Multimedia Appendix 2
Open R code.
[ZIP File (Zip Archive), 2 KB - mental_v10i1e46550_app2.zip]

Multimedia Appendix 3
Results of model 3 and 4 of the attention-deficit/hyperactivity disorder sample and control sample.
[PDF File (Adobe PDF File), 418 KB - mental_v10i1e46550_app3.pdf]

References

1. Fayyad J, Sampson NA, Hwang I, Adamowski T, Aguilar-Gaxiola S, Al-Hamzawi A, WHO World Mental Health Survey Collaborators. The descriptive epidemiology of DSM-IV Adult ADHD in the World Health Organization world mental health surveys. Atten Defic Hyperact Disord 2017 Mar;9(1):47-65 [FREE Full text] [doi: 10.1007/s12402-016-0208-3] [Medline: 27866355]
2. Simon V, Czobor P, Bálint S, Mészáros A, Bitter I. Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis. Br J Psychiatry 2009 Mar;194(3):204-211. [doi: 10.1192/bjp.bp.107.048827] [Medline: 19252145]
3. Daley D, Birchwood J. ADHD and academic performance: why does ADHD impact on academic performance and what can be done to support ADHD children in the classroom? Child Care Health Dev 2010 Jul;36(4):455-464. [doi: 10.1111/j.1365-2214.2009.01046.x] [Medline: 20074251]
4. Brunkhorst-Kanaan N, Libutzki B, Reif A, Larsson H, McNeill RV, Kittel-Schneider S. ADHD and accidents over the life span - a systematic review. Neurosci Biobehav Rev 2021 Jun;125:582-591 [FREE Full text] [doi: 10.1016/j.neubiorev.2021.02.002] [Medline: 33582234]
5. Altszuler AR, Page TF, Gnagy EM, Coxe S, Arrieta A, Molina BS, et al. Financial dependence of young adults with childhood ADHD. J Abnorm Child Psychol 2016 Aug;44(6):1217-1229 [FREE Full text] [doi: 10.1007/s10802-015-0093-9] [Medline: 26542688]
6. Kooij JJ, Bijlenga D, Salerno L, Jaeschke R, Bitter I, Balázs J, et al. Updated European consensus statement on diagnosis and treatment of adult ADHD. Eur Psychiatry 2019 Feb;56:14-34 [FREE Full text] [doi: 10.1016/j.eurpsy.2018.11.001] [Medline: 30453134]
7. Asherson P, Buitelaar J, Faraone SV, Rohde LA. Adult attention-deficit hyperactivity disorder: key conceptual issues. Lancet Psychiatry 2016 Jun;3(6):568-578. [doi: 10.1016/S2215-0366(16)30032-3] [Medline: 27183901]
8. Cortese S, Adamo N, Del Giovane C, Mohr-Jensen C, Hayes AJ, Carucci S, et al. Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. Lancet Psychiatry 2018 Sep;5(9):727-738 [FREE Full text] [doi: 10.1016/S2215-0366(18)30269-4] [Medline: 30097390]
9. Spencer T, Biederman J, Wilens T, Doyle R, Surman C, Prince J, et al. A large, double-blind, randomized clinical trial of methylphenidate in the treatment of adults with attention-deficit/hyperactivity disorder. Biol Psychiatry 2005 Mar 01;57(5):456-463. [doi: 10.1016/j.biopsych.2004.11.043] [Medline: 15737659]
10. van Lieshout M, Luman M, Twisk JW, van Ewijk H, Groenman AP, Thissen AJ, et al. A 6-year follow-up of a large European cohort of children with attention-deficit/hyperactivity disorder-combined subtype: outcomes in late adolescence and young adulthood. Eur Child Adolesc Psychiatry 2016 Sep;25(9):1007-1017 [FREE Full text] [doi: 10.1007/s00787-016-0820-y] [Medline: 26837866]
11. Mick E, McManus DD, Goldberg RJ. Meta-analysis of increased heart rate and blood pressure associated with CNS stimulant treatment of ADHD in adults. Eur Neuropsychopharmacol 2013 Jun;23(6):534-541 [FREE Full text] [doi: 10.1016/j.euroneuro.2012.06.011] [Medline: 22796229]

12. Adler LA, Orman C, Starr HL, Silber S, Palumbo J, Cooper K, et al. Long-term safety of OROS methylphenidate in adults with attention-deficit/hyperactivity disorder: an open-label, dose-titration, 1-year study. J Clin Psychopharmacol 2011 Feb;31(1):108-114. [doi: 10.1097/JCP.0b013e318203ea0a] [Medline: 21192153]
13. Stein MA, Weiss M, Hlavaty L. ADHD treatments, sleep, and sleep problems: complex associations. Neurotherapeutics 2012 Jul;9(3):509-517 [FREE Full text] [doi: 10.1007/s13311-012-0130-0] [Medline: 22718078]
14. Biederman J, Fried R, DiSalvo M, Woodworth KY, Biederman I, Driscoll H, et al. Further evidence of low adherence to stimulant treatment in adult ADHD: an electronic medical record study examining timely renewal of a stimulant prescription. Psychopharmacology (Berl) 2020 Sep;237(9):2835-2843. [doi: 10.1007/s00213-020-05576-y] [Medline: 32591937]
15. Breda V, Cerqueira RO, Ceolin G, Koning E, Fabe J, McDonald A, et al. Is there a place for dietetic interventions in adult ADHD? Prog Neuropsychopharmacol Biol Psychiatry 2022 Dec 20;119:110613. [doi: 10.1016/j.pnpbp.2022.110613] [Medline: 35964708]
16. Lange K. The need for alternative treatments for attention-deficit/hyperactivity disorder. J Dis Prev Health Promot 2020;4:1-9 [FREE Full text] [doi: 10.1093/pch/8.4.243]
17. Pinto S, Correia-de-Sá T, Sampaio-Maia B, Vasconcelos C, Moreira P, Ferreira-Gomes J. Eating patterns and dietary interventions in ADHD: a narrative review. Nutrients 2022 Oct 16;14(20):4332 [FREE Full text] [doi: 10.3390/nu14204332] [Medline: 36297016]
18. Adan RA, van der Beek EM, Buitelaar JK, Cryan JF, Hebebrand J, Higgs S, et al. Nutritional psychiatry: towards improving mental health by what you eat. Eur Neuropsychopharmacol 2019 Dec;29(12):1321-1332 [FREE Full text] [doi: 10.1016/j.euroneuro.2019.10.011] [Medline: 31735529]
19. Lassale C, Batty GD, Baghdadli A, Jacka F, Sánchez-Villegas A, Kivimäki M, et al. Healthy dietary indices and risk of depressive outcomes: a systematic review and meta-analysis of observational studies. Mol Psychiatry 2019 Jul;24(7):965-986 [FREE Full text] [doi: 10.1038/s41380-018-0237-8] [Medline: 30254236]
20. Jacka FN, O'Neil A, Opie R, Itsiopoulos C, Cotton S, Mohebbi M, et al. A randomised controlled trial of dietary improvement for adults with major depression (the 'SMILES' trial). BMC Med 2017 Jan 30;15(1):23 [FREE Full text] [doi: 10.1186/s12916-017-0791-y] [Medline: 28137247]
21. Del-Ponte B, Quinte GC, Cruz S, Grellert M, Santos IS. Dietary patterns and attention deficit/hyperactivity disorder (ADHD): a systematic review and meta-analysis. J Affect Disord 2019 Jun 01;252:160-173. [doi: 10.1016/j.jad.2019.04.061] [Medline: 30986731]
22. Shareghfarid E, Sangsefidi ZS, Salehi-Abargouei A, Hosseinzadeh M. Empirically derived dietary patterns and food groups intake in relation with Attention Deficit/Hyperactivity Disorder (ADHD): a systematic review and meta-analysis. Clin Nutr ESPEN 2020 Apr;36:28-35. [doi: 10.1016/j.clnesp.2019.10.013] [Medline: 32220366]
23. Farsad-Naeimi A, Asjodi F, Omidian M, Askari M, Nouri M, Pizarro AB, et al. Sugar consumption, sugar sweetened beverages and attention deficit hyperactivity disorder: a systematic review and meta-analysis. Complement Ther Med 2020 Sep;53:102512. [doi: 10.1016/j.ctim.2020.102512] [Medline: 33066852]
24. Koch ED, Moukhtarian TR, Skirrow C, Bozhilova N, Asherson P, Ebner-Priemer UW. Using e-diaries to investigate ADHD - state-of-the-art and the promising feature of just-in-time-adaptive interventions. Neurosci Biobehav Rev 2021 Aug;127:884-898. [doi: 10.1016/j.neubiorev.2021.06.002] [Medline: 34090919]
25. Smyth JM, Smyth JM. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003 Mar;4(1):35-52 [FREE Full text] [doi: 10.1023/A:1023657221954]
26. Evenden JL. Varieties of impulsivity. Psychopharmacology (Berl) 1999 Oct;146(4):348-361. [doi: 10.1007/pl00005481] [Medline: 10550486]
27. Kooij SJ, Bejerot S, Blackwell A, Caci H, Casas-Brugué M, Carpentier PJ, et al. European consensus statement on diagnosis and treatment of adult ADHD: the European network adult ADHD. BMC Psychiatry 2010 Sep 03;10:67 [FREE Full text] [doi: 10.1186/1471-244X-10-67] [Medline: 20815868]
28. Malloy-Diniz L, Fuentes D, Leite WB, Correa H, Bechara A. Impulsive behavior in adults with attention deficit/ hyperactivity disorder: characterization of attentional, motor and cognitive impulsiveness. J Int Neuropsychol Soc 2007 Jul;13(4):693-698. [doi: 10.1017/S1355617707070889] [Medline: 17521490]
29. Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull 1997 Jan;121(1):65-94. [doi: 10.1037/0033-2909.121.1.65] [Medline: 9000892]
30. Garza KB, Ding M, Owensby JK, Zizza CA. Impulsivity and fast-food consumption: a cross-sectional study among working adults. J Acad Nutr Diet 2016 Jan;116(1):61-68. [doi: 10.1016/j.jand.2015.05.003] [Medline: 26095434]
31. Lumley J, Stevenson R, Oaten M, Mahmut M, Yeomans M. Individual differences in impulsivity and their relationship to a western-style diet. Pers Individ Dif 2016 Jul;97:178-185 [FREE Full text] [doi: 10.1016/j.paid.2016.03.055]
32. Limbers CA, Young D. Executive functions and consumption of fruits/ vegetables and high saturated fat foods in young adults. J Health Psychol 2015 May;20(5):602-611. [doi: 10.1177/1359105315573470] [Medline: 25903247]
33. Steele CC, Pirkle JR, Kirkpatrick K. Diet-induced impulsivity: effects of a high-fat and a high-sugar diet on impulsive choice in rats. PLoS One 2017 Jun 29;12(6):e0180510 [FREE Full text] [doi: 10.1371/journal.pone.0180510] [Medline: 28662133]

34. Steele CC, Pirkle JR, Davis IR, Kirkpatrick K. Dietary effects on the determinants of food choice: impulsive choice, discrimination, incentive motivation, preference, and liking in male rats. Appetite 2019 May 01;136:160-172 [FREE Full text] [doi: 10.1016/j.appet.2019.01.023] [Medline: 30721744]
35. Jakobsen LH, Kondrup J, Zellner M, Tetens I, Roth E. Effect of a high protein meat diet on muscle and cognitive functions: a randomised controlled dietary intervention trial in healthy men. Clin Nutr 2011 Jun;30(3):303-311. [doi: 10.1016/j.clnu.2010.12.010] [Medline: 21239090]
36. Javelle F, Li D, Zimmer P, Johnson SL. Dietary intake of tryptophan tied emotion-related impulsivity in humans. Int J Vitam Nutr Res 2021 Jan;91(1-2):69-76. [doi: 10.1024/0300-9831/a000608] [Medline: 31584346]
37. Brandley ET, Holton KF. Breakfast positively impacts cognitive function in college students with and without ADHD. Am J Health Promot 2020 Jul;34(6):668-671. [doi: 10.1177/0890117120903235] [Medline: 32013526]
38. Verburgh L, Königs M, Scherder EJ, Oosterlaan J. Physical exercise and executive functions in preadolescent children, adolescents and young adults: a meta-analysis. Br J Sports Med 2014 Jun;48(12):973-979. [doi: 10.1136/bjsports-2012-091441] [Medline: 23467962]
39. Den Heijer AE, Groen Y, Tucha L, Fuermaier AB, Koerts J, Lange KW, et al. Sweat it out? The effects of physical exercise on cognition and behavior in children and adults with ADHD: a systematic literature review. J Neural Transm (Vienna) 2017 Feb;124(Suppl 1):3-26 [FREE Full text] [doi: 10.1007/s00702-016-1593-7] [Medline: 27400928]
40. Suarez-Manzano S, Ruiz-Ariza A, De La Torre-Cruz M, Martínez-López EJ. Acute and chronic effect of physical activity on cognition and behaviour in young people with ADHD: a systematic review of intervention studies. Res Dev Disabil 2018 Jun;77:12-23. [doi: 10.1016/j.ridd.2018.03.015] [Medline: 29625261]
41. Vysniauske R, Verburgh L, Oosterlaan J, Molendijk ML. The effects of physical exercise on functional outcomes in the treatment of ADHD: a meta-analysis. J Atten Disord 2020 Mar;24(5):644-654. [doi: 10.1177/1087054715627489] [Medline: 26861158]
42. Mehren A, Reichert M, Coghill D, Müller HH, Braun N, Philipsen A. Physical exercise in attention deficit hyperactivity disorder - evidence and implications for the treatment of borderline personality disorder. Borderline Personal Disord Emot Dysregul 2020 Jan 06;7:1 [FREE Full text] [doi: 10.1186/s40479-019-0115-2] [Medline: 31921425]
43. Abramovitch A, Goldzweig G, Schweiger A. Correlates of physical activity with intrusive thoughts, worry and impulsivity in adults with attention deficit/hyperactivity disorder: a cross-sectional pilot study. Isr J Psychiatry Relat Sci 2013;50(1):47-54 [FREE Full text] [Medline: 24029111]
44. Fuermaier AB, Tucha L, Koerts J, van den Bos M, Regterschot GR, Zeinstra EB, et al. Whole-body vibration improves cognitive functions of an adult with ADHD. Atten Defic Hyperact Disord 2014 Sep;6(3):211-220. [doi: 10.1007/s12402-014-0149-7] [Medline: 25031090]
45. Fuermaier AB, Tucha L, Koerts J, van Heuvelen MJ, van der Zee EA, Lange KW, et al. Good vibrations--effects of whole body vibration on attention in healthy individuals and individuals with ADHD. PLoS One 2014 Feb 28;9(2):e90747 [FREE Full text] [doi: 10.1371/journal.pone.0090747] [Medline: 24587412]
46. Gapin JI, Labban JD, Bohall SC, Wooten JS, Chang YK. Acute exercise is associated with specific executive functions in college students with ADHD: a preliminary study. J Sport Health Sci 2015 Mar;4(1):89-96. [doi: 10.1016/j.jshs.2014.11.003]
47. Mehren A, Özyurt J, Lam AP, Brandes M, Müller HH, Thiel CM, et al. Acute effects of aerobic exercise on executive function and attention in adult patients with ADHD. Front Psychiatry 2019 Mar 26;10:132 [FREE Full text] [doi: 10.3389/fpsyt.2019.00132] [Medline: 30971959]
48. Miguelez-Fernandez C, de Leon SJ, Baltasar-Tello I, Peñuelas-Calvo I, Barrigon ML, Capdevila AS, et al. Evaluating attention-deficit/hyperactivity disorder using ecological momentary assessment: a systematic review. Atten Defic Hyperact Disord 2018 Dec;10(4):247-265. [doi: 10.1007/s12402-018-0261-1] [Medline: 30132248]
49. Hinton PS, Johnstone B, Blaine E, Bodling A. Effects of current exercise and diet on late-life cognitive health of former college football players. Phys Sportsmed 2011 Sep;39(3):11-22. [doi: 10.3810/psm.2011.09.1916] [Medline: 22030936]
50. Molteni R, Wu A, Vaynman S, Ying Z, Barnard RJ, Gómez-Pinilla F. Exercise reverses the harmful effects of consumption of a high-fat diet on synaptic and behavioral plasticity associated to the action of brain-derived neurotrophic factor. Neuroscience 2004;123(2):429-440. [doi: 10.1016/j.neuroscience.2003.09.020] [Medline: 14698750]
51. Ruf A, Koch ED, Ebner-Priemer U, Knopf M, Reif A, Matura S. Studying microtemporal, within-person processes of diet, physical activity, and related factors using the APPetite-mobile-app: feasibility, usability, and validation study. J Med Internet Res 2021 Jul 05;23(7):e25850 [FREE Full text] [doi: 10.2196/25850] [Medline: 34342268]
52. Ruf A, Neubauer AB, Ebner-Priemer U, Reif A, Matura S. Studying dietary intake in daily life through multilevel two-part modelling: a novel analytical approach and its practical application. Int J Behav Nutr Phys Act 2021 Sep 27;18(1):130 [FREE Full text] [doi: 10.1186/s12966-021-01187-8] [Medline: 34579744]
53. Ruf A, Neubauer AB, Koch ED, Ebner-Priemer U, Reif A, Matura S. Individual differences in the dietary response to stress in ecological momentary assessment: does the individual-difference model need expansion? Appl Psychol Health Well Being 2023 May;15(2):629-649. [doi: 10.1111/aphw.12400] [Medline: 36082743]
54. Mayer JS, Hees K, Medda J, Grimm O, Asherson P, Bellina M, et al. Bright light therapy versus physical exercise to prevent co-morbid depression and obesity in adolescents and young adults with attention-deficit / hyperactivity disorder: study

protocol for a randomized controlled trial. Trials 2018 Feb 26;19(1):140 [FREE Full text] [doi: 10.1186/s13063-017-2426-1] [Medline: 29482662]
55. Ritter PS, Bermpohl F, Gruber O, Hautzinger M, Jansen A, Juckel G, et al. Aims and structure of the German research consortium BipoLife for the study of bipolar disorder. Int J Bipolar Disord 2016 Dec;4(1):26 [FREE Full text] [doi: 10.1186/s40345-016-0066-0] [Medline: 27873290]
56. Pfennig A, Leopold K, Martini J, Boehme A, Lambert M, Stamm T, et al. Improving early recognition and intervention in people at increased risk for the development of bipolar disorder: study protocol of a prospective-longitudinal, naturalistic cohort study (Early-BipoLife). Int J Bipolar Disord 2020 Jul 01;8(1):22 [FREE Full text] [doi: 10.1186/s40345-020-00183-4] [Medline: 32607662]
57. Arteaga-Henríquez G, Rosales-Ortiz SK, Arias-Vásquez A, Bitter I, Ginsberg Y, Ibañez-Jimenez P, et al. Treating impulsivity with probiotics in adults (PROBIA): study protocol of a multicenter, double-blind, randomized, placebo-controlled trial. Trials 2020 Feb 11;21(1):161 [FREE Full text] [doi: 10.1186/s13063-019-4040-x] [Medline: 32046750]
58. Chmitorz A, Neumann RJ, Kollmann B, Ahrens KF, Öhlschläger S, Goldbach N, et al. Longitudinal determination of resilience in humans to identify mechanisms of resilience to modern-life stressors: the longitudinal resilience assessment (LORA) study. Eur Arch Psychiatry Clin Neurosci 2021 Sep;271(6):1035-1051 [FREE Full text] [doi: 10.1007/s00406-020-01159-2] [Medline: 32683526]
59. Koch SA, Conrad J, Hierath L, Hancock N, Beer S, Cade JE, et al. Adaptation and evaluation of Myfood24-Germany: a web-based self-administered 24-h dietary recall for the German adult population. Nutrients 2020 Jan 06;12(1):160 [FREE Full text] [doi: 10.3390/nu12010160] [Medline: 31935885]
60. Tomko RL, Solhan MB, Carpenter RW, Brown WC, Jahng S, Wood PK, et al. Measuring impulsivity in daily life: the momentary impulsivity scale. Psychol Assess 2014 Jun;26(2):339-349 [FREE Full text] [doi: 10.1037/a0035083] [Medline: 24274047]
61. Whiteside SP, Lynam DR, Miller JD, Reynolds SK. Validation of the UPPS impulsive behaviour scale: a four ‐ factor model of impulsivity. Eur J Pers 2005 Dec 1;19(7):559-574. [doi: 10.1002/per.556]
62. Cyders MA, Smith GT, Spillane NS, Fischer S, Annus AM, Peterson C. Integration of impulsivity and positive mood to predict risky behavior: development and validation of a measure of positive urgency. Psychol Assess 2007 Mar;19(1):107-118. [doi: 10.1037/1040-3590.19.1.107] [Medline: 17371126]
63. Schmidt RE, Gay P, d'Acremont M, Van der Linden M. A German adaptation of the UPPS impulsive behavior scale: psychometric properties and factor structure. Swiss J Psychol 2008;67(2):107-112 [FREE Full text] [doi: 10.1024/1421-0185.67.2.107]
64. Wang LP, Maxwell SE. On disaggregating between-person and within-person effects with longitudinal data using multilevel models. Psychol Methods 2015 Mar;20(1):63-83. [doi: 10.1037/met0000030] [Medline: 25822206]
65. Viechtbauer W. Statistical methods for ESM data. In: Myin-Germeys I, Kuppens P, editors. The Open Handbook of Experience Sampling Methodology: A Step-By-Step Guide to Designing, Conducting, and Analyzing ESM Studies. 2nd edition. Chicago, IL: Independently published; 2022:153-183 URL: https://ppw.kuleuven.be/okp/esmhandbook.php
66. Thanarajah SE, Backes H, DiFeliceantonio AG, Albus K, Cremer AL, Hanssen R, et al. Food intake recruits orosensory and post-ingestive dopaminergic circuits to affect eating desire in humans. Cell Metab 2019 Mar 05;29(3):695-706.e4 [FREE Full text] [doi: 10.1016/j.cmet.2018.12.006] [Medline: 30595479]
67. Bürkner PC. brms: an R package for Bayesian multilevel models using stan. J Stat Softw 2017 Aug 29;80(1):1-28 [FREE Full text] [doi: 10.18637/jss.v080.i01]
68. Bürkner P. Advanced Bayesian multilevel modeling with the R package brms. R J 2018 May 18;10(1):395-411 [FREE Full text] [doi: 10.32614/RJ-2018-017]
69. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing URL: https://www.R-project.org/ [accessed 2023-08-02]
70. RStudio Team. RStudio: integrated development environment for R. RStudio. Boston, MA: RStudio URL: https://www.r-project.org/conferences/useR-2011/abstracts/180111-allairejj.pdf [accessed 2023-08-02]
71. Stan Development Team. RStan: the R interface to Stan. RStan. URL: https://cran.r-project.org/web/packages/rstan/vignettes/rstan.html [accessed 2023-08-02]
72. Beal DJ, Weiss HM. Methods of ecological momentary assessment in organizational research. Organ Res Methods 2003 Oct;6(4):440-464 [FREE Full text] [doi: 10.1177/1094428103257361]
73. Sobolev M, Vitale R, Wen H, Kizer J, Leeman R, Pollak JP, et al. The Digital Marshmallow Test (DMT) diagnostic and monitoring mobile health app for impulsive behavior: development and validation study. JMIR Mhealth Uhealth 2021 Jan 22;9(1):e25018 [FREE Full text] [doi: 10.2196/25018] [Medline: 33480854]
74. MacLean RR, Pincus AL, Smyth JM, Geier CF, Wilson SJ. Extending the balloon analogue risk task to assess naturalistic risk taking via a mobile platform. J Psychopathol Behav Assess 2018 Mar;40(1):107-116 [FREE Full text] [doi: 10.1007/s10862-017-9628-4] [Medline: 30505069]
75. Muth AK, Park SQ. The impact of dietary macronutrient intake on cognitive function and the brain. Clin Nutr 2021 Jun;40(6):3999-4010. [doi: 10.1016/j.clnu.2021.04.043] [Medline: 34139473]

76. Chang YK, Liu S, Yu HH, Lee YH. Effect of acute exercise on executive function in children with attention deficit hyperactivity disorder. Arch Clin Neuropsychol 2012 Mar;27(2):225-237. [doi: 10.1093/arclin/acr094] [Medline: 22306962]
77. Chuang LY, Tsai YJ, Chang YK, Huang CJ, Hung TM. Effects of acute aerobic exercise on response preparation in a Go/No Go task in children with ADHD: an ERP study. J Sport Health Sci 2015 Mar;4(1):82-88 [FREE Full text] [doi: 10.1016/j.jshs.2014.11.002]
78. Ludyga S, Brand S, Gerber M, Weber P, Brotzmann M, Habibifar F, et al. An event-related potential investigation of the acute effects of aerobic and coordinative exercise on inhibitory control in children with ADHD. Dev Cogn Neurosci 2017 Dec;28:21-28 [FREE Full text] [doi: 10.1016/j.dcn.2017.10.007] [Medline: 29100212]
79. Piepmeier AT, Shih CH, Whedon M, Williams LM, Davis ME, Henning DA, et al. The effect of acute exercise on cognitive performance in children with and without ADHD. J Sport Health Sci 2015 Mar;4(1):97-104 [FREE Full text] [doi: 10.1016/j.jshs.2014.11.004]
80. Pontifex MB, Saliba BJ, Raine LB, Picchietti DL, Hillman CH. Exercise improves behavioral, neurocognitive, and scholastic performance in children with attention-deficit/hyperactivity disorder. J Pediatr 2013 Mar;162(3):543-551 [FREE Full text] [doi: 10.1016/j.jpeds.2012.08.036] [Medline: 23084704]
81. Reichert M, Tost H, Reinhard I, Schlotz W, Zipf A, Salize HJ, et al. Exercise versus nonexercise activity: e-diaries unravel distinct effects on mood. Med Sci Sports Exerc 2017 Apr;49(4):763-773. [doi: 10.1249/MSS.0000000000001149] [Medline: 27824691]
82. Bell BM, Alam R, Alshurafa N, Thomaz E, Mondol AS, de la Haye K, et al. Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review. NPJ Digit Med 2020 Mar 13;3:38 [FREE Full text] [doi: 10.1038/s41746-020-0246-2] [Medline: 32195373]
83. Wen H, Sobolev M, Vitale R, Kizer J, Pollak JP, Muench F, et al. mPulse mobile sensing model for passive detection of impulsive behavior: exploratory prediction study. JMIR Ment Health 2021 Jan 27;8(1):e25019 [FREE Full text] [doi: 10.2196/25019] [Medline: 33502330]
84. Ravelli MN, Schoeller DA. Traditional self-reported dietary instruments are prone to inaccuracies and new approaches are needed. Front Nutr 2020 Jul 03;7:90 [FREE Full text] [doi: 10.3389/fnut.2020.00090] [Medline: 32719809]
85. Adamo KB, Prince SA, Tricco AC, Connor-Gorber S, Tremblay M. A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: a systematic review. Int J Pediatr Obes 2009;4(1):2-27. [doi: 10.1080/17477160802315010] [Medline: 18720173]
86. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
EMA: ecological momentary assessment
FFQ: food frequency questionnaire
MIS: Momentary Impulsivity Scale
OR: odds ratio
PA: physical activity
UPPS-P Impulsive Behavior Scale: Urgency-Premeditation-Perseverance-Sensation Seeking-Positive Urgency Impulsive Behavior Scale

Edited by J Torous; submitted 17.02.23; peer-reviewed by M Sobolev, AK Arend, H Yin; comments to author 14.05.23; revised version received 24.05.23; accepted 17.06.23; published 17.08.23.

Please cite as:
Ruf A, Neubauer AB, Koch ED, Ebner-Priemer U, Reif A, Matura S
Microtemporal Dynamics of Dietary Intake, Physical Activity, and Impulsivity in Adult Attention-Deficit/Hyperactivity Disorder: Ecological Momentary Assessment Study Within Nutritional Psychiatry
JMIR Ment Health 2023;10:e46550
URL: https://mental.jmir.org/2023/1/e46550
doi:10.2196/46550
PMID:37590053

Review

Effectiveness of Technology-Based Interventions for School-Age Children With Attention-Deficit/Hyperactivity Disorder: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Ka Po Wong[1], PhD; Jing Qin[2], PhD; Yao Jie Xie[2], PhD; Bohan Zhang[2], MSc

[1]Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[2]School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

Corresponding Author:
Ka Po Wong, PhD
Department of Applied Social Sciences
The Hong Kong Polytechnic University
11 Yuk Choi Road
Hong Kong
China (Hong Kong)
Phone: 852 64860214
Email: portia.wong@polyu.edu.hk

Abstract

Background: Attention-deficit/hyperactivity disorder (ADHD) is relatively common among school-age children. Technology-based interventions, such as computer-assisted training programs, neurofeedback training, and virtual reality, show promise in regulating the behaviors and cognitive functions of children with ADHD. An increasing number of randomized controlled trials have been conducted to evaluate the effectiveness of these technologies in improving the conditions of children with ADHD.

Objective: This study aims to conduct a systematic review of technological interventions for school-age children with ADHD and perform a meta-analysis of the outcomes of technology-based interventions.

Methods: A total of 19 randomized controlled studies involving 1843 participants were selected from a pool of 2404 articles across 7 electronic databases spanning from their inception to April 2022. ADHD behaviors, cognitive functions, learning ability, and quality of life were addressed in this study.

Results: Random effects meta-analyses found that children with ADHD receiving technology-based intervention showed small and significant effect sizes in computer-rated inattention (standardized mean difference [SMD] −0.35; P<.04), parent-rated overall executive function measured by the Behavior Rating Inventory of Executive Function (SMD −0.35; P<.04), parent-rated disruptive behavior disorder measured by the Child Behavior Checklist (SMD −0.50; P<.001) and Disruptive Behavior Disorder Rating Scale (SMD −0.31; P<.02), and computer-rated visual attention measured by the Continuous Performance Test (SMD −0.42; P<.001) and Reaction Time (SMD −0.43; P<.02).

Conclusions: Technology-based interventions are promising treatments for improving certain ADHD behaviors and cognitive functions among school-age children with ADHD.

Trial Registration: PROSPERO CRD42023446924; https://tinyurl.com/7ee5t24n

(JMIR Ment Health 2023;10:e51459) doi:10.2196/51459

KEYWORDS

attention-deficit/hyperactivity disorder; school-age children; computer-assisted training program; ADHD; neurofeedback training; virtual reality; cognitive functions

Introduction

Background

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorders among school-age children [1], with a prevalence of 7.6% [2]. Meanwhile, according to Diagnostic and Statistical Manual of Mental Disorders Text Revision Fourth Edition [3], the prevalence of ADHD in school-age children is 3% to 7% [3]. The 3 main categories of ADHD symptoms are inattention, hyperactivity, and impulsivity, which usually manifest in the school-age period [4]. These symptoms have detrimental impacts on the quality of life and functioning, including self-esteem, academic performance, social functioning, and relationship building [5]. ADHD is usually associated with long-term disability [6]. The types of treatments vary in different stages of life [7]. Behavioral parenting training and medication are the common approaches used to improve the behaviors and self-control of school-age children with ADHD [7]. ADHD medications are associated with an increased risk of headache, anxiety, and sleep disturbances [8,9]. Behavioral therapies are generally limited by time and space [10]. Therefore, feasible nonpharmacological approaches are recommended as alternatives to regulate the behaviors, executive functions, and well-being of children with ADHD.

In the last 10 years, human-computer interaction has widely been recognized in psychiatric and mental health research [11,12]. Digital technologies, neurofeedback systems, and virtual reality for health support, care, and treatment have increasingly been adopted, thus successfully gaining psychological health advantage [13,14]. ADHD is one of the common psychiatric disorders for which technology-based treatments are often used as therapeutic tools [15,16]. The application and effectiveness of technology-based interventions have been evaluated in ADHD treatment [4,17,18]. The advantages of technology-based treatment include improved executive functions and increased physiological and mental well-being [19,20]. However, inconsistent results have been reported regarding the efficiency of technology-based intervention in school-age children with ADHD. Regarding ADHD behaviors, Dovis et al [19], Egeland et al [21], Steiner et al [22], and van der Oord et al [23] found that computer-based training improved the inattention and hyperactivity in children with ADHD. However, some studies did not find significant results between technological treatment and ADHD behaviors [24-26]. Regarding executive functions, technology-based treatments improved inhibition [27], working memory [19,28], flexibility [19,27], emotional control [19], initiation [19], planning and organization [19], organizing materials [19], monitoring [19], and metacognition [22,23]. However, several studies determined that technology-based interventions have no effect on executive functions [21,24,25]. Regarding disruptive behavior disorder, Dovis et al [19], Lim et al [29], Steiner et al [22], and van der Oord et al [23] reported that computer-assisted training and neurofeedback training regulated oppositional defiant disorder and conduct disorder, whereas Bikic et al [24] and Breider et al [30] found no significant effect. Some studies have discovered that technology-based interventions can significantly improve visual attention [20,25,31], yet numerous studies indicated no significant effects [21,24,28,32,33]. These contradictory results make it difficult to examine the effectiveness of technology-based interventions in school-age children with ADHD.

Previous reviews illustrated the effectiveness of different types of technologies on children with ADHD [34-36]. Cibrian et al [34] and Powell et al [35] summarized this topic through a narrative description. Although Powell et al [35] adopted a meta-analysis to synthesize previous studies, their research focused on the use of virtual reality among children and adolescents with ADHD.

Objective

These existing studies based on nonrandomized, cross-sectional, and observational designs have added to the knowledge base and identified the potential implications of technology in enhancing the ability and functions of children. In addition, more randomized controlled trials (RCTs) have been conducted in the recent decade, as a growing number of researchers have shown interest in the use of technology-based interventions to improve the capability and well-being of children with ADHD. A meta-analysis of RCTs can provide strong and robust evidence regarding the effectiveness of technology-based interventions in improving children with ADHD. Therefore, a systematic review and meta-analysis of the existing evidence of RCTs are needed to explicate the advantages of technology to school-age children with ADHD.

Methods

Search Strategy

This review focused on RCTs using technologies to regulate the ability and alleviate the well-being of children with ADHD. This review was registered in the PROSPERO International Prospective Register of Systematic Reviews (registration number CRD42023446924). The studies included in this review were searched from electronic databases, including PubMed, MEDLINE, ScienceDirect, Web of Science, CINAHL (via EBSCO), PsycINFO (via OVID), and Scopus, in April 2022. The keywords used in the search engines were as follows:

- Population: “children with ADHD” or “school-age children with ADHD” or “students with ADHD”
- Intervention: “technology” or “computer” or “robots” or “virtual reality” or “VR” or “augmented reality” or “AR” or “web-based” or “serious games”
- Outcomes: “inattentive” or “hyperactive-impulsive” or “hyperactivity” or “impulsivity” or “executive functions” or “executive functioning” or “inhibition” or “working memory” or “emotional control” or “flexibility” or “attention” or “initiation” or “planning” or “organization” or “time management” or “metacognition” or “quality of life” or “performance”

Multimedia Appendix 1 shows the full search strategy.

Eligibility Criteria

The titles and abstracts of the selected papers were screened to identify relevant papers for this review. The analysis was

performed using Population, Intervention, Control, and Outcomes framework (1) population: patients who were diagnosed with ADHD and aged between 6 and 12 years; (2) intervention: using technology (ie, the method of applying scientific knowledge for practical purposes) without restriction of technology type or frequency of intervention; (3) comparison: technology-based interventions for managing ADHD versus no interventions, interventions with placebo effect, and treatment as usual or waitlist control; and (4) outcome: behaviors, cognitive functions, and whole well-being evaluation results for patients with ADHD. Studies were selected using the following inclusion criteria: (1) articles published in English and (2) RCTs. Exclusion criteria included (1) teenagers, adults, and older adults; (2) comorbidity with autism spectrum disorder, psychosis, and affective or anxiety disorder; (3) consumption of toxic substances; (4) diagnosed with learning disorder; (5) non–peer reviewed studies; and (6) qualitative studies, reviews, cross-sectional studies, case studies, observational studies, study protocols, pre-post studies without a control group, or conference abstracts without full text. Furthermore, the references of the included papers were manually checked for eligibility. After removing duplicate articles, the studies were independently screened by 2 reviewers (KPW and BZ). The selected full-text articles were retrieved and reviewed by 4 reviewers (KPW, JQ, YJX, and BZ).

Quality Assessment

The risk of bias in each study were independently evaluated by 3 reviewers (KPW, JQ, and BZ) using the Cochrane Collaboration tool for assessing the risk of bias [37]. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) criteria were adopted for conducting this review [38]. Multimedia Appendix 2 demonstrates the PRISMA checklist. The criteria of the tool included (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of outcome assessment, (5) incomplete outcome data, (6) selective reporting, and (7) other bias. In addition, judgment has 3 levels, including “low risk of bias,” “high risk of bias,” and “unclear risk of bias.” The conflicting results were settled by 4 reviewers (KPW, JQ, YJX, and BZ) through discussion.

Data Extraction

The information of the selected studies was extracted and coded into different categories, including study characteristics (first author’s name, publication year, country, and setting), characteristics of participants (sample size, sex, and age), intervention and control condition (type of technology used, frequency, length, and duration), outcome measurement (rating scale, test, and questionnaire), and result (mean and SD).

Data Synthesis

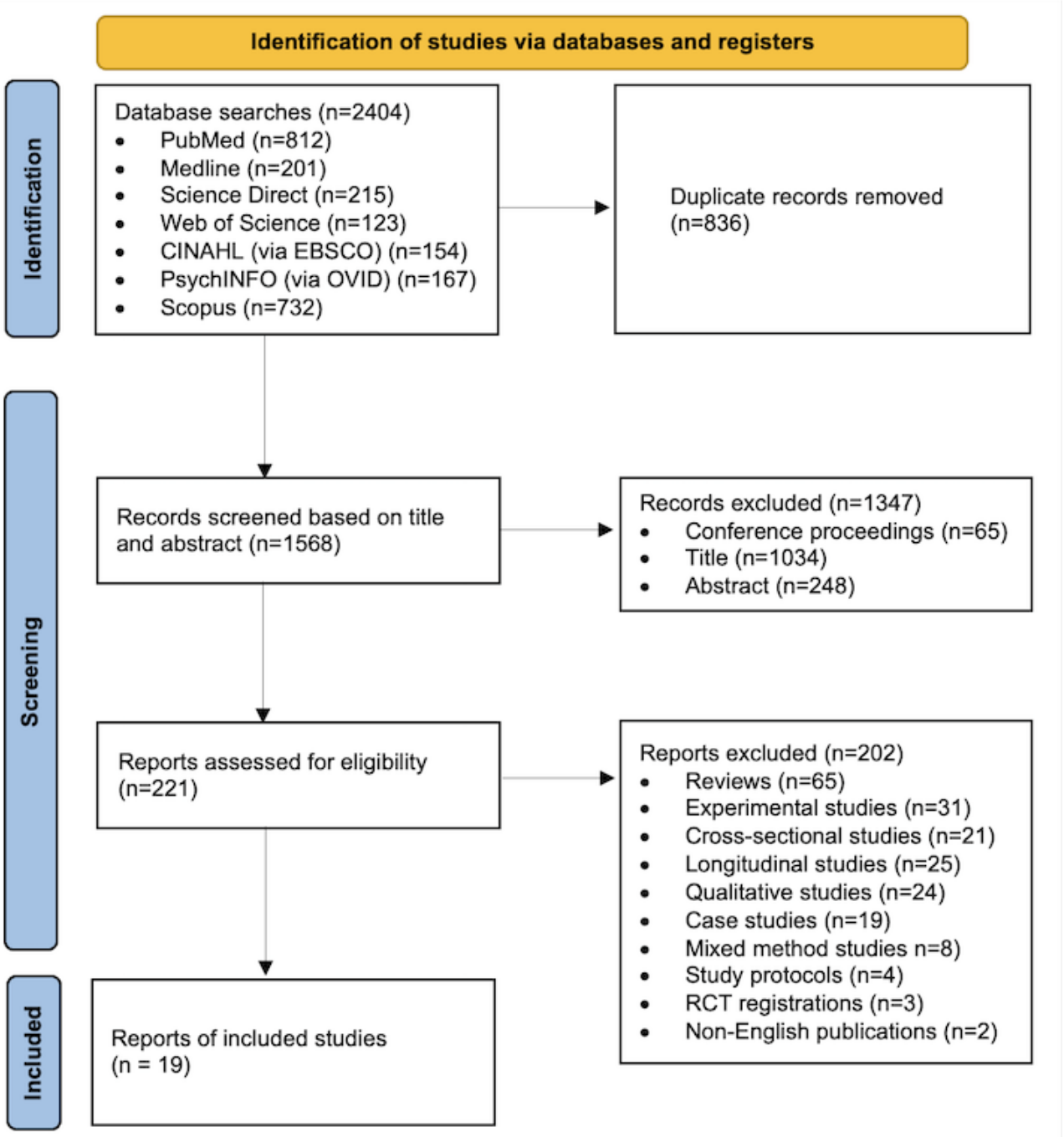
Data processing and analysis were conducted using the Review Manager Software RevMan (version 5.4), Cochrane Collaboration. The standardized mean difference (SMD) with a 95% CI was used to compute the effect size of the continuous outcomes of the interventions. The mean value of the baseline and posttest with SDs and the number of participants in the intervention groups and control groups were selected for the effect size calculation (ie, effect size of group differences). If reported, we selected the results estimated by the analysis of covariance, which treats individual baseline scores as covariates to correct for regression to baseline imbalanced means [39]. If the analysis of covariance had not been reported, change from baseline with SDs and posttest with SDs were selected. Multiple effect sizes were included in the same study, which contradicted the assumption that the effect sizes are independent of each other in the conventional meta-analytic procedures because the effect sizes in one study may be more correlated than those in other studies [40]. The results would become biased if this dependency was not considered. Hence, effect sizes assessed by the same measures were clustered to estimate the association between technology-based interventions and the conditions of school-age children with ADHD. The random effects model was applied in the meta-analysis, given the methodological diversity across the studies. Heterogeneity was measured using the I^2 value ($P<.10$; $I^2>50\%$), and the higher the value of I^2, the higher the level of heterogeneity. Subgroup analysis of parent-rated, teacher-rated, computer-rated, and self-rated results was performed, where applicable. To test moderating effects, 5 study-level characteristics, including the number of sessions, sample size, setting, game elements, and types of control group, were selected to calculate the meta-regression for each moderator.

Results

Selected Articles

A total of 2404 articles were retrieved from electronic databases. After removing duplicates, the titles and abstracts of 1568 articles were screened. In total, 1347 articles were excluded, and the remaining 221 articles were selected for full-text screening. At this stage, we excluded reviews (n=65), experimental studies (n=31), cross-sectional studies (n=21), longitudinal studies (n=25), qualitative studies (n=24), case studies (n=19), mixed methods studies (n=8), study protocols (n=4), RCT registrations (n=3), and non-English publications (n=2). Finally, 19 RCTs that met the inclusion and exclusion criteria were identified. A flowchart of the study selection process is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the study selection process. RCT: randomized controlled trial.



Risk of Bias

A low risk of random sequence generation was recorded from 8 trials (ie, the use of a computer random number generator and web-based system). Seven studies reported a low risk of allocation concealment. Meanwhile, 18 studies had a low risk of blinding participants and personnel. Although some studies had no complete blinding of participants and personnel [20-22,24,30,33,41,42], the reviewers determined that the outcomes were not likely to be affected by a lack of blinding. Six studies were judged to have a low risk of blinding the outcome assessment. Seven studies indicated no blinding of assessors, and 6 studies did not clearly indicate the blinding of assessors. All studies were judged to have a low risk of incomplete outcome data because the attrition rate of all studies was <20%. All studies were judged to have a low risk of selective reporting, as nearly all studies had the protocol and all studies reported the primary and secondary outcomes. Ten studies described the way to manage missing data (ie, intention-to-treat analysis) [19,22-25,27,29,30,41,43]. The risk of bias assessment is shown in Figures 2 and 3 [19,22-25, 27,29,30,41,43].

Figure 2. Risk of bias summary.

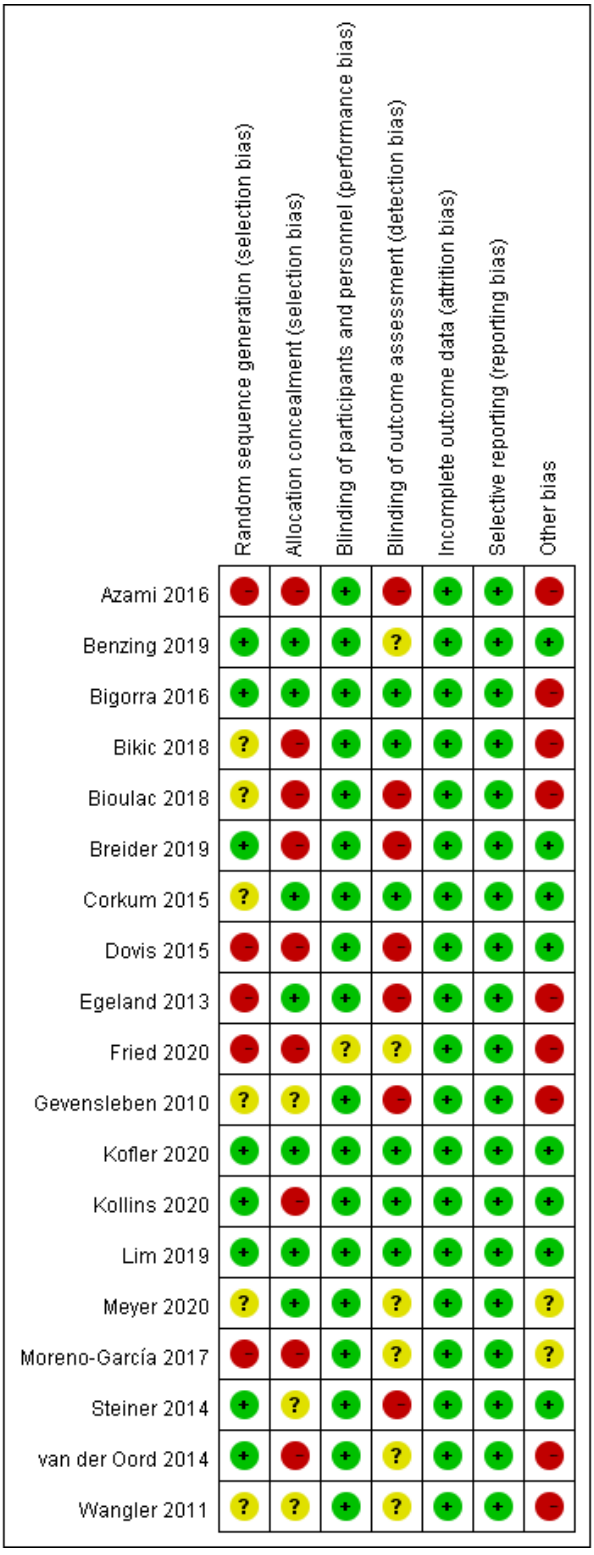
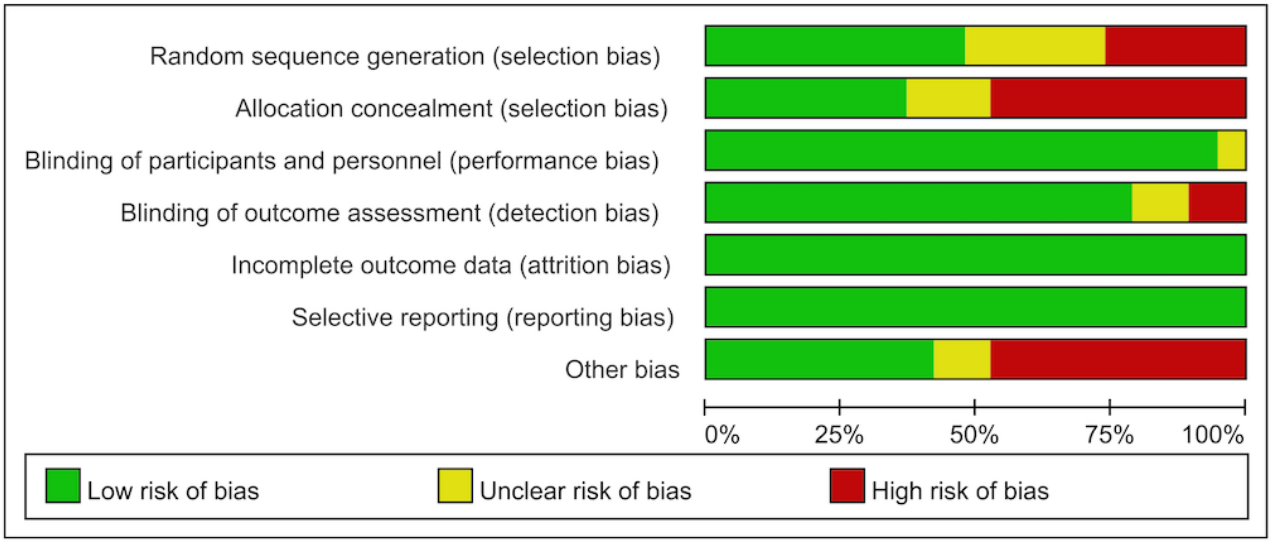


Figure 3. Risk of bias graph.



Study Characteristics

Fourteen studies used a 2-arm trials design [19-27,29-32,41-44], and the remaining 5 studies had a 3-arm trial design [19,20,22,28,33]. These studies were conducted in 11 countries: Canada [41], Denmark [24], France [33], Germany [31,42], Iran [28], the Netherlands [19,23,30], Norway [21], Singapore [29], Spain [20,32], Switzerland [27], and the United States [22,25,26,43,44].

Participants Characteristics

A total of 1843 school-age children were included in this review, all of whom were diagnosed with ADHD. The number of participants included in each trial ranged from 10 to 246. Most studies were conducted in participants' homes, and a small number of studies were conducted in schools, classrooms, and clinics. Detailed information is presented in Table 1.

Table 1. Participants' characteristics of the selected studies (n=19).

Study	Country	Setting	Sample size, n	Sex (female/male), n/n	Age (years), mean (range)[a]
Azami et al [28]	Iran	Participants' home	34	0/34	7-12
Benzing and Schmidt [27]	Switzerland	Participants' home	51	10/41	10.43 (8-12)
Bigorra et al [32]	Spain	Participants' home	66	36/24	8.9
Bikic et al [24]	Denmark	Participants' home	70	11/59	9.96
Bioulac et al [33]	France	No mention	51	10/41	8.9
Breider et al [30]	Netherlands	Participants' home	21	6/15	7.76
Corkum et al [41]	Canada	Classroom-based	58	7/51	8.83 (6-12)
Dovis et al [19]	Netherlands	Home-based training	89	18/71	10.50
Egeland et al [21]	Norway	School	67	18/49	10.4
Fried et al [44]	The United States	Participants' home	333	88/245	9.13
Gevensleben et al [42]	Germany	No mention	94	17/77	9.21
Kofler et al [25]	The United States	Office and participants' home	54	12/42	10.41
Kollins et al [43]	The United States	Participants' home	348	100/248	9.65
Lim et al [29]	Singapore	Clinic	172	25/147	8.6
Meyer et al [26]	The United States	Home	40	12/28	10.15
Moreno-García et al [20]	Spain	Home and school	57	13/44	8.84
Steiner et al [22]	The United States	Classroom	104	34/70	8.56
van der Oord et al [23]	Netherlands	Home	40	7/33	9.75
Wangler et al [31]	Germany	No mention	94	17/77	9.64

[a]If studies did not provide mean values and/or ranges for participant ages, this information was not shown in the table.

Technology-Based Intervention and Control Condition

Different types of technologies were adopted in the intervention group among the included studies, including computer-assisted training programs, neurofeedback training, and virtual reality. For the control groups, 6 studies used treatment or medication as the usual approach. The participants in the control groups of 3 included studies did not receive any training. Two studies used stimulants, 7 studies used placebo cognitive training, and 1 study used a digital game. Detailed information is presented in Table 2.

Table 2. Characteristics of intervention of the selected studies (n=19).

Study	Intervention	Control	Type of intervention	Outcome measurement
Azami et al [28]	Computer-assisted cognitive rehabilitation	Psychostimulants	Individual: 20 sessions in 3 mo, 90 min/session	Continuous Performance Test, Tower-of-London, Forward/Backward Digit Span From WISC-R[a], Raven Progressive Matrices, and web-based version of Span Board Task Progressive Matrices
Benzing and Schmidt [27]	Exergame training	Not receiving training	Individual: 8 wk, 3 times a week for at least 30 min	Conners - 3 Rating Scales and German Motor Test
Bigorra et al [32]	Computerized working memory training	Nonadaptive work memory training	Individual: 5 wk, 5 sessions per week, 30-45 min/session	Backward Digit Span, Letter-number Sequencing of WISC-IV[b], Backward Spatial Span of WMS-III[c], Lowa Gambling Task, Happé Strange Stories, and Folk Psychology Test
Bikic et al [24]	Cognitive computer games of the ACTIVATE program	Treatment as usual	Individual: 6 times a week, 8 wk	Cambridge Neuropsychological Test Automated Battery, Motor Screening Task, Attention Switching Task, Rapid Visual Information Processing, Intraextra Dimensional Set Shift, Reaction Time, ADHD[d] Rating Scale, and Behavior Rating Inventory of Executive Function
Bioulac et al [33]	Virtual classroom cognitive remediation program	Psychotherapy placebo training	Individual: twice a week for 6 wk, 12 sessions, 30 min/session	ADHD Rating Scale and Continuous Performance Test Task Assessment
Breider et al [30]	Web-based program and supportive therapist contact	Face-to-face parent training	Parents: 17 sessions, 45-60 min/session	Child Behavior Checklist
Corkum et al [41]	Web-based learning and blackboard learning	Treatment as usual	Teachers: 6 training sessions	Conners-3 Parent and Teacher Rating Scales and Impairment Rating Scale
Dovis et al [19]	Braingame Brian training	Placebo-mode working memory training	Individual: 25 training sessions in 5 wk, 35-50 min per session	Disruptive Behavior Disorder Rating Scale, Behavior Rating Inventory of Executive Function, Sensitivity to Punishment and Sensitivity to Reward Questionnaire for Children, Pediatric Quality of Life Inventory, and Home Situations Questionnaire
Egeland et al [21]	Cogmed RoboMemo program	Treatment as usual	Individual: daily basis for 5-7 wk, 30-45 min	Color Word Test, Trail Making Test, Conners Continuous Performance Test-II, Key Math, Logometrica, Benton Visual Retention Test, Children's Auditory Verbal Learning Test-2, ADHD Rating Scale, Strengths and Difficulties Questionnaire, and Behavior Rating Inventory of Executive Function
Fried et al [44]	Text messaging	Treatment as usual	Individual: 45 d	Adherence to stimulants
Gevensleben et al [42]	Neurofeedback training	Attention skills training	Individual: 6 mo, 25-30 min/session	ADHD Rating Scale, German Rating Scale for Oppositional Defiant/Conduct Disorders, Strength and Difficulties Questionnaire, Home Situations Questionnaire, and Homework Problem Checklist
Kofler et al [25]	Central executive training	Inhibitory control training	Group: 10 wk, 1 h/wk; individual: 10 wk, 2-3 d/wk, 15 min/d	Behavior Assessment System for Children, ADHD Rating Scale, and Phonological and Visuospatial Reordering

Study	Intervention	Control	Type of intervention	Outcome measurement
Kollins et al [43]	AKL-T01 (a digital therapeutic)	Digital game	Individual: 5 d/wk for 4 wk, 25 min/d	ADHD Rating Scale, Test of Variables of Attention, and Attention Performance Index
Lim et al [29]	Brain computer interface-based attention training program	Not receiving training	Individual: 3 sessions/wk in the first 8 wk and 4 sessions/wk in the next 12 wk	ADHD Rating Scale Inattention Score and Child Behavior Checklist
Meyer et al [26]	Computerized training	Medication as usual	Individual: at least 5 d a week for 4 wk, 15 min/d	Swanson, Nolan, and Pelham-IV Questionnaire and Conners Parent Rating Scale and Conners Teacher Rating Scale
Moreno-García et al [20]	Neurofeedback training	Pharmacology	Individual: 40 sessions	ADHD Rating Scale and Integrated Visual and Auditory Continuous Performance Test
Steiner et al [22]	Computer attention training using neurofeedback or cognitive training	Not receiving training	Individual: 40 sessions over 5 mo, 3 times/wk, 45 min/time	Conners 3–Parent Assessment Report, Behavior Rating Inventory of Executive Function, and Behavioral Observation of Students in Schools
van der Oord et al [23]	Executive functioning training	Treatment as usual	Individual: 25 sessions over 5 wk, 40 min/session	Behavior Rating Inventory of Executive Functioning and Disruptive Behavior Disorder Rating Scale
Wangler et al [31]	Neurofeedback training	Computerized attention skills training	Individual: 3-4 wk, 25-30 min/session	Attention Network Test and ADHD Rating Scale

[a]WISC-R: Wechsler Intelligence Scale for Children—Revised.

[b]WISC-IV: Backward Digit Span of the Wechsler Intelligence Scale for Children-IV.

[c]WMS-III: Wechsler Memory Scale–III.

[d]ADHD: attention-deficit/hyperactivity disorder.

Intervention Duration and Length

The intervention duration of 3 studies was 4 weeks. Meanwhile, the intervention duration of 9 studies was 5 to 8 weeks. One study had an intervention duration of 12 weeks, 2 studies had an intervention duration of 5 months, and 1 study had an intervention duration of 6 months. For the length of intervention, 1 trial was conducted for 15 minutes per session, 10 studies varied from 25 to 45 minutes per session, 2 studies varied from 45 to 60 minutes per session, and 1 study was conducted for 90 minutes per session.

Meta-Analysis Results of the Technology-Based Intervention

ADHD Behaviors

ADHD behaviors included in this study were inattention and hyperactivity or impulsivity. Moderate heterogeneity was observed among the studies examining ADHD behaviors (I^2=26.4%). Table 3 shows the pooled results of ADHD behavior. Corkum et al [41] evaluated the overall ADHD behaviors of participants, whereas the other 10 studies classified the results of the ADHD behaviors into inattention and hyperactivity or impulsivity. Corkum et al [41] reported the Conners 3 Parent Rating Scale (Conners 3-P) and Conners 3 Teacher Rating Scale (Conners 3-T) scores. The SMD of ADHD behaviors measured with parent-rated evaluation was −0.21 (95% CI −0.73 to 0.31), and the teacher-rated evaluation of ADHD behaviors was −0.46 (95% CI −0.98 to 0.06). However, no significant effect was observed for ADHD behavior.

A total of 12 studies reported inattention, which was measured using the ADHD Rating Scale (ADHD-RS), Disruptive Behavior Disorder Rating Scale (DBDRS), Conners 3-P, and Conners 3-T. Parent-rated inattention measured by ADHD-RS (*K*=7), Conners 3-P (*K*=4), and DBDRS (*K*=2) had no statistically significant effect. Teacher-rated inattention measured by ADHD-RS (*K*=6), Conners 3-T (*K*=1), and DBDRS (*K*=2) had no significant effect. Computer-rated inattention measured by ADHD-RS (*K*=2; SMD −0.35, 95% CI −0.68 to −0.01) had a small and statistically significant effect (*P*<.04).

Ten studies explored the effectiveness of technology-based interventions on hyperactivity or impulsivity. Three different scales were used: ADHD-RS, DBDRS, and Conners 3-P and Conners 3-T. No significant effect was found in parent-rated hyperactivity or impulsivity measured by ADHD-RS (*K*=7), Conners 3-P (*K*=4), and DBDRS (*K*=2) and teacher-rated hyperactivity or impulsivity measured by ADHD-RS (*K*=4), Conners 3-T (*K*=1), and DBDRS (*K*=2).

The results of moderator analysis demonstrated that the sample size, setting, game elements, and type of control group moderated the effect size (Table 4). A sample size of ≤50 (*K*=10; SMD −0.25, 95% CI −0.47 to −0.03; *P*<.03), nonhome setting (ie, clinic or school; *K*=12; SMD −0.24, 95% CI −0.37 to −0.11; *P*<.001), game elements excluded (*K*=40; SMD −0.15, 95% CI −0.23 to −0.07; *P*<.001), and control groups with no treatment (*K*=28; SMD −0.22, 95% CI −0.32 to −0.13; *P*<.001) and nonequivalent treatment had a significant moderating effect on the effect size.

Table 3. Meta-analyses results of technology-based interventions for school-age children with attention-deficit/hyperactivity disorder.

Variables	Number of effect sizes	Standardized mean difference	SE	95% CI	*z* score	*P* value
Parent-rated inattention						
ADHD-RS[a]	7	−0.190	0.099	−0.384 to 0.004	−1.92	.05
Conners 3-P[b]	4	−0.373	0.144	−0.655 to −0.092	−2.60	.10
DBDRS[c]	2	−0.462	0.212	−0.878 to −0.045	−2.17	.09
Teacher-rated inattention						
ADHD-RS	6	0.052	0.077	−0.098 to 0.202	0.678	.50
Conners 3-T[d]	1	−0.150	0.318	−0.774 to 0.474	−0.472	.64
DBDRS	2	−0.113	0.201	−0.507 to 0.280	−0.564	.57
Computer-rated inattention						
ADHD-RS	2	−0.345	0.17	−0.679 to −0.011	−2.025	.043
Parent-rated hyperactivity or impulsivity						
ADHD-RS	7	0.032	0.14	−0.243 to 0.306	0.225	.82
Conners 3-P	4	−0.16	0.184	−0.521 to 0.200	−0.871	.38
DBDRS	2	−0.311	0.267	−0.834 to 0.212	−1.164	.24
Teacher-rated hyperactivity or impulsivity						
ADHD-RS	4	0.097	0.133	−0.163 to 0.358	0.732	.46
Conners 3-T	1	0.312	0.318	−0.312 to 0.935	0.979	.33
DBDRS	2	−0.13	0.2	−0.521 to 0.261	−0.651	.52
Parent-rated executive functions						
BRIEF[e]	2	−0.347	0.17	−0.681 to −0.013	−2.038	.04
Parent-rated inhibition						
BRIEF	3	−0.192	0.192	−0.568 to 0.185	−0.998	.32
German Motor Test	1	−0.686	0.349	−1.370 to −0.002	−1.967	.05
Computer-rated inhibition						
BASC[f]	1	0.231	0.274	−0.306 to 0.768	0.844	.40
BRIEF	1	0.125	0.273	−0.410 to 0.661	0.458	.65
Parent-rated working memory						
BRIEF	3	−0.032	0.154	−0.333 to 0.269	−0.206	.84
German Motor Test	1	0.285	0.283	−0.269 to 0.840	1.009	.31
WISC-IV[g]	1	0.788	0.266	0.267 to 1.309	2.965	<.001
Computer-rated working memory						
BRIEF	2	0.261	0.449	−0.620 to 1.141	0.580	.56
WISC-IV	2	1.486	0.534	0.439 to 2.534	2.782	.01
Parent-rated flexibility						
BRIEF	3	−0.142	0.259	−0.649 to 0.365	−0.549	.58
German Motor Test	1	−0.566	0.457	−1.462 to 0.331	−1.237	.22
Parent-rated emotional control						
BRIEF	2	0.035	0.419	−0.786 to 0.856	0.084	.93
Parent-rated initiation						
BRIEF	2	0.108	0.244	−0.370 to 0.586	0.442	.66

Variables	Number of effect sizes	Standardized mean difference	SE	95% CI	*z* score	*P* value
Teacher-rated initiation						
BRIEF	2	0.196	0.169	−0.136 to 0.528	1.156	.25
Parent-rated planning and organization						
BRIEF	2	0.057	0.253	−0.438 to 0.553	0.227	.82
Parent-rated organizing materials						
BRIEF	2	0.118	0.177	−0.230 to 0.465	0.663	.51
Parent-rated monitoring						
BRIEF	2	0.291	0.327	−0.349 to 0.932	0.891	.37
Parent-rated metacognition						
BRIEF	5	−0.144	0.131	−0.400 to 0.113	−1.099	.27
Teacher-rated metacognition						
BRIEF	2	−0.116	0.171	−0.451 to 0.220	−0.676	.50
Parent-rated disruptive behavior disorder						
ADHD-RS	1	0.350	0.242	−0.124 to 0.825	1.447	.15
BOSS[h]	2	−0.187	0.170	−0.521 to 0.147	−1.096	.27
CBC[i]	3	−0.504	0.111	−0.720 to −0.287	−4.550	<.001
DBDRS	5	−0.306	0.125	−0.552 to −0.060	−2.441	.02
FBB-SSV[j]	3	−0.127	0.154	−0.428 to 0.174	−0.826	.41
Teacher-rated disruptive behavior disorder						
ADHD-RS	1	0.076	0.239	−0.392 to 0.545	0.319	.75
DBDRS	4	0.009	0.141	−0.267 to 0.286	0.065	.95
Computer-rated visual attention						
ANT[k]	5	0.014	0.112	−0.207 to 0.234	0.121	.90
AST[l]	2	−0.114	0.183	−0.473 to 0.244	−0.624	.53
BVRT[m]	1	0.000	0.263	−0.516 to 0.516	0.000	>.99
CPT[n]	8	−0.419	0.124	−0.662 to −0.176	−3.377	<.001
IED[o]	1	−0.313	0.260	−0.822 to 0.196	−1.205	.23
IVA/CPT[p]	6	−0.250	0.132	−0.508 to 0.009	−1.892	.06
RTI[q]	2	−0.425	0.186	−0.790 to −0.060	−2.280	.02
RVIP[r]	4	−0.021	0.130	−0.275 to 0.233	−0.165	.87
Computer-rated auditory attention						
CAVLT-2[s]	5	0.011	0.147	−0.276 to 0.298	0.074	.94
IVA/CPT	4	−0.084	0.195	−0.467 to 0.298	−0.431	.67
Parent-rated sensitivity to punishment and sensitivity to reward						
SPSRQ-C[t]	4	−0.147	0.128	−0.399 to 0.104	−1.148	.25
Parent-rated quality of life						
PedsQL[u]	1	0.646	0.263	0.131 to 1.161	2.458	.01
Self-rated quality of life						
PedsQL	1	0.042	0.256	−0.460 to 0.544	0.163	.87

Variables	Number of effect sizes	Standardized mean difference	SE	95% CI	*z* score	*P* value
Computer-rated reading fluency						
LOGOS[v]	2	0.014	0.186	−0.351 to 0.379	0.073	.94
Parent-rated adherence to stimulants						
Adherence to stimulants	1	−0.113	0.125	−0.358 to 0.132	−0.906	.37

[a]ADHD-RS: ADHD Rating Scale.
[b]Conners 3-P: Conners 3 Parent Rating Scale.
[c]DBDRS: Disruptive Behavior Disorder Rating Scale.
[d]Conners 3-T: Conners 3 Teacher Rating Scale.
[e]BRIEF: Behavior Rating Inventory of Executive Function.
[f]BASC: Behavior Assessment System for Children.
[g]WISC-IV: Backward Digit Span of the Wechsler Intelligence Scale for Children-IV.
[h]BOSS: Behavioral Observation of Students in Schools.
[i]CBC: Child Behavior Checklist.
[j]FBB-SSV: Fremdbeurteilungsbogen für Störungen des Sozialverhaltens.
[k]ANT: Attention Network Test.
[l]AST: Attention Switching Task.
[m]BVRT: Benton Visual Retention Test.
[n]CPT: Continuous Performance Test.
[o]IED: Intraextra Dimensional Set Shift.
[p]IVA/CPT: Integrated Visual and Auditory Continuous Performance Test.
[q]RTI: Reaction Time.
[r]RVIP: Rapid Visual Information Processing.
[s]CAVLT-2: Children's Auditory Verbal Learning Test-2.
[t]SPSRQ-C: Sensitivity to Punishment and Sensitivity to Reward Questionnaire for Children.
[u]PedsQL: Pediatric Quality of Life Inventory.
[v]LOGOS: Logometrica.

Table 4. Results of moderators between technology-based intervention for school-age children with attention-deficit/hyperactivity disorder (ADHD).

	Number of effect sizes	Standardized mean difference	SE	95% CI	*z* score	*P* value	Heterogeneity	
							Q-statistics	I^2
ADHD behavior	46	−0.105	0.037	−0.177 to −0.033	−2.853	<.001	61.123	26.378
Number of sessions	N/A[a]	N/A	N/A	N/A	−2.312	.02	N/A	N/A
≤20	11	−0.012	0.094	−0.195 to 0.172	−0.124	.90		
>20 to 40	26	−0.111	0.060	−0.229 to 0.007	−1.841	.07		
>40 to 60	4	−0.163	0.143	−0.443 to 0.117	−1.140	.25		
>60	2	−0.254	0.153	−0.554 to 0.045	−1.663	.10		
Sample size	N/A	N/A	N/A	N/A	−2.651	.01	N/A	N/A
≤50	10	−0.245	0.112	−0.465 to −0.025	−2.183	.03		
>50 to 100	29	−0.076	0.057	−0.188 to 0.037	−1.320	.19		
>100	7	−0.143	0.092	−0.324 to 0.037	−1.555	.12		
Setting	N/A	N/A	N/A	N/A	−2.651	.01	N/A	N/A
Home	21	−0.102	0.056	−0.212 to 0.007	−1.827	.07		
Mixed	8	0.204	0.107	−0.006 to 0.414	1.905	.06		
Nonhome	12	−0.242	0.066	−0.372 to −0.113	−3.670	<.001		
Game elements	N/A	N/A	N/A	N/A	−2.800	.01	N/A	N/A
Game elements excluded	40	−0.151	0.043	−0.234 to −0.068	−3.546	<.001		
Game elements included	1	0.241	0.131	−0.015 to 0.497	1.845	.07		
Types of control group	N/A	N/A	N/A	N/A	−2.756	.01	N/A	N/A
Equivalent treatment	9	0.083	0.071	−0.057 to 0.223	1.161	.25		
No treatment	28	−0.225	0.047	−0.317 to −0.132	−4.749	<.001		
Nonequivalent treatment	4	0.337	0.165	0.013 to 0.661	2.041	.04		
Executive functions	48	0.013	0.037	−0.060 to 0.087	0.357	.72	99.361	52.698
Number of sessions	N/A	N/A	N/A	N/A	0.257	.80	N/A	N/A
≤20	3	0.777	0.303	0.183 to 1.371	2.564	.01		
>20 to 40	43	−0.012	0.056	−0.121 to 0.097	−0.214	.83		
>40 to 60	2	0.012	0.253	−0.484 to 0.508	0.048	.96		
Sample size	N/A	N/A	N/A	N/A	0.260	.80	N/A	N/A
≤50	7	−0.046	0.168	−0.376 to 0.284	−0.272	.79		
>50 to 100	37	0.063	0.058	−0.051 to 0.177	1.082	.28		
>100	4	−0.363	0.174	−0.704 to −0.023	−2.093	.04		
Setting	N/A	N/A	N/A	N/A	0.257	.80	N/A	N/A
Home	40	0.047	0.059	−0.068 to 0.163	0.802	.42		
Mixed	2	0.178	0.268	−0.347 to 0.703	0.664	.51		
Nonhome	6	−0.239	0.146	−0.525 to 0.047	−1.641	.10		
Game elements	N/A	N/A	N/A	N/A	0.260	.80	N/A	N/A

	Number of effect sizes	Standardized mean difference	SE	95% CI	*z* score	*P* value	Heterogeneity	
							Q-statistics	I^2
Game elements excluded	42	0.009	0.055	−0.099 to 0.116	0.161	.87		
Game elements included	3	−0.318	0.219	−0.748 to 0.112	−1.448	.15		
Types of control group	N/A	N/A	N/A	N/A	0.257	.80	N/A	N/A
Equivalent treatment	11	−0.022	0.111	−0.240 to 0.196	−0.196	.84		
No treatment	34	−0.007	0.062	−0.129 to 0.115	−0.118	.91		
Nonequivalent treatment	3	0.777	0.303	0.183 to 1.370	2.565	.01		
Disruptive behavior disorder	19	−0.212	0.057	−0.323 to −0.101	−3.735	<.001	27.160	33.725
Number of sessions	N/A	N/A	N/A	N/A	−3.307	<.001	N/A	N/A
≤20	1	−1.070	0.475	−2.001 to −0.138	−2.251	.02		
>20 to 40	13	−0.102	0.078	−0.254 to 0.050	−1.312	.19		
>60	2	−0.470	0.128	−0.722 to −0.219	−3.668	<.001		
Sample size	N/A	N/A	N/A	N/A	−3.422	<.001	N/A	N/A
≤50	5	−0.395	0.156	−0.702 to −0.089	−2.527	.01		
>50 to 100	10	−0.035	0.084	−0.200 to 0.131	−0.411	.68		
>100	4	−0.377	0.102	−0.577 to −0.176	−3.675	<.001		
Setting	N/A	N/A	N/A	N/A	−2.714	.01	N/A	N/A
Home	12	−0.127	0.096	−0.314 to 0.061	−1.322	.19		
Nonhome	4	−0.361	0.130	−0.017 to −0.616	−2.781	.01		
Game elements	N/A	N/A	N/A	N/A	−2.738	.01	N/A	N/A
Game elements excluded	19	−0.196	0.072	−0.337 to −0.056	−2.738	.01		
Types of control group	N/A	N/A	N/A	N/A	−2.911	<.001	N/A	N/A
Equivalent treatment	8	−0.106	0.106	−0.314 to 0.102	−1.001	.32		
No treatment	10	−0.235	0.090	−0.412 to −0.058	−2.606	.01		
Nonequivalent treatment	1	−1.070	0.491	−2.032 to −0.108	−2.179	.03		
Visual attention	29	−0.174	0.050	−0.273 to −0.076	−3.487	<.001	36.296	22.856
Number of sessions	N/A	N/A	N/A	N/A	−3.415	<.001	N/A	N/A
≤20	5	−0.355	0.185	−0.718 to 0.009	−1.914	.06		
>20 to 40	18	−0.232	0.077	−0.383 to −0.080	−3.000	<.001		
>40 to 60	1	0.000	0.307	−0.602 to 0.602	0.000	>.99		
Sample size	N/A	N/A	N/A	N/A	−3.112	<.001	N/A	N/A
≤50	2	−0.025	0.315	−0.642 to 0.592	−0.080	.94		
>50 to 100	27	−0.187	0.059	−0.303 to −0.071	−3.152	<.001		
Setting	N/A	N/A	N/A	N/A	−2.810	.01	N/A	N/A
Home	12	−0.162	0.095	−0.349 to 0.025	−1.700	.10		

	Number of effect sizes	Standardized mean difference	SE	95% CI	*z* score	*P* value	Heterogeneity	
							Q-statistics	I^2
Mixed	8	−0.323	0.127	−0.571 to −0.075	−2.550	.01		
Nonhome	1	0.000	0.312	−0.612 to 0.612	0.000	>.99		
Game elements	N/A	N/A	N/A	N/A	−3.146	<.001	N/A	N/A
Game elements excluded	27	−0.187	0.059	−0.303 to −0.071	−3.152	<.001		
Game elements included	2	−0.025	0.315	−0.642 to 0.592	−0.080	.94		
Types of control group	N/A	N/A	N/A	N/A	−3.146	<.001	N/A	N/A
Equivalent treatment	13	−0.267	0.086	−0.435 to −0.099	−3.109	<.001		
No treatment	10	−0.147	0.089	−0.322 to 0.028	−1.647	.10		
Nonequivalent treatment	6	−0.003	0.155	−0.306 to 0.300	−0.017	.99	__[b]	—
Auditory attention	9	−0.020	0.091	−0.199 to 0.159	−0.218	.83	11.641	31.279
Number of sessions	N/A	N/A	N/A	N/A	−0.200	.84	N/A	N/A
>20 to 40	4	−0.084	0.195	−0.467 to 0.298	−0.431	.67		
>40 to 60	5	0.011	0.147	−0.276 to 0.298	0.074	.94		
Sample size	N/A	N/A	N/A	N/A	−0.205	.84	N/A	N/A
>50 to 100	9	−0.023	0.111	−0.240 to 0.195	−0.205	.84		
Setting	N/A	N/A	N/A	N/A	−0.200	.84	N/A	N/A
Mixed	4	−0.084	0.195	−0.467 to 0.298	−0.431	.67		
Nonhome	5	0.011	0.147	−0.276 to 0.298	0.074	.94		
Game elements	N/A	N/A	N/A	N/A	−0.205	.84	N/A	N/A
Game elements excluded	9	−0.023	0.111	−0.240 to 0.195	−0.205	.84		
Types of control group	N/A	N/A	N/A	N/A	−0.200	.84	N/A	N/A
No treatment	5	0.011	0.147	−0.276 to 0.298	0.074	.94		
Nonequivalent treatment	4	−0.084	0.195	−0.467 to 0.298	−0.431	.67		
Sensitivity to punishment and sensitivity to reward	4	−0.147	0.128	−0.399 to 0.104	−1.148	>.99	0.000	0.000
Quality of life	2	0.336	0.183	−0.023 to 0.696	1.833	.07	2.712	63.124
Number of sessions	N/A	N/A	N/A	N/A	1.129	.26	N/A	N/A
>20 to 40	2	0.341	0.302	−0.251 to 0.933	1.129	.26		
Sample size	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
>50 to 100	2	0.341	0.302	−0.251 to 0.933	1.129	.26		
Setting	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Home	2	0.341	0.302	−0.251 to 0.933	1.129	.26		
Game elements	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Game elements excluded	2	0.341	0.302	−0.251 to 0.933	1.129	.26		

	Number of effect sizes	Standardized mean difference	SE	95% CI	*z* score	*P* value	Heterogeneity	
							Q-statistics	I^2
Types of control group	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Equivalent treatment	2	0.341	0.302	−0.251 to 0.933	1.129	.26		
Reading fluency	2	0.014	0.173	−0.326 to 0.353	0.079	.94	1.156	13.528
Number of sessions	N/A	N/A	N/A	N/A	0.073	.94	N/A	N/A
>40 to 60	2	0.014	0.186	−0.351 to 0.379	0.073	.94		
Sample size	N/A	N/A	N/A	N/A	0.073	.94	N/A	N/A
>50 to 100	2	0.014	0.186	−0.351 to 0.379	0.073	.94		
Setting	N/A	N/A	N/A	N/A	0.073	.94	N/A	N/A
Nonhome	2	0.014	0.186	−0.351 to 0.379	0.073	.94		
Game elements	N/A	N/A	N/A	N/A	0.073	.94	N/A	N/A
Game elements excluded	2	0.014	0.186	−0.351 to 0.379	0.073	.94		
Types of control group	N/A	N/A	N/A	N/A	0.073	.94	N/A	N/A
No treatment	2	0.014	0.186	−0.351 to 0.379	0.073	.94		

[a]N/A: not applicable.

[b]Not available.

Executive Functions

The fundamental skills of executive functions include inhibition, working memory, flexibility, emotional control, initiation, planning and organization, organizing materials, monitoring, and metacognition. Substantial heterogeneity was observed among the studies that examined executive functions (I^2=52.7%). Table 3 shows the pooled results of the fundamental skills of executive functions. Steiner et al [22] reported an overall executive function of school-aged children with ADHD and SMD using the Behavior Rating Inventory of Executive Function (BRIEF) score. The SMD was −0.35 (95% CI −0.68 to −0.01), indicating a small and statistically significant effect was observed for the overall executive function.

Five studies reported inhibition through BRIEF, Behavior Assessment System for Children, and German Motor Test. Parent-rated inhibition using BRIEF (*K*=3) and German Motor Test (*K*=1) showed no significant effect. No significant effect was observed for computer-rated inhibition measured by BRIEF (*K*=1) and Behavior Assessment System for Children (*K*=1).

Six studies investigated working memory. Three different scales were used: the Backward Digit Span of the Wechsler Intelligence Scale for Children-IV (WISC-IV), BRIEF, and German Motor Test. No significant effect was observed for parent-rated working memory measured by BRIEF (*K*=3), German Motor Test (*K*=1), and WISC-IV (*K*=1). A large and statistically significant effect favoring the control group was observed for computer-rated working memory measured by WISC-IV (*K*=2; SMD 1.49, 95% CI 0.44-2.53).

Four studies evaluated flexibility using BRIEF and German Motor Test. No significant result was observed for parent-rated flexibility measured by BRIEF (*K*=3) and German Motor Test (*K*=1).

Two studies reported emotional control, and the BRIEF score was used for the measurement. No significant effect was observed for parent-rated emotional control and teacher-rated emotional control.

Initiation was reported by 2 studies using BRIEF to score. No statistically significant result was found for parent-rated initiation and teacher-rated initiation.

Three studies used the BRIEF assessment to score and report the impact of technology-based interventions on planning and organization. No significant effect was found for planning and organization for parent-rated planning and organization.

Two studies reported the BRIEF scores for organizing materials. No significant result was observed for parent-rated organizing materials.

Two studies reported the BRIEF scores for monitoring. The results of parent-rated monitoring were not statistically significant.

Four studies explored the effects of technology-based interventions on metacognition using BRIEF. The results for parent-rated and teacher-rated metacognition were not statistically significant.

The results of moderator analysis showed no moderating effects among the moderators (Table 4).

Disruptive Behavior Disorder

Conduct disorder and oppositional defiant disorder are the most common disruptive behavior disorders. Seven studies reported the effectiveness of technology-based interventions on disruptive behavior disorder. Moderate heterogeneity was observed among these studies (I^2=33.7%). Table 3 shows the pooled results of disruptive behavior disorder. Five different scales were adopted: ADHD-RS, Behavioral Observation of Students in Schools, Child Behavior Checklist, DBDRS, and German Rating Scale for Oppositional Defiant and Conduct Disorders. Small and significant effects were observed for parent-rated disruptive behavior disorder measured by Child Behavior Checklist (*K*=3; SMD −0.50, 95% CI −0.72 to −0.29) and DBDRS (*K*=5; SMD −0.31, 95% CI −0.55 to −0.06). Parent-rated disruptive behavior disorder measured by ADHD-RS (*K*=1), Behavioral Observation of Students in Schools (*K*=2), and German Rating Scale for Oppositional Defiant and Conduct Disorders (*K*=3) had no significant effect. No significant effect was observed for teacher-rated disruptive behavior disorder measured using ADHD-RS and DBDRS.

The results of the moderator analysis demonstrated that the number of sessions, sample size, setting, game elements, and type of control group moderated the effect size (Table 4). The number of sessions of >60 (*K*=2; SMD −0.47, 95% CI −0.72 to −0.22; *P*<.001), sample size of ≤50 (*K*=5; SMD −0.40, 95% CI −0.70 to −0.09; *P*<.01), and >100 (*K*=4; SMD −0.38, 95% CI −0.58 to −0.18; *P*<.001); nonhome setting (*K*=4; SMD −0.36, 95% CI −0.02 to −0.62; *P*<.005); game elements excluded (*K*=19; SMD −0.20, 95% CI −0.34 to −0.06; *P*<.006); and no treatment (*K*=10; SMD −0.24, 95% CI −0.41 to −0.06; *P*<.009) moderated the effect size.

Visual Attention

Eight studies explored the effects of technology-based visual attention. There was a small heterogeneity among these studies (I^2=22.9%). Table 3 shows the pooled results of visual attention. Four studies used the Continuous Performance Test as measurement, whereas the other study used the Benton Visual Retention Test, Phonological and Visuospatial Reordering, and Integrated Visual and Auditory Continuous Performance Test (IVA/CPT). In addition, 1 study used the Attention Network Test, Attention Switching Task, Rapid Visual Information Processing, Intraextra Dimensional Set Shift, and Reaction Time to score visual attention. Technology-based intervention had small and significant effects on visual attention measured by Continuous Performance Test (*K*=8; SMD −0.42, 95% CI −0.66 to −0.18) and Reaction Time (*K*=2; SMD −0.43, 95% CI −0.79 to −0.06). No significant effect was observed for visual attention measured using Attention Network Test, Attention Network Test, Benton Visual Retention Test, Intraextra Dimensional Set Shift, IVA/CPT, and Rapid Visual Information Processing.

The results of the moderator analysis showed that the number of sessions, sample size, setting, game elements, and type of control group moderated the effect size (Table 4). The number of sessions of >20 to 40 (*K*=18; SMD −0.23, 95% CI −0.38 to −0.08; *P*<.003), sample size of >than 50 to 100 (*K*=27; SMD −0.19, 95% CI −0.30 to −0.07; *P*<.002), mixed setting (ie, combining both home and school or clinic; *K*=8; SMD −0.32, 95% CI −0.57 to −0.08; *P*<.01), game elements excluded (*K*=27; SMD −0.19, 95% CI −0.30 to −0.07; *P*<.002), and equivalent treatment (*K*=13; SMD −0.27, 95% CI −0.44 to −0.10; *P*<.002) moderated the effect size.

Auditory Attention

Two studies evaluated auditory attention, which was scored using 2 scales, namely, Children's Auditory Verbal Learning Test-2 (CAVLT-2) and IVA/CPT. No statistically significant effect was observed for auditory attention measured using these 2 scales (Table 3). Moderate heterogeneity was observed among these studies (I^2=31.3%).

The results of moderator analysis showed no moderating effects among the moderators (Table 4).

Sensitivity to Punishment and Sensitivity to Reward

The study by Dovis et al [19] reported the Sensitivity to Punishment and Sensitivity to Reward Questionnaire for Children score. A nonsignificant effect was found for parent-rated sensitivity to punishment and sensitivity to reward (Table 3). Homogeneity was observed for these records (Table 4).

Quality of Life

Dovis et al [19] reported the Pediatric Quality of Life Inventory (PedsQL) score. No significant effect was observed for parent-rated quality of life and self-rated quality of life (Table 3). Substantial heterogeneity was observed among these records (I^2=63.1%).

The results of moderator analysis showed no moderating effects among the moderators (Table 4).

Reading Fluency

Egeland et al [21] investigated the effectiveness of a technology-based intervention on reading fluency. Logometrica was used to score reading fluency. The result of reading fluency was not at a statistically significant level (Table 3). Small heterogeneity was observed among these records (I^2=13.5%).

The results of moderator analysis showed no moderating effects among the moderators (Table 4).

Adherence to Stimulants

Fried et al [44] evaluated adherence to stimulants. There was no significant effect on adherence to stimulants (Table 3). Heterogeneity estimation was not conducted because of 1 effect size in adherence to stimulants. No moderator analysis was conducted because of only 1 effect size for adherence to stimulants (Table 4).

Discussion

Principal Findings

The adoption of technologies in school-age children with ADHD has become prevalent in this decade and has been applied in many countries. Nevertheless, the efficacy of applying technology-based treatments for children has not been well established. To the best of our knowledge, this work is the first

systematic review and meta-analysis of RCTs to assess the efficacy of technology-based interventions for school-age children, which provides strong evidence for health care practitioners. The pooled results indicated that technology-based treatment has the potential to regulate the inattention, overall executive function, and visual attention of school-age children with ADHD.

Quality of Evidence

Most of the included studies were of moderate quality, and only 2 studies [25,29] fulfilled the 7 criteria for randomized trials. More than half of the included studies had an unblinded or unclear method of randomization and allocation concealment. Although the blinding of participants was unpreventably broken owing to the nature of the intervention in most studies, the level of influence on the outcomes was considered low in most studies. Only 6 studies were assessed by blinded assessors. The attrition rate in all interventions did not exceed 20%, which met the criteria of the dropout rate. Only 8 studies reported the intention-to-treat analysis. Thus, more rigorous methods should be designed to scientifically justify the effects of technology-based interventions for children with ADHD.

Effects of Technology-Based Intervention on ADHD Behaviors

The pooled results indicated that technology-based treatment may have the potential to improve computer-rated inattention. No significant results were found for parent-rated and teacher-rated inattention and parent-rated and teacher-rated hyperactivity or impulsivity. This finding is consistent with that of previous studies, which showed that computer-assisted training had a positive effect on inattention [45,46]. Moreover, neurofeedback training and brain-computer interfaces training involve feeding brain activity into a computer, which maps the areas of the brain [47]. The patient controls a computer or video game by producing a brief, sustained brainwave activity in the target area, thereby training the brain regions that are aroused. These treatments include cognitive training to remediate the inattention behaviors of patients with ADHD. The insignificant finding on hyperactivity or impulsivity may be explained by the harmful effect of screen time and the frequent digital media use, which seem to increase the risk of the symptoms of hyperactivity or impulsivity. A meta-analysis conducted by Nikkelen et al [17] identified a modest association between the use of traditional digital media and ADHD behaviors. Notifications and invitation messages may pop up in these training programs. Exposure to these notifications and messages may take away the attention of a child from the task, possibly disrupting the normative development of patience and impulse control [48]. The precise descriptions of computer training, neurofeedback training, and virtual reality have not been completely demonstrated. Thus, relevant information is needed to provide further studies to assess the impact of the messages that suddenly pop up regarding the symptoms of hyperactivity and impulsivity among children with ADHD. Furthermore, it was found that a smaller sample size, nonhome setting, game elements excluded, and no treatment and nonequivalent treatment as the control group tended to have more improving effects on ADHD behavior while conducting the technology-based intervention. This suggests that fewer participants, conducting the intervention at clinic and school, intervention without game elements, and using no treatment or nontechnological intervention as control groups may increase the potential of the technology-based intervention to alleviate ADHD behaviors.

Effects of Technology-Based Intervention on Executive Functions

Nine categories, namely, inhibition, working memory, flexibility, emotional control, initiation, planning and organization, organizing materials, monitoring, and metacognition, were used to evaluate the executive functions. The pooled result showed a statistically significant effect of the technology-based intervention on parent-rated executive functions. A large significant effect of the control group on computer-rated working memory compared with the technological intervention was found, indicating computed-rated working memory would favor the control group instead of technological intervention. No significant effects were found from other categories related to executive functions. These findings were unexpected, as most previous studies reported positive effects of technology-based training on the improvement of executive functions [18,49,50].

Effects of Technology-Based Intervention on Disruptive Behavior Disorder

Evidence indicated that technology-based interventions and computer-assisted training may regulate both parent-rated disruptive behavior disorders. This finding is consistent with the previous case study by Kotwal et al [51], who found that the frequency of parent-reported disruptive behaviors in a boy with ADHD was reduced at home and school after conducting a 3-month computer-assisted training. More number of sessions, small and large sample sizes, nonhome settings, game elements excluded, and no treatment as the control group seemed to improve the disruptive behavior disorder when adopting the technology-based intervention. It is unexpected that small and large sample sizes might moderate the association between technology-based interventions and disruptive behavior disorder. Thus, further investigations need to be conducted to evaluate the moderating effect of different sample sizes.

Effects of Technology-Based Intervention on Visual Attention

Participants' ability to take in important visual information seems to be significantly improved after conducting computer-assisted training. Ordikhani-Seyedlar et al [52] used brain-computer interfaces with electroencephalograms to generate neurofeedback that significantly improved the visual attention of patients with ADHD. In addition, numerous studies have reported the positive outcomes of adopting neurofeedback training in patients with ADHD [53,54]. Alpha-beta activities in the electroencephalogram have been used to evaluate the attention level of the participants [52]. Furthermore, moderate sessions, moderate sample size, mixed setting, and game elements excluded and using equivalent treatment as a control might improve visual attention under the technology-based interventions. It is somewhat surprising that a mixed setting and

equivalent treatment as the control group may have a moderating effect; thus, more RCTs should be conducted to identify the underlying reasons.

Effects of Technology-Based Intervention on Auditory Attention

The pooled result showed a nonsignificant result on auditory attention. However, some experimental studies have reported the significant effect of biofeedback technology on strengthening auditory attention in children with ADHD [53]. Auditory attention among children with ADHD may be influenced by the intensity and frequency of sound [54,55]. White noise therapy has the potential to improve speech recognition and auditory attention in children with ADHD [56,57]. White noise, which is a random signal with equal intensity at different frequencies, is a steady, unchanging, and unobtrusive sound from certain machines such as a whirring fan and a static radio. The intensity and frequency of sound can remarkably influence the auditory attention of children with ADHD.

Effects of Technology-Based Intervention on Reading Fluency

No statistically significant effect was found for reading fluency. Nevertheless, previous experiments reported an improvement in oral reading fluency through a computerized program, Headsprout [58]. The computerized program could enhance the engagement of students with ADHD. Hence, learning among these students becomes more effective. This observation also implies that technology, such as a computerized program, has a high potential to enhance oral reading fluency. More RCTs with a larger pediatric population are needed to evaluate the effects of technology on reading fluency among children with ADHD.

Effects of Technology-Based Intervention on Adherence to Stimulants

Our result regarding the effects of the technology-based intervention on adherence to stimulants was not statistically significant. The literature review conducted by Chacko et al [59] indicated that technology-based interventions could reduce environmental barriers, such as caregivers' forgetfulness, to improve medication adherence for the pediatric population with ADHD. This nonsignificant finding may be explained by the fact that only a small number of RCTs were conducted to determine the effects of using technologies on stimulant adherence in children with ADHD.

Effects of Technology-Based Intervention on Sensitivity to Punishment, Sensitivity to Reward, and Quality of Life

A narrative review and pooled results found that technology had no significant effects on sensitivity to punishment and sensitivity to reward and quality of life. Few qualitative and quantitative studies have been conducted to investigate the effectiveness of technology-based interventions on sensitivity to punishment and sensitivity to reward and quality of life among children with ADHD. This circumstance indicates that more trials should be conducted.

Limitations and Future Research

This study is the first systematic review and meta-analysis of RCTs to examine the effectiveness of technology-based treatment for school-age children with ADHD. The strengths of this review include the use of well-defined inclusive and exclusive criteria, the application of a rigorous and well-constructed search strategy from the 7 electronic databases, and the stringent quality assessment of selected RCTs. However, this study has some limitations. First, the small sample size of the included trials implies the sparsity of RCTs in the children-computer interaction research field. Second, the pooled results may be influenced by variations in intervention designs (ie, intervention intervals and duration) because of the considerable clinical heterogeneity. Third, more than half of the included RCTs were designed using a nonblind approach owing to the unavoidable physical component of the interventions. To minimize these biases, blinded outcome assessors are strongly recommended for open-label RCTs. Fourth, the age selection was between 6 and 12 years, which may exclude preschool age children and adolescents with ADHD benefiting from technology-based treatment. Finally, RCTs published in languages other than English were excluded. Thus, language bias may be premeditated. Although RCT provide excellent internal validity and the most reliable evidence, the results are needed to be interpreted with caution owing to the inclusion of the designated population in this review. Furthermore, the use of virtual reality–based treatments for children with ADHD has been gradually implemented in the field of developmental psychology. RCTs assessing the effects of virtual reality–based treatments on children are still limited; thus, more relevant research is recommended. Moreover, the safety issues of children's interactions with technological interventions have rarely been discussed in RCTs. Measurements of side effects can be adopted in studies to further optimize the designs of the technological treatment, which can facilitate the universality of technological therapies in patients of different ages.

Conclusions

This review synthesized evidence from 19 RCTs on the application of technology-based interventions in school-age children with ADHD. The results indicated that the existing RCTs were mainly of low to moderate quality, particularly in random sequence generation, allocation of concealment and blinding, and intention-to-treat analysis. Further well-designed and rigorous trials should be conducted to determine the effectiveness of technology-based interventions in children with ADHD. The pooled results indicated that children's intervention with technological treatments has effects on computer-rated inattention, parent-rated overall executive functions, parent-rated disruptive behavior disorder, and visual attention. The number of sample sizes, setting of the intervention, exclusion of game elements, and type of control group had moderating effects on ADHD behaviors, disruptive behavior disorder, and visual attention. The potential of technology-based therapies for hyperactivity or impulsivity, inhibition, working memory, flexibility, emotional control, initiation, planning and organization, organizing materials, monitoring, metacognition, disruptive behavior disorder, visual attention, auditory attention,

sensitivity to punishment and reward, quality of life, adherence to stimulants, and reading fluency requires further evaluation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategy.
[DOCX File , 15 KB - mental_v10i1e51459_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[PDF File (Adobe PDF File), 81 KB - mental_v10i1e51459_app2.pdf]

References

1. Willcutt EG. The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics 2012 Jul;9(3):490-499 [FREE Full text] [doi: 10.1007/s13311-012-0135-8] [Medline: 22976615]
2. Salari N, Ghasemi H, Abdoli N, Rahmani A, Shiri MH, Hashemian AH, et al. The global prevalence of ADHD in children and adolescents: a systematic review and meta-analysis. Ital J Pediatr 2023 Apr 20;49(1):48 [FREE Full text] [doi: 10.1186/s13052-023-01456-1] [Medline: 37081447]
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th edition. New York City, NY: American Psychiatric Association; 2000.
4. Colomer C, Berenguer C, Roselló B, Baixauli I, Miranda A. The impact of inattention, hyperactivity/impulsivity symptoms, and executive functions on learning behaviors of children with ADHD. Front Psychol 2017 Apr 12;8:540 [FREE Full text] [doi: 10.3389/fpsyg.2017.00540] [Medline: 28446885]
5. Van der Oord S, Van der Meulen EM, Prins PJ, Oosterlaan J, Buitelaar JK, Emmelkamp PM. A psychometric evaluation of the social skills rating system in children with attention deficit hyperactivity disorder. Behav Res Ther 2005 Jun;43(6):733-746. [doi: 10.1016/j.brat.2004.06.004] [Medline: 15890166]
6. Biederman J, Faraone S, Milberger S, Curtis S, Chen L, Marrs A, et al. Predictors of persistence and remission of ADHD into adolescence: results from a four-year prospective follow-up study. J Am Acad Child Adolesc Psychiatry 1996 Mar;35(3):343-351. [doi: 10.1097/00004583-199603000-00016] [Medline: 8714323]
7. Barbaresi WJ, Colligan RC, Weaver AL, Voigt RG, Killian JM, Katusic SK. Mortality, ADHD, and psychosocial adversity in adults with childhood ADHD: a prospective study. Pediatrics 2013 Apr;131(4):637-644 [FREE Full text] [doi: 10.1542/peds.2012-2354] [Medline: 23460687]
8. Briars L, Todd T. A review of pharmacological management of attention-deficit/hyperactivity disorder. J Pediatr Pharmacol Ther 2016 May;21(3):192-206 [FREE Full text] [doi: 10.5863/1551-6776-21.3.192] [Medline: 27453697]
9. Lim CG, Ooi YP, Fung DS, Mahendran R, Kaur A. Sleep disturbances in Singaporean children with attention deficit hyperactivity disorder. Ann Acad Med Singap 2008 Aug;37(8):655-661 [FREE Full text] [Medline: 18797558]
10. Goldfried MR, Castonguay LG. Behavior therapy: redefining strengths and limitations. Behav Ther 1993;24(4):505-526 [FREE Full text] [doi: 10.1016/s0005-7894(05)80315-6]
11. Farzanfar R, Frishkopf S, Friedman R, Ludena K. Evaluating an automated mental health care system: making meaning of human–computer interaction. Comput Human Behav 2007 May;23(3):1167-1182 [FREE Full text] [doi: 10.1016/j.chb.2004.11.015]
12. Olson GM, Olson JS. Human-computer interaction: psychological aspects of the human use of computing. Annu Rev Psychol 2003 Feb;54(1):491-516. [doi: 10.1146/annurev.psych.54.101601.145044] [Medline: 12209025]
13. Alneyadi M, Drissi N, Almeqbaali M, Ouhbi S. Biofeedback-based connected mental health interventions for anxiety: systematic literature review. JMIR Mhealth Uhealth 2021 Apr 22;9(4):e26038 [FREE Full text] [doi: 10.2196/26038] [Medline: 33792548]
14. Schueller SM, Stiles-Shields C, Yarosh L. Online treatment and virtual therapists in child and adolescent psychiatry. Child Adolesc Psychiatr Clin N Am 2017 Jan;26(1):1-12 [FREE Full text] [doi: 10.1016/j.chc.2016.07.011] [Medline: 27837935]
15. Baum RA, Epstein JN, Kelleher K. Healthcare reform, quality, and technology: ADHD as a case study. Curr Psychiatry Rep 2013 Jul;15(7):369 [FREE Full text] [doi: 10.1007/s11920-013-0369-0] [Medline: 23712720]
16. Rapport MD, Chung KM, Shore G, Denney CB, Isaacs P. Upgrading the science and technology of assessment and diagnosis: laboratory and clinic-based assessment of children with ADHD. J Clin Child Psychol 2000 Dec;29(4):555-568. [doi: 10.1207/S15374424JCCP2904_8] [Medline: 11126633]
17. Nikkelen SW, Valkenburg PM, Huizinga M, Bushman BJ. Media use and ADHD-related behaviors in children and adolescents: a meta-analysis. Dev Psychol 2014 Sep;50(9):2228-2241. [doi: 10.1037/a0037318] [Medline: 24999762]

18. Scionti N, Cavallero M, Zogmaister C, Marzocchi GM. Is cognitive training effective for improving executive functions in preschoolers? A systematic review and meta-analysis. Front Psychol 2019 Jan 10;10:2812 [FREE Full text] [doi: 10.3389/fpsyg.2019.02812] [Medline: 31998168]
19. Dovis S, Van der Oord S, Wiers RW, Prins PJ. Improving executive functioning in children with ADHD: training multiple executive functions within the context of a computer game. a randomized double-blind placebo controlled trial. PLoS One 2015 Apr 06;10(4):e0121651 [FREE Full text] [doi: 10.1371/journal.pone.0121651] [Medline: 25844638]
20. Moreno-García I, Meneres-Sancho S, Camacho-Vara de Rey C, Servera M. A randomized controlled trial to examine the posttreatment efficacy of neurofeedback, behavior therapy, and pharmacology on ADHD measures. J Atten Disord 2019 Feb 01;23(4):374-383. [doi: 10.1177/1087054717693371] [Medline: 29254414]
21. Egeland J, Aarlien AK, Saunes BK. Few effects of far transfer of working memory training in ADHD: a randomized controlled trial. PLoS One 2013 Oct 4;8(10):e75660 [FREE Full text] [doi: 10.1371/journal.pone.0075660] [Medline: 24124503]
22. Steiner NJ, Frenette EC, Rene KM, Brennan RT, Perrin EC. In-school neurofeedback training for ADHD: sustained improvements from a randomized control trial. Pediatrics 2014 Mar;133(3):483-492. [doi: 10.1542/peds.2013-2059] [Medline: 24534402]
23. van der Oord S, Ponsioen AJ, Geurts HM, Ten Brink EL, Prins PJ. A pilot study of the efficacy of a computerized executive functioning remediation training with game elements for children with ADHD in an outpatient setting: outcome on parent- and teacher-rated executive functioning and ADHD behavior. J Atten Disord 2014 Nov;18(8):699-712. [doi: 10.1177/1087054712453167] [Medline: 22879577]
24. Bikic A, Leckman JF, Christensen T, Bilenberg N, Dalsgaard S. Attention and executive functions computer training for attention-deficit/hyperactivity disorder (ADHD): results from a randomized, controlled trial. Eur Child Adolesc Psychiatry 2018 Dec;27(12):1563-1574. [doi: 10.1007/s00787-018-1151-y] [Medline: 29644473]
25. Kofler MJ, Wells EL, Singh LJ, Soto EF, Irwin LN, Groves NB, et al. A randomized controlled trial of central executive training (CET) versus inhibitory control training (ICT) for ADHD. J Consult Clin Psychol 2020 Aug;88(8):738-756 [FREE Full text] [doi: 10.1037/ccp0000550] [Medline: 32700955]
26. Meyer KN, Santillana R, Miller B, Clapp W, Way M, Bridgman-Goines K, et al. Computer-based inhibitory control training in children with Attention-Deficit/Hyperactivity Disorder (ADHD): evidence for behavioral and neural impact. PLoS One 2020 Nov 30;15(11):e0241352 [FREE Full text] [doi: 10.1371/journal.pone.0241352] [Medline: 33253237]
27. Benzing V, Schmidt M. The effect of exergaming on executive functions in children with ADHD: a randomized clinical trial. Scand J Med Sci Sports 2019 Aug;29(8):1243-1253 [FREE Full text] [doi: 10.1111/sms.13446] [Medline: 31050851]
28. Azami S, Moghadas A, Sohrabi-Esmrood F, Nazifi M, Mirmohamad M, Hemmati F, et al. A pilot randomized controlled trial comparing computer-assisted cognitive rehabilitation, stimulant medication, and an active control in the treatment of ADHD. Child Adolesc Ment Health 2016 Nov;21(4):217-224. [doi: 10.1111/camh.12157] [Medline: 32680341]
29. Lim CG, Poh XW, Fung SS, Guan C, Bautista D, Cheung YB, et al. A randomized controlled trial of a brain-computer interface based attention training program for ADHD. PLoS One 2019 May 21;14(5):e0216225 [FREE Full text] [doi: 10.1371/journal.pone.0216225] [Medline: 31112554]
30. Breider S, de Bildt A, Nauta MH, Hoekstra PJ, van den Hoofdakker BJ. Self-directed or therapist-led parent training for children with attention deficit hyperactivity disorder? A randomized controlled non-inferiority pilot trial. Internet Interv 2019 Aug 08;18:100262 [FREE Full text] [doi: 10.1016/j.invent.2019.100262] [Medline: 31890615]
31. Wangler S, Gevensleben H, Albrecht B, Studer P, Rothenberger A, Moll GH, et al. Neurofeedback in children with ADHD: specific event-related potential findings of a randomized controlled trial. Clin Neurophysiol 2011 May;122(5):942-950. [doi: 10.1016/j.clinph.2010.06.036] [Medline: 20843737]
32. Bigorra A, Garolera M, Guijarro S, Hervas A. Impact of working memory training on hot executive functions (decision-making and theory of mind) in children with ADHD: a randomized controlled trial. Neuropsychiatry 2016;6(5). [doi: 10.4172/neuropsychiatry.1000147]
33. Bioulac S, Micoulaud-Franchi JA, Maire J, Bouvard MP, Rizzo AA, Sagaspe P, et al. Virtual remediation versus methylphenidate to improve distractibility in children with ADHD: a controlled randomized clinical trial study. J Atten Disord 2020 Jan;24(2):326-335. [doi: 10.1177/1087054718759751] [Medline: 29562853]
34. Cibrian FL, Lakes KD, Schuck SE, Hayes GR. The potential for emerging technologies to support self-regulation in children with ADHD: a literature review. Int J Child Comput Interact 2022 Mar;31:100421 [FREE Full text] [doi: 10.1016/j.ijcci.2021.100421]
35. Powell L, Parker J, Harpin V. What is the level of evidence for the use of currently available technologies in facilitating the self-management of difficulties associated with ADHD in children and young people? A systematic review. Eur Child Adolesc Psychiatry 2018 Nov;27(11):1391-1412 [FREE Full text] [doi: 10.1007/s00787-017-1092-x] [Medline: 29222634]
36. Romero-Ayuso D, Toledano-González A, Rodríguez-Martínez MD, Arroyo-Castillo P, Triviño-Juárez JM, González P, et al. Effectiveness of virtual reality-based interventions for children and adolescents with ADHD: a systematic review and meta-analysis. Children (Basel) 2021 Jan 21;8(2):70 [FREE Full text] [doi: 10.3390/children8020070] [Medline: 33494272]

37. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
38. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
39. Twisk J, Proper K. Evaluation of the results of a randomized controlled trial: how to define changes between baseline and follow-up. J Clin Epidemiol 2004 Mar;57(3):223-228. [doi: 10.1016/j.jclinepi.2003.07.009] [Medline: 15066681]
40. Hox J, Moerbeek M, van de Schoot R. Multilevel Analysis: Techniques and Applications. 3rd edition. Oxfordshire, UK: Routledge; 2002.
41. Corkum P, Elik N, Blotnicky-Gallant PA, McGonnell M, McGrath P. Web-based intervention for teachers of elementary students with ADHD: randomized controlled trial. J Atten Disord 2019 Feb;23(3):257-269. [doi: 10.1177/1087054715603198] [Medline: 26362259]
42. Gevensleben H, Holl B, Albrecht B, Schlamp D, Kratz O, Studer P, et al. Neurofeedback training in children with ADHD: 6-month follow-up of a randomised controlled trial. Eur Child Adolesc Psychiatry 2010 Sep;19(9):715-724 [FREE Full text] [doi: 10.1007/s00787-010-0109-5] [Medline: 20499120]
43. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, Keefe RS, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. Lancet Digit Health 2020 Apr;2(4):e168-e178 [FREE Full text] [doi: 10.1016/S2589-7500(20)30017-0] [Medline: 33334505]
44. Fried R, DiSalvo M, Kelberman C, Adler A, McCafferty D, Woodworth KY, et al. An innovative SMS intervention to improve adherence to stimulants in children with ADHD: preliminary findings. J Psychopharmacol 2020 Aug;34(8):883-890. [doi: 10.1177/0269881120908014] [Medline: 32077768]
45. Brehmer Y, Westerberg H, Bäckman L. Working-memory training in younger and older adults: training gains, transfer, and maintenance. Front Hum Neurosci 2012 Mar 27;6:63 [FREE Full text] [doi: 10.3389/fnhum.2012.00063] [Medline: 22470330]
46. Hardy KK, Willard VW, Bonner MJ. Computerized cognitive training in survivors of childhood cancer: a pilot study. J Pediatr Oncol Nurs 2011 Jan;28(1):27-33. [doi: 10.1177/1043454210377178] [Medline: 20966158]
47. Park K, Kihl T, Park S, Kim MJ, Chang J. Fairy tale directed game-based training system for children with ADHD using BCI and motion sensing technologies. Behav Inf Technol 2018;38(6):564-577 [FREE Full text] [doi: 10.1080/0144929x.2018.1544276]
48. Ra CK, Cho J, Stone MD, De La Cerda J, Goldenson NI, Moroney E, et al. Association of digital media use with subsequent symptoms of attention-deficit/hyperactivity disorder among adolescents. JAMA 2018 Jul 17;320(3):255-263 [FREE Full text] [doi: 10.1001/jama.2018.8931] [Medline: 30027248]
49. Mayer RE, Parong J, Bainbridge K. Young adults learning executive function skills by playing focused video games. Cogn Dev 2019 Jan;49:43-50 [FREE Full text] [doi: 10.1016/j.cogdev.2018.11.002]
50. Wexler BE, Vitulano LA, Moore C, Katsovich L, Smith SD, Rush C, et al. An integrated program of computer-presented and physical cognitive training exercises for children with attention-deficit/hyperactivity disorder. Psychol Med 2021 Jul;51(9):1524-1535. [doi: 10.1017/S0033291720000288] [Medline: 32090720]
51. Kotwal DB, Burns WJ, Montgomery DD. Computer-assisted cognitive training for ADHD: a case study. Behav Modif 1996;20(1):85-96 [FREE Full text] [doi: 10.1177/01454455960201004]
52. Ordikhani-Seyedlar M, Lebedev MA, Sorensen HB, Puthusserypady S. Neurofeedback therapy for enhancing visual attention: state-of-the-art and challenges. Front Neurosci 2016 Aug 03;10:352 [FREE Full text] [doi: 10.3389/fnins.2016.00352] [Medline: 27536212]
53. Leins U, Goth G, Hinterberger T, Klinger C, Rumpf N, Strehl U. Neurofeedback for children with ADHD: a comparison of SCP and Theta/Beta protocols. Appl Psychophysiol Biofeedback 2007 Jun;32(2):73-88. [doi: 10.1007/s10484-007-9031-0] [Medline: 17356905]
54. Lofthouse N, Arnold LE, Hersch S, Hurt E, DeBeus R. A review of neurofeedback treatment for pediatric ADHD. J Atten Disord 2012 Jul;16(5):351-372. [doi: 10.1177/1087054711427530] [Medline: 22090396]
55. Gomes H, Duff M, Ramos M, Molholm S, Foxe JJ, Halperin J. Auditory selective attention and processing in children with attention-deficit/hyperactivity disorder. Clin Neurophysiol 2012 Feb;123(2):293-302 [FREE Full text] [doi: 10.1016/j.clinph.2011.07.030] [Medline: 21839675]
56. Liu T, Wang J, Chen Y, Wang R, Song M. Neurofeedback treatment experimental study for ADHD by using the brain-computer interface neurofeedback system. In: Proceedings of the 2012 World Congress on Medical Physics and Biomedical Engineering. 2012 Presented at: WCMPBE '12; May 26-31, 2012; Beijing, China p. 1537-1540 URL: https://link.springer.com/chapter/10.1007/978-3-642-29305-4_404 [doi: 10.1007/978-3-642-29305-4_404]
57. Pickens TA, Khan SP, Berlau DJ. White noise as a possible therapeutic option for children with ADHD. Complement Ther Med 2019 Feb;42:151-155. [doi: 10.1016/j.ctim.2018.11.012] [Medline: 30670235]
58. Clarfield J, Stoner G. The effects of computerized reading instruction on the academic performance of students identified with ADHD. Sch Psychol Rev 2019;34(2):246-254 [FREE Full text] [doi: 10.1080/02796015.2005.12086286]

59. Chacko A, Newcorn JH, Feirsen N, Uderman JZ. Improving medication adherence in chronic pediatric health conditions: a focus on ADHD in youth. Curr Pharm Des 2010 Jul 01;16(22):2416-2423. [doi: 10.2174/138161210791959908] [Medline: 20513228]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ADHD-RS: ADHD Rating Scale
BRIEF: Behavior Rating Inventory of Executive Function
CAVLT-2: Children's Auditory Verbal Learning Test-2
Conners 3-P: Conners 3 Parent Rating Scale
Conners 3-T: Conners 3 Teacher Rating Scale
DBDRS: Disruptive Behavior Disorder Rating Scale
IVA/CPT: Integrated Visual and Auditory Continuous Performance Test
PedsQL: Pediatric Quality of Life Inventory
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference
WISC-IV: Backward Digit Span of the Wechsler Intelligence Scale for Children-IV

Edited by J Torous; submitted 01.08.23; peer-reviewed by HF Hsieh, M Sakata; comments to author 14.09.23; revised version received 05.10.23; accepted 22.10.23; published 21.11.23.

Please cite as:
Wong KP, Qin J, Xie YJ, Zhang B
Effectiveness of Technology-Based Interventions for School-Age Children With Attention-Deficit/Hyperactivity Disorder: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Ment Health 2023;10:e51459
URL: https://mental.jmir.org/2023/1/e51459
doi:10.2196/51459
PMID:37988139

Original Paper

Predicting Generalized Anxiety Disorder From Impromptu Speech Transcripts Using Context-Aware Transformer-Based Neural Networks: Model Evaluation Study

Bazen Gashaw Teferra[1], PhD; Jonathan Rose[1,2], PhD

[1]The Edward S Rogers Sr Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada

[2]The Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
Bazen Gashaw Teferra, PhD
The Edward S Rogers Sr Department of Electrical and Computer Engineering
University of Toronto
10 King's College Road
Toronto, ON, M5S3G4
Canada
Phone: 1 4169786992
Email: bazen.teferra@mail.utoronto.ca

Abstract

Background: The ability to automatically detect anxiety disorders from speech could be useful as a screening tool for an anxiety disorder. Prior studies have shown that individual words in textual transcripts of speech have an association with anxiety severity. Transformer-based neural networks are models that have been recently shown to have powerful predictive capabilities based on the context of more than one input word. Transformers detect linguistic patterns and can be separately trained to make specific predictions based on these patterns.

Objective: This study aimed to determine whether a transformer-based language model can be used to screen for generalized anxiety disorder from impromptu speech transcripts.

Methods: A total of 2000 participants provided an impromptu speech sample in response to a modified version of the Trier Social Stress Test (TSST). They also completed the Generalized Anxiety Disorder 7-item (GAD-7) scale. A transformer-based neural network model (pretrained on large textual corpora) was fine-tuned on the speech transcripts and the GAD-7 to predict whether a participant was above or below a screening threshold of the GAD-7. We reported the area under the receiver operating characteristic curve (AUROC) on the test data and compared the results with a baseline logistic regression model using the Linguistic Inquiry and Word Count (LIWC) features as input. Using the integrated gradient method to determine specific words that strongly affect the predictions, we inferred specific linguistic patterns that influence the predictions.

Results: The baseline LIWC-based logistic regression model had an AUROC value of 0.58. The fine-tuned transformer model achieved an AUROC value of 0.64. Specific words that were often implicated in the predictions were also dependent on the context. For example, the first-person singular pronoun "I" influenced toward an anxious prediction 88% of the time and a nonanxious prediction 12% of the time, depending on the context. Silent pauses in speech, also often implicated in predictions, influenced toward an anxious prediction 20% of the time and a nonanxious prediction 80% of the time.

Conclusions: There is evidence that a transformer-based neural network model has increased predictive power compared with the single word–based LIWC model. We also showed that the use of specific words in a specific context—a linguistic pattern—is part of the reason for the better prediction. This suggests that such transformer-based models could play a useful role in anxiety screening systems.

(JMIR Ment Health 2023;10:e44325) doi:10.2196/44325

KEYWORDS

mental health; generalized anxiety disorder; impromptu speech; linguistic features; anxiety prediction; neural networks; natural language processing; transformer models; mobile phone

Introduction

Background

The screening, diagnosis, and tracking of mental health disorders require frequent interactions with psychiatrists or psychologists. However, the high cost [1] and low availability of mental health professionals make frequent interactions difficult [2]. This shortage could be addressed, in part, if there is an ability to assess a mental health disorder automatically through a passive and frequent collection of patient data. One possible way to do such monitoring may be through speech, as the presence of a mental health disorder has been shown to be associated with changes in human speech [3,4].

In this study, we focused on anxiety disorders, specifically on generalized anxiety disorder (GAD) [5]. Anxiety disorders are characterized by an excessive and uncontrollable fear of what is to come and are among the most common mental health issues, with an incidence of approximately 10% in the Canadian population [6]. It may be possible to reach a much greater proportion of the population using methods that automate some aspects of the measurement and diagnosis of anxiety disorders, such as the detection of anxiety from speech.

The current gold standard diagnosis for GAD requires multiple sessions with a mental health professional where the professional compares the different symptoms exhibited by the patient with the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition diagnostic criteria for GAD [7]. One place to look for symptoms is in the linguistic patterns used by the patient as the choice of words by anxious individuals tends to be different from that of nonanxious individuals [4]. The goal of this study was to determine the accuracy of a method for the automatic detection of anxiety from the transcript of impromptu speech. We were motivated to pursue this goal, in part, because it should be possible to frequently collect speech-to-text (STT) transcripts using smartphones or other wearable devices and, therefore, to enable a system for monitoring symptoms during or after treatment.

In recent years, transformer-based [8-12] neural network models [13] have been shown to have a strong capability to predict from language, including tasks such as next-word prediction, machine translation, and sequence classification. In this study, we leveraged this capability to predict whether a participant is above or below the screening threshold for GAD.

This paper is organized as follows: the *Prior Work* subsection summarizes related work in anxiety prediction from language and provides a brief overview of transformer language models. The *Methods* section describes the speech sample collection methods and the construction, training, and evaluation of the prediction model. The *Results* section presents the prediction model's performance, whereas the *Discussion* section discusses specific patterns that were influential in the prediction.

Prior Work

Previous Work on the Automatic Prediction of Anxiety From Speech

Several prior studies have explored the automatic prediction of anxiety from speech. These studies have used both the acoustic properties as well as the linguistic features of speech and have shown some ability to detect anxiety. Most prior studies have focused on the acoustic structure of speech, that is, the nature of the audio signal itself. Comparatively less work has been done on the linguistic aspects of speech, the focus of this paper, which we describe in the subsequent paragraphs.

Di Matteo et al [14] explored the relationship between passively collected audio data and anxiety and depression. A total of 84 participants installed an Android app on their smartphone for 2 weeks. During this period, the app passively collected intermittent samples of audio data from the participants' smartphones. The audio was then converted to text, and the Linguistic Inquiry and Word Count (LIWC) [15] was used to classify the words into 67 different categories. The correlation between the LIWC scores and self-report measures was calculated for social anxiety disorder (SAD), GAD, depression, and functional impairment. A significant correlation was observed between words related to the perceptual process ("See" in the LIWC) and SAD (r=0.31; P=.003). In addition, words related to reward were significantly correlated with GAD (r=−0.29; P=.007).

Anderson et al [16] recruited 42 participants diagnosed with SAD and 27 healthy controls to explore the differences in the words used between these 2 groups using the LIWC features. An anxiety-stimulating task was performed in which the participants were asked to write about an autobiographical and socially painful memory, which required them to recall a social humiliation, embarrassment, or shame. The word count in each of the LIWC categories was generated, including first-person singular pronouns, anxiety-related words, and fear-related words. The patients with SAD used more first-person singular pronouns (I, me, and mine), anxiety-related words, sensory or perceptual words, and words denoting physical touch but made fewer references to other people than the healthy controls.

Hofmann et al [17] examined the association between linguistic features and SAD. They recruited 24 participants diagnosed with SAD and 21 healthy controls. The participants were asked to provide a speech on any topic of their choice for a total of 4 minutes in front of an experimenter while being video recorded. To induce stress and anxiety in the participants, they were told that a panel of judges would rate their speech after it was recorded on the basis of poise, social confidence, and general presentation skills. The speech was transcribed, and LIWC was used to extract the count of the words in the following categories: first-person pronouns, negative emotion words, and positive emotion words. The results showed that the patients with SAD used more positive emotion words than the healthy controls. The authors did not observe any significant difference for the other explored LIWC categories.

Sonnenschein et al [18] explored the transcripts from passively recorded therapy sessions of 85 patients. These patients were

categorized into 3 groups: those diagnosed with anxiety but not depression, those diagnosed with depression but not anxiety, and those diagnosed with both anxiety and depression. From the transcripts, the LIWC score was generated in 4 categories: first-person singular, sad, anxiety, and filler. The group with depression but not anxiety showed a higher use of sad words than the group with anxiety but not depression. The group with anxiety but not depression showed a higher use of anxiety-related words than the group with depression but not anxiety. The *both anxious and depressed* group also showed a higher use of "sad" words than the group with anxiety but not depression. None of the other LIWC categories explored showed a significant difference.

Rook et al [19] attempted to predict GAD from linguistic patterns because they believed that the worrying behavior in GAD comes from the verbal linguistic process. A total of 142 undergraduate participants (n=56, 39.4% men and n=86, 60.6% women) were recruited for their study and were asked to recall and write about an anxious experience during their university life. The Generalized Anxiety Disorder 7-item (GAD-7) scale score and behavioral inhibition/behavioral approach system (BIS/BAS) scale score were used as the label for each of the participants. The LIWC features [15] were extracted from the texts written by the participants. Another set of features was also used by combining the LIWC features with the BIS/BAS scores. Several machine learning models were explored, including support vector machine (SVM) with linear kernel, logistic regression, naive Bayes, and random forest. Their results showed that all models built using the LIWC features performed significantly better than a random model (average precision~0.61; average recall~0.6) and achieved a higher performance (except for the SVM model) when the LIWC features were used together with the BIS/BAS scores as input features (average precision~0.65; average recall~0.64).

Gruda and Hasan [20] explored the prediction of anxiety from microblogs such as tweets using machine learning approaches. The authors started by labeling 600 tweets on a 4-point anxiety level using the short version of the traditional full-scale State-Trait Anxiety Inventory [21]. Then, a machine learning model was trained using features extracted from the textual content. The features used include a semantic embedding vector, which is the mean of multiple word vectors that map words to a vector. They also used the count of specific words and emojis as another type of features. They achieved an R^2 of 0.49 between the human label and the predicted label after training a Bayesian ridge regression [22] model. The authors then compared their model with a model that classifies a tweet as anxious or not based on the presence of anxiety-type words and negative emotion–type words, which was acquired using the LIWC library. The method that used the LIWC features to classify between anxious and nonanxious tweets achieved an R^2 of 0.21, indicating the importance of the meaning of words represented by word vectors.

A precursor study to this work [23] identified both acoustic features and linguistic features using LIWC that significantly correlated with the GAD-7. Using these features, in another study [24], a logistic regression model was trained to predict whether a participant was above or below the screening threshold for GAD based on the GAD-7. Using both the acoustic and linguistic features, we achieved a mean area under the receiver operating characteristic curve (AUROC) of 0.59.

Note that this previous study [24] and the other prior work, described in the previous paragraphs, explored the count of single words (using the LIWC) to find an association with anxiety or to predict anxiety. However, there are some studies that found specific word categories to be associated with anxiety, whereas others found no such association. For example, the studies by both Di Matteo et al [14] and Anderson et al [16] found that the word categories for "perceptual process" were associated with anxiety, whereas no other prior studies did so. Similarly, the first-person singular pronoun category was associated with anxiety only in the studies by Anderson et al [16] and Teferra et al [23] and nowhere else. These inconsistencies may be explained if the context for the specific words is taken into account—or, in other words, if the evaluation model is context aware. In this study, we hypothesized that there is a greater predictive power in examining the larger context of multiple words than in examining single words using LIWC. The former can be done using recent advances in natural language processing (NLP) [8], which has new powerful methods of converting language into numerical quantities that represent meaning and learning features that are patterns of those meanings.

Furthermore, note that the largest sample size among the previously explored studies (excluding our own [23,24]) was 142. This limits the potential for generalizability to a larger population. In this study, we used a much larger data set based on speech samples from a total of 2000 people.

Transformers and NLP

Over the last 5 years, substantial advances have been made in the field of NLP [25]. A key advance was the invention of limited-size word *vectors* or *embeddings*, through which it has been shown that a small-sized (from 50 to 300) vector of real numbers was capable of representing the meaning of individual words or parts of words [26-28]. Note that sometimes, words are divided into subparts and then converted into tokens, which can represent either a full or a partial word. These word or token vectors make it possible to determine whether 2 words have similar meaning through a numerical comparison, as well as other encapsulations of meaning through calculation. This invention also permitted the use of neural networks to process language in a far more effective way and has led to major advances in the subfields of speech recognition, natural language understanding, question answering, and language generation [26,29].

Another important step that has dramatically improved the state of the art in these fields is the advent of the transformer-based neural network models [8,10-12,30]. These so-called large language models are trained using massive corpora of text, often obtained from the internet. More specifically, the "learning" (in the machine learning sense [31]) is done by either predicting the next word in sequence or predicting intentionally missing words. The architecture of a transformer-style neural network has 2 important properties. First, it "transforms" a sequence of

words or parts of words, represented as vectors, into another sequence of vectors. The output vectors account for additional meaning inferred from the full sequence of words and thus create a sequence of so-called contextual embeddings that better encapsulate the meaning of the full input sequence. Second, it makes use of an important neural network mechanism known as "attention" [8,32]. In this, a part of the network learns several different ways in which parts of a sentence or paragraph are related to other parts of the sentence. For example, a certain word or meaning may typically be connected with specific other words in a sentence. A transformer learns many such relationships, which makes it capable of classifying the broader meaning of a sentence or paragraph. It is this capability that we leverage in this study to look for patterns of language that indicate the presence of anxiety.

There now exist many such large language models that have already been fully "pretrained" on massive corpora of text gathered from a number of sources on the internet and elsewhere [10,12,30]. A common use case in the field of deep learning and NLP is to take such pretrained models and "fine-tune" them for a specific prediction task that takes language as input. To "fine-tune" a model means to train it on a (typically much smaller) data set to learn the task at hand. The task described in the subsequent section is the classification of participants into anxious or nonanxious categories.

Methods

Data Collection

Recruitment and Demographics

We note that this study used the same participants and data as 2 earlier studies [23,24]. This study performed a novel analysis of these data using a transformer-based neural network.

The participants were recruited using Prolific [33], a web-based human participant recruitment platform. The inclusion criteria were an age range of 18 to 65 years; fluency in English; English as a first language; and the completion of at least 10 previous studies on Prolific, with 95% of these previous Prolific tasks completed satisfactorily (as labeled by the study author). The data set was also balanced for sex (n=1000, 50% female and n=1000, 50% male).

The participants who completed the study were paid £2 (approximately CAD $3.41 and US $2.74) for approximately 15 minutes of work. They completed the entire study remotely using their PCs.

Ethics Approval

This study was approved by the University of Toronto Research Ethics Board (protocol #37584).

Study Procedure

The participants were recruited for a 10- to 15-minute task implemented through a custom website. An earlier study that determined the correlates of anxiety [23] described the data collection procedure in detail. The parts of the data collection procedure that are relevant for the purpose of this study are presented in the following paragraphs.

On the Prolific platform, the participants who met the inclusion criteria were presented with the opportunity to participate in this study. Those who wished to participate clicked on the study link, which brought them to a consent form that described the procedure and goals of the study and provided information on data privacy. If a participant granted consent, a hyperlink brought them to an external web application that implemented the tasks described subsequently.

The participants were asked to fill out the standard GAD-7 questionnaire [34], which is described in more detail in the *Anxiety Measures* section. Then, they were asked to perform a speech task, which was both audio and video recorded using their computer's microphone and camera. The speech task followed a modified version of the widely used Trier Social Stress Test (TSST) [35], which aims to evoke a moderate amount of stress from each participant. Prior studies [36,37] have shown a higher activation in participants with relatively higher anxiety after they experienced moderate stress induced by the TSST.

In the modified version of the TSST, the participants were told to imagine that they were a job applicant invited for an interview with a hiring manager. They were told to imagine that it was a job that they really wanted—their so-called dream job. They were given a few minutes to prepare—to choose their dream job—and to think about how they would convince an interviewer that they were the right person for that position. The participants were also told that the recorded video would be viewed by researchers studying their behavior and language. The participants were then asked to speak for 5 minutes, making the case for themselves to be hired for that dream job.

Note that, in the original TSST [35], participants would normally deliver their speech in front of a live panel of judges. If a participant finished their delivery in <5 minutes, the judges in the original TSST design would encourage the participant to keep speaking for the full 5 minutes. For example, in the original TSST, to encourage the participants, they were asked the following question: "What are your personal strengths?" In the modified TSST, we implemented a similar method to encourage the participants to speak for the full 5 minutes: when our system detects silence (defined as the absence of speech for >6 seconds), it will display several different prompts inviting the participants to keep speaking on different topics relating to the task. Finally, note that the modified TSST only included the first task of the original TSST, not the second task, which involves the performance of mental arithmetic.

Anxiety Measures

Our goal was to predict, based on the transcript of the language spoken, whether a participant was above or below the screening threshold for GAD based on the GAD-7 scale. The GAD-7 [34] scale is a 7-item questionnaire that asks participants how often they were bothered by anxiety-related problems during the previous 2 weeks. Although the 2-week period suggests that the GAD-7 measures a temporary condition, a GAD diagnosis requires a 6-month duration of symptoms [7,38]. However, the GAD-7 has been validated as a diagnostic tool for GAD using a value of 10 as the cutoff threshold, with a sensitivity of 89% and a specificity of 82% [34]. Thus, we chose to use the GAD-7

threshold of 10 to obtain a binary label of GAD as our indicator of anxiety.

Each of the 7 questions on the GAD-7 has 4 options for the participant to select from, indicating how often they have been "bothered" by the 7 problems listed. These options and their numerical ratings are 0=not at all, 1=several days, 2=more than half the days, and 3=nearly every day. The final GAD-7 score is a summation of the values for all the questions, giving a severity measure for GAD in the range from 0 (no anxiety symptoms) to 21 (severe anxiety symptoms).

Construction and Evaluation of the Baseline Classification Model

In this section, the inputs, structure, and evaluation of a baseline model are described. The inputs to this model were the linguistic features acquired using LIWC [15]. LIWC is based on the count of words from a given transcript that fall into different preset categories. An example category is "negemo," which comprises words (such as hurt, ugly, and nasty) that are associated with negative emotion. The full set of categories in LIWC can be found in the study by Pennebaker et al [15].

The transcript was generated from the speech samples using Amazon Web Services STT system (Amazon.com, Inc) [39]—the transcription accuracy on a written text had an average word error rate (WER) of 7% (SD 4.6%). In our earlier study [23], we identified LIWC features that had a significant (P<.05) correlation with the GAD-7. These features are listed in Table 1. These were the features that were used as the input to the baseline prediction model.

A logistic regression model was trained to make predictions between the anxious and nonanxious classes. The construction and evaluation steps were as follows. First, the input features were normalized so that each feature would have a mean of 0 and an SD of 1. Next, the data were undersampled to equalize representation from both the anxious and nonanxious classes. This avoids the problem of class imbalance, which, if it occurs, causes low predictive accuracy for the minority class (which is the anxious class in our case). To undersample the data, samples were randomly selected and removed from the majority class until the majority class had an equal number of samples as the minority class.

The model construction and training steps used 3 data sets: a training data set (80% of the entire subsampled data), which was used to train the model; a validation data set (20% of the training data), which was used to select the best hyperparameters during training; and a test data set (20% of the entire subsampled data that were not included in the training data set), which was used to evaluate the performance of the trained model using the AUROC metric. This methodology—the careful separation of the training and validation data from the test data—is standard in the machine learning community [31].

Table 1. Correlation of significant Linguistic Inquiry and Word Count features with the Generalized Anxiety Disorder 7-item scale.

Feature	*r*	*P* value
AllPunc	0.13	<.001
Word Count	−0.12	<.001
Period	0.12	<.001
Assent	0.10	<.001
Negemo	0.10	<.001
Relativ	−0.09	<.001
Motion	−0.08	<.001
Swear	0.08	<.001
Anger	0.08	<.001
Focusfuture	−0.07	.003
Adverb	−0.07	.004
Time	−0.07	.004
Function	−0.07	.005
Negate	0.07	.006
Prep	−0.06	.007
WPS[a]	−0.06	.007
Anx	0.06	.008
Hear	0.06	.01
Death	0.06	.01
Ipron	−0.06	.01
See	−0.06	.01
Affect	0.06	.02
I	0.05	.02
Family	0.05	.02
Sad	0.05	.03
Ppron	0.05	.03
Space	−0.05	.04
Article	−0.05	.04
Leisure	0.05	.04
Friend	0.05	.047

[a]WPS: words per sentence.

Construction and Evaluation of the Transformer-Based Model

The advent and remarkable success of transformer-based neural networks for NLP is discussed in the *Prior Work* section. A property that distinguishes different transformer models is the number of textual words or tokens that will fit into the contextual window that the model can consider at one time, which itself is limited by the computational burden of the key method of attention [8]. These windows range in size from 512 tokens [10] to 4096 tokens [30].

The modified TSST that provided the input to our model required the participants to speak for 5 minutes, which produced transcripts ranging in size from 15 to 1190 (mean 707, SD 183) tokens. Therefore, our model required a transformer model that can process sequences of this length. We selected the transformer model known as Longformer (obtained from the HuggingFace model hub [40]) because it has a contextual window of size 4096 tokens (recall that tokens are either words or parts of a word).

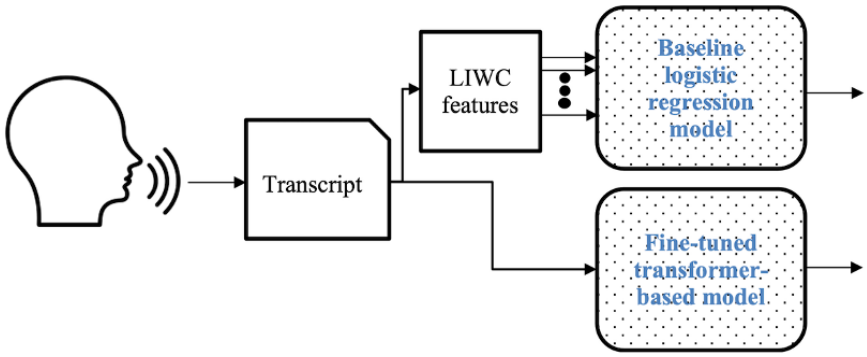
We *fine-tuned* a pretrained version of Longformer (as described in the *Prior Work* section) to create a classifier for the anxiety classification task. This process took a pretrained model and attached it to an untrained (and much smaller) neural network called a "classification head." The pretrained model together with the sequence classification head was then *fine-tuned* on

the specific task of predicting whether a participant is above or below the screening threshold for GAD based on the GAD-7 scale.

The input data set was processed in a similar way to how the baseline model was processed. Beginning with the set of transcripts from all the participants, the data were first undersampled to equalize the representation from both the anxious and nonanxious classes. The model fine-tuning step also used 3 data sets: a training data set (80% of the full data); a validation data set (20% of the training data); and a test data set (20% of the full data set), which was used to evaluate the performance of the trained model using the AUROC metric.

The overall structure of both the baseline logistic regression model and the fine-tuned transformer-based model is shown in Figure 1.

Figure 1. Overall structure of the baseline classification model and fine-tuned transformer-based model. LIWC: Linguistic Inquiry and Word Count.



Transformer Model Interpretation

Deep neural networks [13], including the transformer network used in this study, do not lend themselves to an easy explanation of which features or factors are important for any specific prediction. This contrasts with the logistic regression model (the baseline) in which the weights on each feature are informative. This study endeavored to provide some interpretation of the results of the transformer model, particularly, to provide insights into which words or group of words were the most influential in the model's prediction of anxious and nonanxious classes when given a specific transcript.

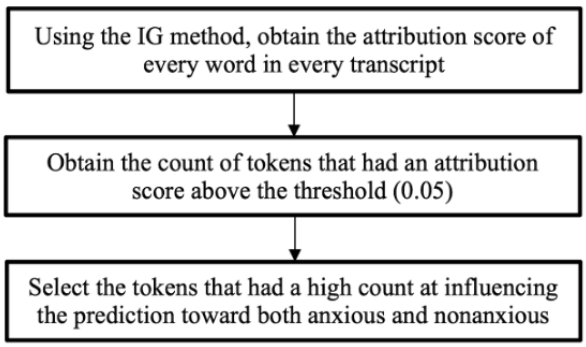
To achieve this model interpretation, we used a method known as integrated gradient (IG) [41]. IG computes a score for every input (word or token) to the model. The score is a function of the rate of change of the prediction with respect to that specific input. When the score of specific input is higher and positive, it is an indication that the input had more influence toward producing a positive classification (which is the anxious class in our case). Similarly, a high negative score indicates a strong influence toward the negative, nonanxious case. This score is referred to as the *attribution* score of the input token. We used a library called Transformer Interpret [42] to compute the attribution score for each word in a given transcript.

Using the attribution score, we can report specific words or tokens that are influential in the prediction of both anxious and nonanxious cases. From there, we explored the specific context of those words to look for patterns of language that were influential. The description in the following paragraphs provides the specific method for selecting words and identifying patterns.

First, the attribution score of each word or token in all the transcripts from all the participants was computed. In the plot of the distribution of the number of words with each score, the knee of the distribution appeared around a threshold attribution score of 0.05, which provided a tractable number of words to explore. The tokens with scores above the threshold of 0.05 are presented in the *Results* section. A summary of the steps we took to get the list of words is shown in Figure 2.

To determine whether there were patterns in the context surrounding the high-attribution words, we manually reviewed the surrounding context of each high-attribution word. The patterns we observed from these contexts, together with the specific direction of the prediction (anxious or nonanxious), are presented in the *Results* section.

Figure 2. Steps to obtain the list of tokens with a high attribution score and high count at influencing the prediction toward both anxious and nonanxious. IG: integrated gradient.



Results

Recruitment and Data Inclusion

A total of 4542 participants accepted the offer from the Prolific recruitment platform to participate in this study. Of them, 2212 participants finished the study, giving a recruitment yield of 48.7%. Of the 2212 participants who completed the study, 2000 provided acceptable submissions (and thus received payment), giving a submission-to-approval yield of 90.4%. To be clear, the recruitment continued until 2000 acceptable submissions were received. The reasons for which submissions were deemed unacceptable include the following: a missing video, missing or grossly imperfect audio, and failure to complete the task. The recruitment period lasted from November 23, 2020, to May 28, 2021. We note that the recruitment was conducted during the COVID-19 pandemic.

Data Overview

Of the 2000 participants, 620 (31%) were above the GAD-7 screening threshold of 10 and 1380 (69%) were below the screening threshold of 10. Henceforth, the participants with a GAD-7 score ≥10 are referred to as the anxious group, and those with a GAD-7 score <10 are referred to as the nonanxious group. As described in the *Methods* section, to have an equal representation of the anxious and nonanxious classes, the nonanxious group was undersampled, resulting in the inclusion of a total of 1240 participants (620 anxious and 620 nonanxious) in our analysis.

Classification Model Performance

This section presents the AUROC of the 2 binary classification models that classify anxious and nonanxious groups. The first model is the logistic regression model that uses the LIWC features as input, which is the baseline model described earlier. The LIWC features used were the ones shown to be significantly correlated with the GAD-7 in our earlier study [23], as listed in Table 1. Note that we also explored other machine learning models such as SVM, decision tree, random forest, multilayer perceptron, but these did not perform better than the baseline logistic regression model. The second model is the fine-tuned transformer-based model. The AUROC curve value for the logistic regression model that uses the LIWC features as input was 0.58 and for the transformer-based model was 0.64. Figures 3 and 4 present the receiver operating characteristics curves.

Figure 3. Area under the receiver operating characteristic curve (AUROC) of the baseline logistic regression model.

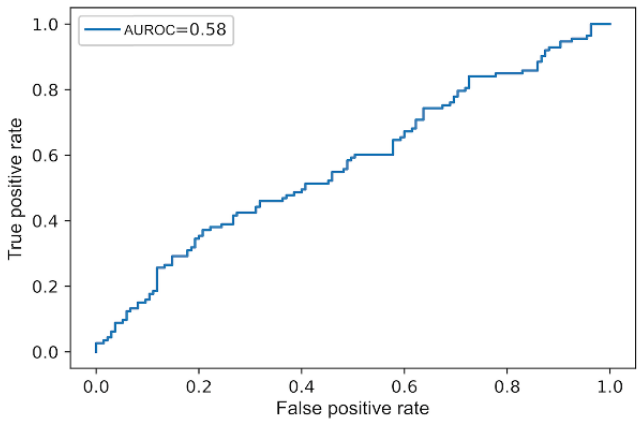
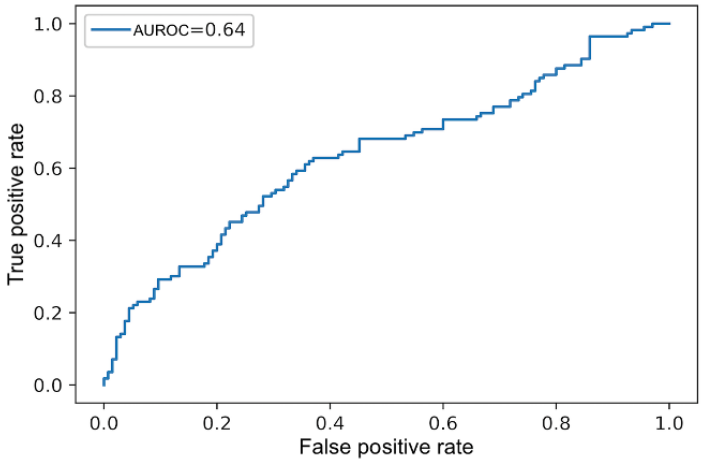


Figure 4. Area under the receiver operating characteristic curve (AUROC) of the fine-tuned transformer-based model.



Model Interpretation: Tokens Used to Predict Both Anxious and Nonanxious

In the *Transformer Model Interpretation* section, we described the IG method that was used to determine an attribution score for each word in a transcript. That score gives an indication of how strongly the word is implicated in the prediction toward anxious (if positive) or nonanxious (if negative). Table 2 presents the number of times (across all transcripts) that a specific token (listed in the first column) had a high attribution score (absolute value >0.05, as described earlier) based on the IG method. The tokens presented in Table 2 were selected because they had a high count in having both high positive and high negative attribution scores, that is, at predicting both anxious and nonanxious. Note that tokens could be words, parts of a word, or characters (eg, the STT system we used generates a "." to indicate silent pauses in speech).

Table 3 presents the patterns we observed with examples taken from the actual transcripts of the recruited participants where the same token influenced the prediction toward anxious in some cases and toward nonanxious in other cases. The first column lists these tokens, indicates the direction (anxious or nonanxious) in which they influenced the prediction, and describes the pattern of the context that we inferred was relevant using the qualitative analysis described in the *Methods* section. The second column provides a specific example of that pattern, taken from the transcripts, and the third column provides the number of occurrences of that pattern across all the transcripts.

Table 2. Tokens with high attribution scores and high counts of prediction influence.

Token	Times influencing toward anxious, n (%)	Times influencing toward nonanxious, n (%)
I (n=3459)	3032 (87.65)	427 (12.35)
[Silent pause][a] (n=14,490)	2933 (20.24)	11,557 (79.76)
[Filled pause][b] (n=3434)	2039 (59.38)	1395 (40.62)
And (n=1595)	913 (57.24)	682 (42.76)

[a][Silent pause]: a silent pause in speech, as determined by the speech-to-text software.

[b][Filled pause]: a pause consisting of filler words such as "um," "mm," "uh," "hmm," or "mhm."

Table 3. Cases in which the tokens in influenced the prediction of both anxious and nonanxious.

Token, prediction class, and definition of pattern	Example of pattern	Occurrences across all transcripts, n
I		
Anxious		
"I" followed by a filled pause[a]	I have um I worked very well	476
"I" is the first word in a sentence but in the middle of the transcript	I get on well with various different groups	1567
Starting a sentence and pausing after just saying "I"	I [Silent pause]	208
"I" used together with am or have	I am able to relate	1515
Nonanxious		
"I" used in a sentence to reference others	I was able to remember all their names	47
"I" is the very first word in the transcript	<speech starts> I think I would be perfect for this job	171
"I" used to describe a positive thing about oneself	I am imaginative	77
[Silent pause][b]		
Anxious		
[Silent pause] used before or after a [Filled pause]	[Silent pause] um mm [Silent pause]	1740
Starting a sentence and pausing within a short period	my [Silent pause]	2057
Nonanxious		
Pauses during speech that are not accompanied by a [Filled pause] and produce a correct sentence	bring a specific [Silent pause] area of expertise of functionality	11,557
[Filled pause][a]		
Anxious		
[Filled pause] used together with a [Silent pause]	[Silent pause] um mm [Silent pause]	1577
[Filled pause] used in the beginning of a speech	<speech starts> hello um I just like to	23
Nonanxious		
Filled pause used in the middle of a sentence without a silent pause	many years playing music at parties *um* starting at the age of	480
And		
Anxious		
Finishing a sentence with "and"	really think about it in detail and	519
Using "and" more than once in a sentence	was tasked in doing that and and I did that successfully and that	187
Starting a sentence and pausing after just saying "and"	and [Silent pause] sometimes things are	282
Nonanxious		
"and" used grammatically correctly in a sentence	eight people for twelve years and after that I managed an additional	572

[a][Filled pause]: a pause consisting of filler words such as "um," "mm," "uh," "hmm," or "mhm."

[b][Silent pause]: a silent pause in speech, as determined by the speech-to-text software.

Discussion

The goal of this study was to determine how well a transformer-based neural network model can predict GAD and compare it to the performance of an LIWC-based logistic regression predictor. In this section, we discuss the implications of the findings presented in the *Results* section, as well as the limitations of the study.

Principal Findings

Recruitment and Data Overview

Results presented in the *Data Overview* section indicates that a substantially larger number of participants screened positive for GAD compared with the prevalence rate of 10% in the general population [6]. This suggests that participants recruited from Prolific are more likely to experience anxiety, which is consistent with previous research using participants from Prolific

[23,43,44]. Another possible reason for a higher number of anxious participants is the recruitment period (November 23, 2020, to May 28, 2021), which coincided with the COVID-19 pandemic. More demographic information can be found in our earlier published papers [23,24].

Classification Model Performance

The logistic regression model with LIWC features is the baseline point of comparison. This model performed better than the random model (as it has an AUROC of >0.5). This indicates that the count and type of words used by individuals do provide some insights into their anxiety, which is in line with prior work [14,16-19] that explored the association between LIWC features and anxiety.

The performance of the fine-tuned transformer model was larger than the baseline model by 10%—suggesting that it is context aware. We believe that a model that considers context can achieve higher predictive performance. This suggests that transformer models, which search for multiword contexts to find patterns, can extract more information for prediction than single word–based models. The results presented in Tables 2 and 3 allow us to understand, in more detail, what the fine-tuned transformer model based its predictions on, as discussed in the subsequent section.

Furthermore, we note that it is possible to increase the probability of correct prediction by incorporating acoustic features in the prediction of the transformer-based model as well as by using multiple measurements if the circumstances of the measurement system permit it. This would be the case if this kind of a model is applied to passively collected speech, and we could sample the speech and measure it over time. In that case, one could survey the multiple measurements and select the majority result (anxious or nonanxious) that has been predicted as the true result. This approach works under the assumption that each measurement from a different speech sample is independent and will work less well as a function of independence. We have discussed this approach in more detail in our earlier paper [24].

Model Interpretation

In this section, we discuss our attempt to provide an interpretation of the results from the transformer model. Table 2 shows the tokens with a high attribution score, as defined earlier, and a high count at influencing the prediction toward anxious and nonanxious. The first entries in Table 3 describe the effects of the singular pronoun "I." Depending on the context, the use of the word "I" influences *either* toward an anxious prediction or toward a nonanxious prediction. By contrast, previous studies have shown an increased use of "I" to be associated only toward the direction of anxiety [16]. A possible reason why "I" is associated with anxiety is because individuals with anxiety will try to divert their attention from anxiety-inducing events by focusing on themselves. This might result in the frequent use of "I" in their speech.

However, this study shows how the context around the word "I" matters—although its presence influenced the prediction toward anxiety for the majority of the cases (88%), it also influenced the prediction toward nonanxious in 12% of the cases. A pattern around "I" that influenced the prediction toward nonanxious is when it was used to reference others (eg, "I was able to remember all *their* names"). This is opposite to the case where anxious individuals tend to focus on themselves and hence a possible reason as to why focusing on others would influence the prediction toward nonanxious. Another pattern of "I" that influenced the prediction toward nonanxious is when it was one of the very first words at the beginning of speech (ie, at the very beginning). This may be because confident people might start their speech by introducing themselves or placing the focus on themselves before proceeding with whatever the subject matter of their speech is. Similarly, relating to confidence, there is a pattern where "I" was used to say something positive about oneself, which influenced the prediction toward nonanxious. These cases suggest that confidence is related to the state of being nonanxious.

Silent pauses ([Silent pause] in Tables 2 and 3) mainly influenced the prediction toward nonanxious, for 80% of the cases. This is in line with prior work [45], which indicated that anxiety is associated with a reduction in the number of silent pauses during speech. The authors suggested that pausing during speech represents a cognitive activity that is observed more in nonanxious individuals than in anxious individuals.

However, there were also times when a silent pause influenced the prediction toward anxiety. The difference was the context: when a silent pause was used together with a filled pause and pausing after saying a single word. These cases hint toward difficulty in producing complete sentences and instead using filler words in the middle of their speech or inability to finish a sentence. This might be because of a higher level of anxiety.

The other 2 types of tokens presented in Table 2 ([Filled Pauses] and "and") had a high count in influencing the prediction toward both anxious and nonanxious. We believe that they have a high count because they are commonly used tokens in STT transcripts. A pattern that stood out around both ([Filled Pauses] and "and") types of tokens is the use of grammatically correct language, which was exhibited more by the participants without anxiety. Prior work [46] suggests that anxiety causes disfluencies in speech, which, therefore, could be a possible explanation for the use of grammatically incorrect language by the participants with anxiety. Our results suggest that the model is picking up on this grammatical incorrectness.

Limitations

One limitation of this study is the accuracy of the STT transcription. In this study, we used Amazon's STT program [39], which had good transcription accuracy, with an average WER of 7% (SD 4.6%). The fact that the WER is not 0 means that we obtain the wrong transcription for some words, and our model might make a wrong prediction based on these words. However, we speculate that because the STT software is improving each year, the WER would become closer and closer to 0, so the prediction of a model based on these transcripts would also improve.

Another limitation of this study is the use of a modified version of the TSST. In the original TSST, participants are asked to describe why they should be hired for their dream job in front

of a live panel of judges. However, in our study, we asked the recruited participants to describe why they should be hired for their dream job in front of a camera at their own location. This is a limitation in achieving the full replication of the TSST as a stress induction task. Nonetheless, we had an internal check where we asked them how anxious they felt before and after the TSST task (more information can be found in our earlier published study [23]), and we observed, on average, a 25% increase in the participants' level of anxiety.

Another limitation is the use of self-report measures to assess GAD. Self-report measures are subjective opinions that individuals have about themselves and may not completely capture clinical symptoms. Ideally, we would want the gold standard label for determining whether a participant has GAD. This is acquired through a one-on-one session between a patient and clinician where the clinician analyzes the patient's behavior to identify possible symptoms of GAD according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [7], but this is clearly much more expensive to acquire.

Another limitation is the subjective or qualitative nature of pattern detection, which is presented in Table 3 and forms the basis of the insights in the *Discussion* section. As described in the *Methods* section, the transcripts were analyzed manually, and instances that we believed exhibited similar patterns across multiple contexts were selected. These were our subjective opinions of what constituted a similar pattern; therefore, other researchers might be able to find other patterns that we might have overlooked. In future studies, we aim to release our transcripts for other researchers to go through as we did and see whether any other interesting patterns could be detected.

Conclusions

In this paper, we have presented the results of a large-sample study that aimed to predict whether participants who provided speech samples fell below or above the screening threshold for GAD based on the GAD-7 scale. More specifically, we investigated the importance of multiword context when predicting the presence or absence of anxiety. Although prior studies have shown that the choice of individual words is a good predictor of mental health disorders, we have shown that the choice of words together with the context is an even better predictor. Furthermore, transformer-based neural network models can be leveraged to find such linguistic patterns that help identify whether a certain word, given the context, would predict anxiety. There is a type of transformer-based model recently published in the literature [47], which is a model pretrained on a mental health corpus (focusing on depression and suicidality). Therefore, we recommend that future studies explore the linguistic patterns of speech identified using transformer models and apply them to the screening of different types of mental health disorders.

Acknowledgments

This research was funded by a University of Toronto XSeed grant, Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2019-04395), and Social Sciences and Humanities Research Council Partnership Engage Grant (892-2019-0011).

The authors are grateful to Professor Ludovic Rheault, Professor Sophie Borwein, and Dr Danielle D DeSouza for their energy, assistance with the launch of this project and data collection, and advice.

Conflicts of Interest

None declared.

References

1. Koerner N, Dugas MJ, Savard P, Gaudet A, Turcotte J, Marchand A. The economic burden of anxiety disorders in Canada. Canadian Psychology / Psychologie canadienne 2004 Aug;45(3):191-201. [doi: 10.1037/h0088236]
2. Roberge P, Fournier L, Duhoux A, Nguyen CT, Smolders M. Mental health service use and treatment adequacy for anxiety disorders in Canada. Soc Psychiatry Psychiatr Epidemiol 2011 Apr;46(4):321-330. [doi: 10.1007/s00127-010-0186-2] [Medline: 20217041]
3. Thompson AR. Pharmacological agents with effects on voice. Am J Otolaryngol 1995;16(1):12-18. [doi: 10.1016/0196-0709(95)90003-9] [Medline: 7717466]
4. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language. use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
5. Hidalgo RB, Sheehan DV. Generalized anxiety disorder. Handb Clin Neurol 2012;106:343-362. [doi: 10.1016/B978-0-444-52002-9.00019-X] [Medline: 22608630]
6. Mental Health - Anxiety Disorders. Government of Canada. URL: https://www.canada.ca/en/health-canada/services/healthy-living/your-health/diseases/mental-health-anxiety-disorders.html [accessed 2022-01-24]
7. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, D.C., United States: American Psychiatric Association; 2013.
8. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. arXiv 2017 [FREE Full text]

9. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: state-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020 Presented at: 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; Nov 8-12, 2020; Online. [doi: 10.18653/v1/2020.emnlp-demos.6]
10. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv 2019 [FREE Full text]
11. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. GitHub. URL: https://life-extension.github.io/2020/05/27/GPT%E6%8A%80%E6%9C%AF%E5%88%9D%E6%8E%A2/language-models.pdf [accessed 2022-11-15]
12. Floridi L, Chiriatti M. GPT-3: its nature, scope, limits, and consequences. Minds Mach 2020 Nov 01;30(4):681-694. [doi: 10.1007/s11023-020-09548-1]
13. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
14. Di Matteo D, Wang W, Fotinos K, Lokuge S, Yu J, Sternat T, et al. Smartphone-detected ambient speech and self-reported measures of anxiety and depression: exploratory observational study. JMIR Form Res 2021 Jan 29;5(1):e22723 [FREE Full text] [doi: 10.2196/22723] [Medline: 33512325]
15. Pennebaker J, Boyd R, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. The University of Texas at Austin. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [accessed 2022-01-24]
16. Anderson B, Goldin PR, Kurita K, Gross JJ. Self-representation in social anxiety disorder: linguistic analysis of autobiographical narratives. Behav Res Ther 2008 Oct;46(10):1119-1125 [FREE Full text] [doi: 10.1016/j.brat.2008.07.001] [Medline: 18722589]
17. Hofmann SG, Moore PM, Gutner C, Weeks JW. Linguistic correlates of social anxiety disorder. Cogn Emot 2012;26(4):720-726 [FREE Full text] [doi: 10.1080/02699931.2011.602048] [Medline: 21851248]
18. Sonnenschein AR, Hofmann SG, Ziegelmayer T, Lutz W. Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. Cogn Behav Ther 2018 Jul;47(4):315-327. [doi: 10.1080/16506073.2017.1419505] [Medline: 29345528]
19. Rook L, Mazza MC, Lefter I, Brazier F. Toward linguistic recognition of generalized anxiety disorder. Front Digit Health 2022;4:779039 [FREE Full text] [doi: 10.3389/fdgth.2022.779039] [Medline: 35493530]
20. Gruda D, Hasan S. Feeling anxious? Perceiving anxiety in tweets using machine learning. Comput Human Behav 2019 Sep;98:245-255. [doi: 10.1016/j.chb.2019.04.020]
21. Skapinakis P. Spielberger state-trait anxiety inventory. In: Encyclopedia of Quality of Life and Well-Being Research. Dordrecht: Springer; 2014.
22. MacKay DJ. Bayesian interpolation. Neural Computation 1992 May;4(3):415-447. [doi: 10.1162/neco.1992.4.3.415]
23. Teferra BG, Borwein S, DeSouza DD, Simpson W, Rheault L, Rose J. Acoustic and linguistic features of impromptu speech and their association with anxiety: validation study. JMIR Ment Health 2022 Jul 08;9(7):e36828 [FREE Full text] [doi: 10.2196/36828] [Medline: 35802401]
24. Teferra BG, Borwein S, DeSouza DD, Rose J. Screening for generalized anxiety disorder from acoustic and linguistic features of impromptu speech: prediction model evaluation study. JMIR Form Res 2022 Oct 28;6(10):e39998 [FREE Full text] [doi: 10.2196/39998] [Medline: 36306165]
25. Jurafsky D, Martin J. Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Hoboken, New Jersey, U.S: Pearson Prentice Hall; 2008.
26. Bengio S, Heigold G. Word embeddings for speech recognition. Proc Interspeech 2014 2014:1053-1057. [doi: 10.21437/interspeech.2014-273]
27. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv 2013 [FREE Full text] [doi: 10.3126/jiee.v3i1.34327]
28. Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1162]
29. Li Y, Yang T. Word embedding for understanding natural language: a survey. In: Guide to Big Data Applications. Cham: Springer; 2018.
30. Beltagy I, Peters M, Cohan A. Longformer: the long-document transformer. arXiv 2020 [FREE Full text]
31. Mitchell T. Machine Learning. New York: McGraw-Hill; 2017.
32. Kim Y, Denton C, Hoang L, Rush A. Structured attention networks. arXiv 2017 [FREE Full text]
33. Palan S, Schitter C. Prolific.ac—A subject pool for online experiments. J Behavioral Experimental Finance 2018 Mar;17:22-27. [doi: 10.1016/j.jbef.2017.12.004]
34. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

35. Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 1993;28(1-2):76-81. [doi: 10.1159/000119004] [Medline: 8255414]
36. Gerra G, Zaimovic A, Zambelli U, Timpano M, Reali N, Bernasconi S, et al. Neuroendocrine responses to psychological stress in adolescents with anxiety disorder. Neuropsychobiology 2000;42(2):82-92. [doi: 10.1159/000026677] [Medline: 10940763]
37. Jezova D, Makatsori A, Duncko R, Moncek F, Jakubek M. High trait anxiety in healthy subjects is associated with low neuroendocrine activity during psychosocial stress. Prog Neuropsychopharmacol Biol Psychiatry 2004 Dec;28(8):1331-1336. [doi: 10.1016/j.pnpbp.2004.08.005] [Medline: 15588760]
38. Endler NS, Kocovski NL. State and trait anxiety revisited. J Anxiety Disord 2001;15(3):231-245. [doi: 10.1016/s0887-6185(01)00060-3] [Medline: 11442141]
39. Hashemipour S, Ali M. Amazon web services (AWS) – an overview of the on-demand cloud computing platform. In: Emerging Technologies in Computing. Cham: Springer International Publishing; 2020.
40. Beltagy I, Peters M, Cohan A. Longformer. Hugging Face. URL: https://huggingface.co/docs/transformers/v4.20.1/en/model_doc/longformer#longformer [accessed 2022-11-15]
41. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. arXiv 2017 [FREE Full text]
42. Pierse C. Transformers interpret. GitHub. URL: https://github.com/cdpierse/transformers-interpret [accessed 2022-11-15]
43. Di Matteo D, Fotinos K, Lokuge S, Yu J, Sternat T, Katzman MA, et al. The relationship between smartphone-recorded environmental audio and symptomatology of anxiety and depression: exploratory study. JMIR Form Res 2020 Aug 13;4(8):e18751 [FREE Full text] [doi: 10.2196/18751] [Medline: 32788153]
44. Di Matteo D, Fotinos K, Lokuge S, Mason G, Sternat T, Katzman MA, et al. Automated screening for social anxiety, generalized anxiety, and depression from objective smartphone-collected data: cross-sectional study. J Med Internet Res 2021 Aug 13;23(8):e28918 [FREE Full text] [doi: 10.2196/28918] [Medline: 34397386]
45. Siegman AW. The meaning of silent pauses in the initial interview. J Nerv Ment Dis 1978 Sep;166(9):642-654. [doi: 10.1097/00005053-197809000-00004] [Medline: 690624]
46. Mahl GF. Disturbances and silences in the patient's speech in psychotherapy. J Abnorm Psychol 1956 Jul;53(1):1-15. [doi: 10.1037/h0047552] [Medline: 13345560]
47. Ji S, Zhang T, Ansari L, Fu J, Tiwari P, Cambria E. MentalBERT: publicly available pretrained language models for mental healthcare. ArXiv 2021:10 [FREE Full text]

Abbreviations

AUROC: area under the receiver operating characteristic curve
BIS/BAS: behavioral inhibition/behavioral approach system
GAD: generalized anxiety disorder
GAD-7: Generalized Anxiety Disorder 7-item
IG: integrated gradient
LIWC: Linguistic Inquiry and Word Count
NLP: natural language processing
SAD: social anxiety disorder
STT: speech-to-text
SVM: support vector machine
TSST: Trier Social Stress Test
WER: word error rate

Edited by J Torous; submitted 15.11.22; peer-reviewed by K Gupta, Z Yang, T Zhang; comments to author 19.12.22; revised version received 21.02.23; accepted 23.02.23; published 28.03.23.

Please cite as:
Teferra BG, Rose J
Predicting Generalized Anxiety Disorder From Impromptu Speech Transcripts Using Context-Aware Transformer-Based Neural Networks: Model Evaluation Study
JMIR Ment Health 2023;10:e44325
URL: https://mental.jmir.org/2023/1/e44325
doi:10.2196/44325
PMID:36976636

Original Paper

Effectiveness and Feasibility of a Self-guided Mobile App Targeting Emotional Well-being in Healthy Adults: 4-Week Randomized Controlled Trial

Emily Brindal[1], PhD; Naomi Kakoschke[1], PhD; Sinead Golley[1], PhD; Megan Rebuli[1], BNutDiet (Hons); Danielle Baird[1], BNutDiet (Hons)

Commonwealth Scientific and Industrial Research Organisation, Adelaide, South Australia, Australia

Corresponding Author:
Emily Brindal, PhD
Commonwealth Scientific and Industrial Research Organisation
Gate 13 Kintore Avenue
Adelaide, South Australia, 5000
Australia
Phone: 61 08 8305 0633
Email: emily.brindal@csiro.au

Abstract

Background: Commercial smartphone apps designed to promote emotional well-being are becoming increasingly popular, but few apps have been empirically validated.

Objective: This study examined the feasibility and effectiveness of a self-guided app designed to reduce daily stress via positive messaging and tailored short inspirational talks (ie, peps).

Methods: A total of 166 participants (n=112, 67.5% female; mean age 38.48, SD 6.73 years) were recruited through social media advertising and randomized into an intervention (Hey Lemonade app plus twice daily mood monitoring using the Multidimensional Mood Questionnaire [MDMQ]) or active control (twice daily mood monitoring [MDMQ]) group. Primary (coping self-efficacy [CSE]; 3 subscales) and secondary outcomes (vitality, satisfaction with life, perceived stress, positive and negative affect, and hassles and uplifts) were measured at the baseline (week 1) and end point (week 4). The app evaluation questions were assessed at week 2. All interactions and measurements were collected on the internet and through the apps.

Results: In total, of 166 participants, 125 (75.3%) completed the trial. There were no differences in dropout rates between the groups (62/81, 76% intervention; 63/85, 74% control). There were significant group-by-time interactions for vitality and hassles but no significant effect for CSE total (P=.05). For the intervention group, the change from baseline to week 4 was significant for vitality (P=.002) and hassles (P=.004), CSE total (P=.008), and CSE Emotional subscale (P=.02). For the control group, any changes over 4 weeks were not significant for any outcome. There was a significant group-by-time interaction for MDMQ calmness (P=.04). By week 4, calmness was significantly higher in the intervention group (P=.046). Of those in the intervention group at week 2 (n=68), 39 (57%) participants recommended the app and 41 (60%) participants wanted to continue using it. Pep talks and customizable voice options were the most popular features.

Conclusions: Participants who had access to the smartphone app on an as-needed basis over the 4-week trial showed significant improvements in emotional well-being indicators. More broadly, this suggests that simple accessible solutions may generate meaningful well-being outcomes. Whether these changes are sustained and can be generalized to other population groups is yet to be determined.

Trial Registration: Australian and New Zealand Clinical Trials Registry (ANZCTR) 12622001005741; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=384304&isReview=true

(JMIR Ment Health 2023;10:e44925) doi:10.2196/44925

KEYWORDS

emotional well-being; self-guided; smartphone app; mobile health; mHealth; affect; mobile phone

Introduction

Background

The COVID-19 pandemic has been linked to measurable decreases in emotional well-being in Australians [1,2], an area of health that is gaining increasing focus worldwide. Although acknowledged as a multidimensional construct, currently, there is no clear consensus on the components of emotional well-being. Emotional well-being often includes mood or affective state, happiness, vitality, and satisfaction with life, with a distinction between psychological and subjective well-being. Subjective well-being captures positive affect and satisfaction with life [3], whereas psychological well-being [4] includes eudemonic aspects, such as purpose, growth, acceptance, and flourishing that align with the self-determination theory [5]. Other terms, such as mental well-being [6] and hedonic well-being, are also used interchangeably with psychological well-being. Although some well-being outcomes are used to capture the presence of mental illness, they are distinct. For current purposes, we use the term emotional well-being to reflect a multidimensional construct that includes a variety of emotional states.

Markers of emotional well-being have important clinical implications, given that they are associated with symptoms of anxiety and depression. There is growing evidence of the connection with other markers of health. For example, in a large sample of older adults in the United Kingdom (n=8780), self-realization (a concept tied to eudemonic well-being) and mood were positively associated with biological markers of inflammation in a model controlling for various confounders, including health conditions and mental health symptoms and conditions [7]. A recent meta-analysis of 17 papers suggested that emotional well-being influences changes in physical health in diseased populations [8]. Recent reviews have also indicated that positive mental well-being facilitates recovery from physical illness including cardiovascular disease [9,10].

Stress is conceptually related to well-being. Stress is a psychological response to a situation that exceeds a person's resources or the product of a person being unable to achieve their immediate goals [11,12]. Stress can be immediate, acute, or chronic. It is widely accepted that stress has a negative impact on both mental and physical health owing to hormonal changes [13] or oxidative stress [14] triggered by psychological stress, which is often suggested as a possible cause. Although stress is often researched in the context of workplaces, it has also been examined in terms of microstressors or daily hassles. Although these may not feel like significant events or intense chronic challenges such as those associated with occupational stress, daily hassles are also likely to be important predictors of health outcomes [15,16].

One theory that combines the concepts of affect, psychological well-being, and stress is the Conservation of Resource (CoR) theory [17,18]. This model suggests that people have certain resources that they can use to regulate their behavior. Resources include those that are psychological or emotional, physical (ie, tools and resources), and relational (ie, social support). For example, positive affect and higher resilience have been theorized to improve a person's ability to manage stress. This theory can also accommodate the experience of daily hassles, which, although seemingly small, also dip into a person's resources. Coping strategies and ability can also become a resource within this framework. Nevertheless, the CoR theory provides a functional descriptive framework with no obvious avenues for intervention.

App-delivered health programs have grown substantially with the popularization of smartphones [19] and have the potential to become physical resources and social support tools within the CoR framework. Apps have recently been developed in the emotional well-being space [20]. A review of apps targeting emotional well-being available in the Apple App Store and Google Play revealed a total of 231 available apps in 2020, which were largely self-guided in nature [21]. Another review [22], conducted in 2021, revealed 4 different mental health apps with associated published scientific articles. Most of these apps delivered a form of therapy that targets depressive symptoms. The authors concluded that more robust research is required in this area.

This Study

Emotional well-being is an important area, and apps have the potential to play a role in enhancing positive mental health through emotional well-being [23]; however, few apps targeting improved well-being have been empirically validated for their effectiveness. Thus, this study aims to examine both the feasibility and effectiveness of a self-guided smartphone app called Hey Lemonade, which was designed to help manage everyday stress and improve resilience and emotional well-being via short, pragmatic, and uplifting talks, which are referred to as peps.

Methods

Participants

We aimed to recruit approximately 150 adults aged 25 to 50 years, including 38 (25%) of whom were men. A comparison of 100 completers would allow the detection of a medium to large effect size (0.4) in coping self-efficacy (CSE; with a magnitude of a 0.84 difference based on a previous study [24]).

The selected age range and gender proportion were chosen, as they reflected the target audience of the Hey Lemonade smartphone app. Other inclusion criteria were as follows: owning a device with an operating system suitable for installing the Hey Lemonade app and currently residing in Australia. Exclusion criteria were self-reported and included having a professionally diagnosed mental health disorder, currently experiencing an abnormal level of life stressors (eg, loss of job or death of a loved one), being related to or close friends with the research team or developers of the Hey Lemonade app, and having fewer than 11 apps of any nature installed on their phone. This final criterion was chosen as an indicator of low app engagement and possibly app literacy [25]. People experiencing considerable life stressors or mental health diagnoses were excluded because the app was not designed to support major and chronic well-being issues.

Design

In this study, we used a 4-week randomized controlled trial design comparing the Hey Lemonade app (intervention plus daily mood monitoring) with an active control group (daily mood monitoring only). Participants were not blinded to the name of the intervention app and were informed that the study would evaluate the effectiveness of an app for managing everyday types of stressors. Those in the control group used only a freely available ecological momentary assessment app (SEMA3) to capture their daily mood data [26] (described further in Table 1). This was considered an active control because regular mood monitoring can itself have mental health benefits [27] and has been included as an aspect of most apps targeting emotional well-being [22].

Table 1. Secondary outcomes measured at baseline and end point.

Construct	Measure	Items and scale	Scoring	Cronbach α
Subjective well-being	Satisfaction with Life Scale [28]	• 5 items rated on a 7-point agreement scale	Scores comprised a summed raw score ranging from 5 to 35, with higher scores representing higher life satisfaction.	.88
Affect	Positive and Negative Affect Schedule [29]	• 20 items rated over the past month from 1 to 5 for the level of extent experienced. Half of the items measure PA[a], and the other half measure NA[b]	Scores are summed across respective items for PA and NA to represent the strength of affect. Scores can range from 10 to 50, with higher scores representing higher levels of each affect type.	PA: .89 and NA: .83
Stress	Perceived Stress Scale [30]	• 14 items rated over past month on a 5-point frequency scale	Positively worded questions were reverse scored. Scores for each item were summed to obtain a total score ranging from 0 to 56. Higher scores indicate higher perceived stress.	.84
Daily hassles and uplifts	Adapted[c] from hassles and uplifts scale [16]	• 40 items rated as hassles. The same items are then rated as uplifts using the following scores: not relevant (score=0), none (score=1), somewhat (score=2), quite a bit (score=3), a great deal (score=4)	Total hassles and uplifts were calculated at each time point through summing all responses.	Hassles: .90 and uplifts: .84
Vitality	Subjective vitality scale [31]	• 7 items rated on a 7-point agreement scale	Item number 2 was reverse scored (“I don’t feel very energetic”). All items were summed to calculate an overall score. Higher scores indicate higher vitality.	.89
Momentary mood	MDMQ[d] [32]	• 6 bipolar items (eg, “Tired–Awake”) rated “at this moment” using a slider scale from –5 (very) to +5 (very). Entered twice per day within certain time frames only (7-11 AM or PM).	Items assess 3 basic mood dimensions: energetic arousal (E), valence (V), and calmness (C). The MDMQ was delivered using the freely available SEMA3 app [32]. Participants were prompted to complete twice daily assessments. Data entry was only possible during prespecified times. Data from the 3 negative items were reverse coded to ensure higher scores indicated positive valance, energetic arousal, and calmness.	N/A[e]
App usability	N/A	• Open-ended questions asking favorite features of the app, • what they would change, and • how they felt after listening to peps. Recommendation to family and friends (yes; no; unsure) • How long do you wish to continue accessing the app? (Not at all; another week or 2; for 2 weeks to a month; for a few months; don’t want to stop) • Select favorite 3 features on a prespecified list of 12 core features	Open-ended responses were grouped thematically and then coded a second time to ensure they fit with the core themes.	N/A

[a]PA: positive affect.

[b]NA: negative affect.

[c]The original 53 items were piloted on a small convenience sample (n=17) and refined to a list of 40 options considered more suitable for a contemporary Australian audience. These details are available from the authors on request.

[d]MDMQ: Multidimensional Mood Questionnaire.

[e]N/A: not applicable.

Intervention: The Hey Lemonade App

The intervention app was developed by Hey Lemonade without input from the authors [33]. It was not publicly available at the time of the trial.

The app's core function is to provide 3- to 4-minute pep talks about daily stresses that can be customized through the choice of different voice options (referred to as companions in the app). The pep talks were written by professional writers and clinicians, including 2 psychologists, a clinical psychologist, and a solution-focused coaching specialist, who was also a provisional psychologist. The peps were grouped into the following 8 themes: All the Relationships, Daily Pep, Work Pressure, Pump-ups, Pepping Domesticity, Navigating the Feels, Looking After You, and Big Life Stress. Preliminary user experience and interface research led to the development of 30-40 uniquely written peps recorded using 7 voice options. The pep talks were voiced by public figures in the Australian community, representing a diverse mixture of characteristics. Each pep had at least 3 different voice options available.

Pep talks were developed using core principles, including positive affirmation, humor, universality through specificity, and encouraging breathing techniques, as well as using motivational language theory [34] and a solution-based coaching approach [35-37]. The app was designed to be light touch, as users could interact with it whenever they felt they needed additional motivation or inspiration. Therefore, there were no prescribed doses for this interaction. Pep talks could be selected based on theme, browsed by voice options, or users could search for peps on the required topics. Daily peps were offered, but not a mandatory feature to engage with.

Other key aspects of the app included the ability to gift relevant peps to people not using the app, daily pop-up inspirational written quotes delivered via an app notification, and a weekly email from the Hey Lemonade team. As the app was designed to have new content added monthly, new peps and a new voice option were added to the app at the start of week 2 to replicate this feature within the trial period. For further details about the intervention, refer to the CONSORT (Consolidated Standards of Reporting Trials)–EHEALTH checklist [38] provided in Multimedia Appendix 1.

Measures

Primary Outcomes

App Engagement

App engagement levels were measured throughout the trial in the intervention group. Data were collected by the app, deidentified, and supplied to us by Hey Lemonade. Overall use was calculated as a summed interaction score (ie, the number of total interactions). This score captures each time a participant engaged with the app in any form. Overall length of any interaction with the app (out of a possible 29 days) was calculated as the difference between the first and the last recorded interaction. If a person had any interaction on day 1 and another on day 28, the length of the interaction was 27 days. This outcome is used to represent an active membership period. Within the app, once a pep had played to completion, users could also rate how they felt: 1=a lot worse; 2=a little worse; 3=about the same; 4=a little better; and 5=a lot better.

Coping Self-efficacy

CSE was measured using the 26-item CSE Scale [24] at week 1 (baseline) and week 4 (end point). Participants were asked, "When things aren't going well for you, or when you're having problems, how confident or certain are you that you can do the following?" For each item, they were asked to rate the extent to which they believed they could perform behaviors important for adaptive coping on an 11-point scale. Anchor points included 0 (*cannot do at all*), 5 (*moderately certain can do*), and 10 (*certain can do*). Scores were summed for a total CSE score (Cronbach α=.95), as well as for specific domains. The subscales showed excellent consistency: CSE Problem-Focus (12 items; Cronbach α=.89), CSE Stop Unpleasant Emotions (9 items; Cronbach α=.89), and CSE Get Support (5 items; Cronbach α=.85).

Secondary Outcomes

Constructs captured as secondary outcomes are described in Table 1. All secondary outcomes were measured at baseline and end point, except for momentary mood, which was captured twice daily, and the app evaluation items measured in the intervention group at week 2.

Ethics Approval

The study was approved by the Commonwealth Scientific and Industrial Research Organisation Human Research Ethics Committee (application 2022_037_LR) and registered at the Australian and New Zealand Clinical Trials Registry (12622001005741).

Procedure

Participants were recruited through paid advertisements of Commonwealth Scientific and Industrial Research Organisation on Facebook, targeting people aged 25 to 50 years in October 2022. The Hey Lemonade company directors also promoted the trial on their social media pages (ie, Facebook, Instagram, and Twitter).

Interested participants were directed to a web-based participant information sheet and consent form, which included screening questions to assess eligibility. Those who passed the screening questionnaire were asked to contact the trial manager to express interest in the study. Eligible participants were contacted and asked to provide verbal consent to participate in the study and demographic information. The participants were then emailed the web-based baseline questionnaire. Participants who completed this step were enrolled in the study, given a participant identifier, randomized via a computer-generated sequence on a 1:1 basis ensuring balanced distribution of gender, and provided with instructions about how to install the SEMA3 and Hey Lemonade (intervention-only) apps. The intervention app was installed from the App Store via TestFlight or Google Play using beta testing links because it was not publicly available at the time of the trial. The trial manager randomized

the participants, and the research team was blinded to the participant allocation and did not have contact with the participants throughout the study.

Trial outcomes were assessed at baseline, week 2, and end point (week 4). The study was conducted without face-to-face contact. All assessments were completed on the web through the web-based survey platform Alchemer via a link sent to the participants at each assessment point. Participants were sent 1 email reminder if they had not completed the survey within 24 to 48 hours. If they failed to complete the survey after this reminder, they were marked as lost to follow-up. Participants were also able to formally withdraw at any point after the baseline. These participants were asked a reason for their withdrawal. Those who completed the final survey were sent an Aus $50 (US $33.50) electronic gift voucher to thank them for their participation. No costs were associated with accessing any of the apps used in this study.

Technical Issues

For the Hey Lemonade app, users reported issues accessing the content on 6 occasions. These issues were usually resolved within 24 hours. For the SEMA3 app, approximately 30 participants reported technical difficulties that affected a higher number of Android users. After 3 to 4 days, these issues could not be resolved, so the trial team created a replacement recurring web-based survey to assess momentary mood using the Alchemer platform, in which participants were asked to complete without reminders.

Data Analysis

To compare the effect of group on change over time for all trial outcome measures, we conducted linear mixed models fitting a marginal model (in which random participant effects are not considered) with unstructured covariance matrices to optimize efficiency [39,40]. Fixed effects included the main effect for group, time, and the interaction effect for group by time, as well as sex. All models were controlled for age as a covariate. The outcomes were analyzed at the intention-to-treat level. Specifically, we used a data augmentation method to analyze the full, incomplete data set, namely the restricted maximum likelihood estimation method [41]. This method does not involve imputing any data but instead uses each participant's available data to derive maximum likelihood estimates, which refers to the value of the parameter that has the highest likelihood of producing the observed data [42]. The models included all available data from each time point measurement of the 166 participants who commenced the study.

The daily mood data comprised a total of 5305 observations with an average of 31.77 (SD 17.48; range 1-59) observations per participant. There was no significant group difference in the number of observations (intervention: mean 32.92, SD 16.88 vs control: mean 30.64, SD 18.07; t_{165}=0.851; P=.40). Owing to the technical issues with the SEMA3 app, the number of missing observations in the first week of the trial was high (570/2338, 24.38%), although this was comparable with previous work [32]. Observations for each participant were aggregated by trial week, and within-participant averages from weeks 1 and 4 were compared in subsequent analyses [43]. An unstructured, linear mixed effects model with main effects of group, time, and the interaction effect for group by time was conducted.

SPSS (version 26.0; IBM Corp) for Microsoft Windows was used to perform all analyses, and statistical tests were 2 tailed, with P<.05 considered statistically significant. Pairwise comparisons were conducted for all the significant interaction effects.

Results

Sample Characteristics

A total of 1066 people accessed consent and screening information, with 572 (53.66%) meeting the eligibility criteria. After being contacted by the trial manager, 30.6% (175/572) of individuals consented to participate in the trial. Of the participants who consented 94.9% (166/175) completed the baseline questionnaire and were randomized into either the intervention (81/166, 48.8%) or control (85/166, 51.2%) condition (Figure 1; Table 2). There were no significant group differences in any sample characteristics.

Figure 1. Participant flow diagram. LTFU: lost to follow-up; WD: withdrawn.

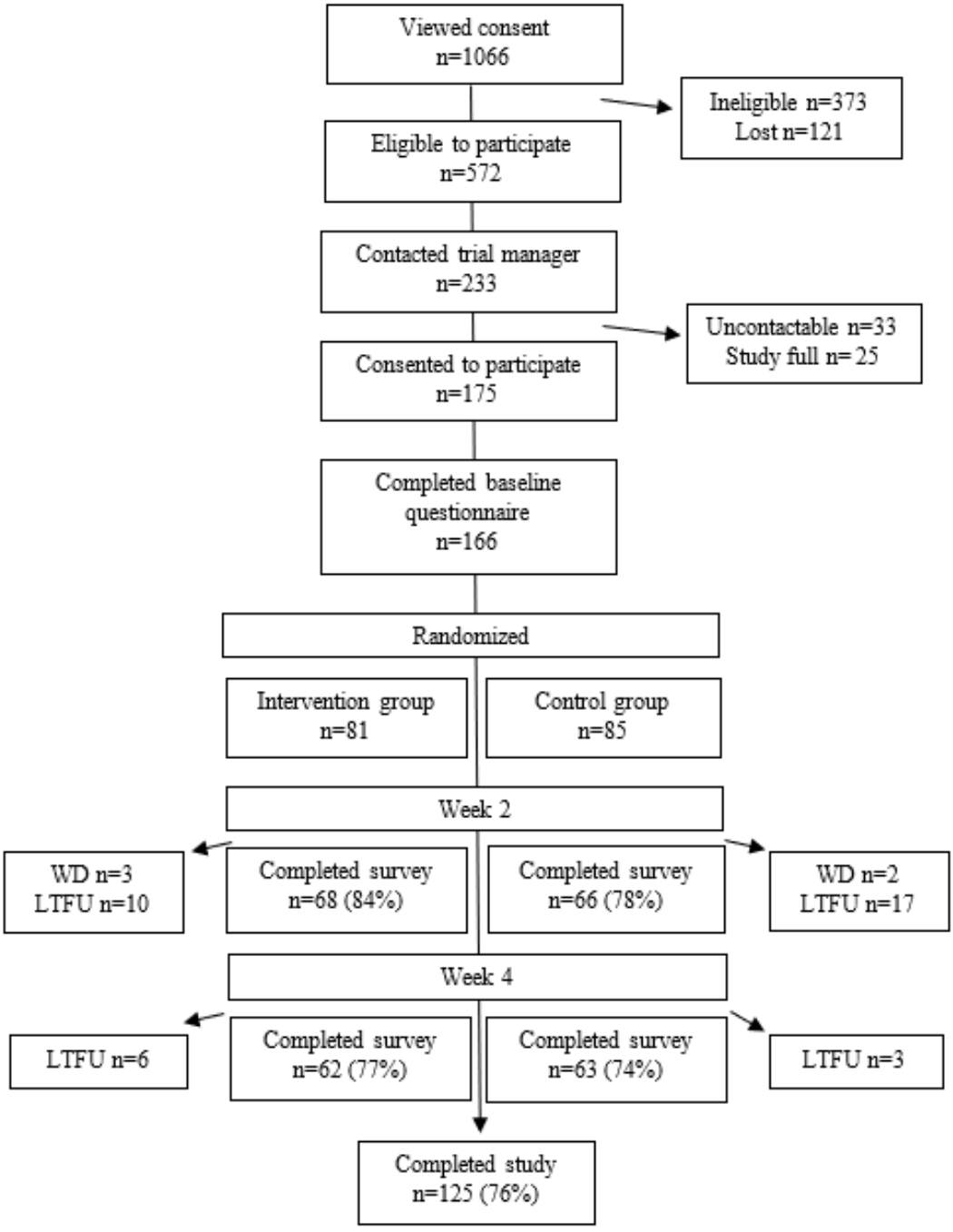


Table 2. Sample characteristics[a] (N=166).

	Intervention (n=81)	Control (n=85)
Female, n (%)	55 (68)	58 (68)
Age (years), mean (SD)	37.63 (6.92)	39.28 (6.48)
Education, n (%)		
High school	5 (6)	5 (6)
TAFE[b], trade or certificate	4 (5)	6 (7)
Diploma	3 (4)	3 (4)
University degree	35 (43)	29 (34)
Postgraduate study	34 (42)	42 (49)
System or device, n (%)		
Apple	56 (69)	54 (64)
Android	25 (31)	31 (37)
Number of wellness apps installed on phone, n (%)		
None	14 (17)	18 (21)
1-3	42 (52)	38 (45)
4-9	19 (24)	34 (40)
10-15	4 (5)	2 (2)
>15	2 (3)	3 (4)

[a]No significant group differences for any sample characteristics.

[b]TAFE: technical and further education.

A total of 41 participants were lost to follow-up or withdrew from the trial following the completion of the baseline questionnaire and group randomization. There were no significant differences between completers and noncompleters in any of the measures, including demographic variables or baseline primary and secondary outcomes.

There were significant, moderate positive correlations between the primary outcome measure, total CSE, with positive affect (r=0.51; P<.001), as well as life satisfaction (r=0.47; P<.001) and vitality (r=0.59; P<.001). Moderate negative relationships were found between CSE and negative affect (r=−0.51; P<.001), hassles (r=−0.44; P<.001), and perceived stress (r=−0.61; P<.001). Uplifts were weakly to moderately positively associated with CSE (r=0.28, P<.001).

Changes in Primary Outcomes

App Engagement

The number of users accessing the app each day fell the most dramatically during the first week and continued to taper off for the remainder of the trial. When interactions were divided by the number of active users per day, the intensity of interaction per user remained stable from day 5 (Table 3). Most participants in the intervention condition accessed the app on the first day (68/81, 84%) or the second day (10/81, 12%) it was available. Overall, app membership (the mean number of days between the first and the last interaction) was 19.15 (SD 8.40; range 0-29) days. One participant did not access the app until 18 days after it became available. A total of 3 users who failed to complete the study outcomes at week 2 had memberships from 19 to 29 days.

Table 3. Total interactions, number of users, and interactions divided by number of users over the 4-week study period.

Date	Total interactions, n	Number of users, n	Interactions/users
August 18, 2022	3401	69	49.29
August 19, 2022	2351	68	34.57
August 20, 2022	969	52	18.63
August 21, 2022	1469	46	31.93
August 22, 2022	941	47	20.02
August 23, 2022	710	36	19.72
August 24, 2022	651	33	19.73
August 25, 2022	377	31	12.16
August 26, 2022	211	23	9.17
August 27, 2022	310	21	14.76
August 28, 2022	375	26	14.42
August 29, 2022	335	27	12.41
August 30, 2022	293	20	14.65
August 31, 2022	248	18	13.78
September 1, 2022	314	21	14.95
September 2, 2022	215	16	13.44
September 3, 2022	129	11	11.73
September 4, 2022	374	16	23.38
September 5, 2022	302	21	14.38
September 6, 2022	271	20	13.55
September 7, 2022	238	12	19.83
September 8, 2022	184	10	18.4
September 9, 2022	178	12	14.83
September 10, 2022	113	7	16.14
September 11, 2022	94	7	13.43
September 12, 2022	105	9	11.67
September 13, 2022	167	7	23.86
September 14, 2022	188	10	18.8
September 15, 2022	323	12	26.92
September 16, 2022	165	15	11

The mean intensity of app use over the intervention period (summed interaction score) for those who completed the intervention (n=62) was 213 (range 9-936). One-third of the users who completed the study had between 100 and 200 interactions with the app, whereas 1 *superuser* had 936 interactions.

In the multiple regression analysis, participant sex, age, number of wellness apps, and device type failed to predict higher total engagement with the app over the trial, explaining 2.7% of the total variance ($F_{4,76}$=1.55; P=.20).

The frequency of pep selection per participant was used as the primary engagement metric. There was a total of 36 different pep talks available for selection, which were played on 914 individual occasions. Of the 81 participants randomized to the intervention app, 80 accessed at least 1 pep. Users accessed peps between 0 and 74 times each throughout the trial, with roughly one-third accessing a pep 1 to 4 times, 5 to 11 times, or 12 to 48 times, and a single user accessing a pep 74 times. On average, all users in the study listened to a pep 11 (mean 11.65, SD 12.69) times per month. Of all peps accessed, 66.2% (605/914) were marked as finished. A total of 14% (11/81) of users failed to play a pep to marked completion at any point throughout the trial.

Once a pep had been played to completion, participants were offered the opportunity to indicate the extent to which listening to the pep had impacted how they were feeling, a feature included for the trial duration only. Of all the peps marked to completion, 81.5% (493/605) received feelings ratings. On the basis of individual pep topics, more than half of the (20/36,

56%) available topics were rated ≥4 on average ("a little better" to "a lot better"), with the mean of all ratings being 4.01 (SD 0.71). On average, each pep was rated ≥3 ("about the same"). The 2 most popular peps were accessed 133 and 128 times throughout the trial, with the third to fifth most popular accessed 47 to 41 times. A total of 9 peps were accessed less than 10 times throughout the trial.

Coping Self-efficacy

The raw means are presented in Table 4, and the linear mixed model results are presented in Table 5. There was a borderline effect for CSE total (P=.05). Exploratory analyses were conducted, and models were created using the 3 subscales of CSE to understand if these changes occurred in specific domains. The results revealed a significant interaction effect only for the CSE Emotional subscale (P=.04). The only significant pairwise comparison was for the intervention group, with a significant improvement in the CSE Emotional subscale from baseline to end point (P=.02; Figure 2).

Table 4. Raw means and SDs for outcome measures at baseline and end point presented by group.

	Baseline, mean (SD)		End point, mean (SD)	
	Intervention (n=81)	Control (n=85)	Intervention (n=62)	Control (n=63)
CSE[a] total	150.11 (37.53)	152.56 (38.08)	162.92 (37.65)	154.21 (41.60)
CSE—problem-focused	73.35 (16.03)	74.87 (17.97)	79.39 (16.03)	75.94 (18.00)
CSE—stop unpleasant emotions	47.44 (15.20)	49.96 (14.32)	52.21 (15.51)	49.84 (16.57)
CSE—get support	29.32 (10.27)	27.73 (9.60)	31.32 (9.92)	28.43 (10.34)
Vitality	27.40 (7.49)	27.80 (7.63)	30.21 (7.41)	27.76 (9.03)
Positive affect	31.49 (7.24)	31.56 (7.20)	32.73 (6.83)	31.22 (7.64)
Negative affect	18.64 (5.65)	19.09 (5.71)	19.44 (6.20)	20.67 (6.98)
Satisfaction with life	24.36 (5.86)	24.02 (5.93)	25.98 (6.02)	24.10 (7.44)
Perceived stress	23.77 (6.50)	24.19 (7.01)	24.35 (5.95)	25.32 (7.99)
Hassles	61.57 (14.18)	62.08 (15.11)	56.82 (11.98)	62.14 (14.90)
Uplifts	67.85 (16.42)	67.72 (14.40)	68.71 (15.05)	65.56 (15.70)
Momentary mood (MDMQ[b,c])				
MDMQ—calmness domain	1.37 (1.55)	1.26 (1.32)	1.60 (1.59)	1.23 (1.43)
MDMQ—valence domain	1.68 (1.64)	1.75 (1.39)	1.81 (1.80)	1.80 (1.52)
MDMQ—energetic arousal domain	−0.01 (1.61)	−0.22 (1.34)	0.13 (1.65)	0.00 (1.57)

[a]CSE: coping self-efficacy.

[b]MDMQ: Multidimensional Mood Questionnaire.

[c]On the basis of 1768 observations at baseline and 1130 observations at end point.

Table 5. Linear mixed model estimates for primary and secondary trial outcomes.

Model estimates	Fixed effects, B (SE)					
	Intercept	Time (baseline)	Group (intervention)	Sex (female)	Age	Group × time
Primary outcomes						
CSE[a] total	120.69 (16.91[b])	0.27 (4.31)	10.79 (6.95)	4.16 (5.99[c])	0.73 (0.42)	−12.02 (6.12[d])
CSE—problem-focused	65.45 (7.39[b])	−0.63 (2.02)	4.06 (2.99)	1.51 (2.62)	0.23 (0.18)	−5.20 (2.87)
CSE—stop unpleasant emotions	36.29 (6.65[b])	1.03 (1.75)	3.23 (2.81)	0.34 (2.335)	0.32 (0.16)	−5.23 (2.49[c])
CSE—get support	20.01 (4.42[b])	−0.20 (1.07)	3.43 (1.77)	2.27 (1.57)	0.16 (0.11)	−1.56 (1.52)
Secondary outcomes						
Vitality	20.11 (3.44[b])	−0.07 (0.74)	2.20 (1.35)	0.09 (1.22)	0.20 (0.09[c])	−2.28 (1.05[c])
Positive affect	20.93 (3.06[b])	0.07 (0.82)	1.62 (1.22)	−1.46 (1.09)	0.29 (0.08[b])	−1.21 (1.16)
Negative affect	25.80 (2.65[b])	−1.69 (0.68[c])	−1.15 (1.16)	−0.52 (0.93)	−0.12 (0.07)	0.50 (0.97)
Satisfaction with life	19.53 (2.81[b])	−0.09 (0.55)	1.33 (1.16)	0.69 (0.99)	0.10 (0.07)	−0.82 (0.78)
Perceived stress	30.06 (3.05[b])	−0.96 (0.70)	−0.92 (1.19)	−0.92 (1.19)	−0.12 (0.08)	0.30 (0.99)
Hassles	67.33 (6.34[b])	−0.66 (1.51)	−5.84 (2.38[c])	0.20 (2.26)	−0.12 (0.16)	5.12 (2.15[c])
Uplifts	51.15 (6.47[b])	2.03 (1.93)	3.39 (2.65)	0.55 (2.29)	0.36 (0.16[c])	−2.66 (2.75)
Momentary mood (MDMQ[e,f])						
MDMQ—calmness domain	1.06 (0.19[b])	0.20 (0.16)	0.54 (0.27[c])	N/A[g]	N/A	−0.05 (0.22[c])
MDMQ—valence domain	1.64 (0.21[b])	0.12 (0.16)	0.16 (0.29)	N/A	N/A	−0.26 (0.23)
MDMQ—energetic arousal domain	−0.07 (0.22)	−0.14 (0.20)	0.21 (0.30)	N/A	N/A	−0.02 (0.27)

[a]CSE: coping self-efficacy.

[b]$P<.01$.

[c]$P<.05$.

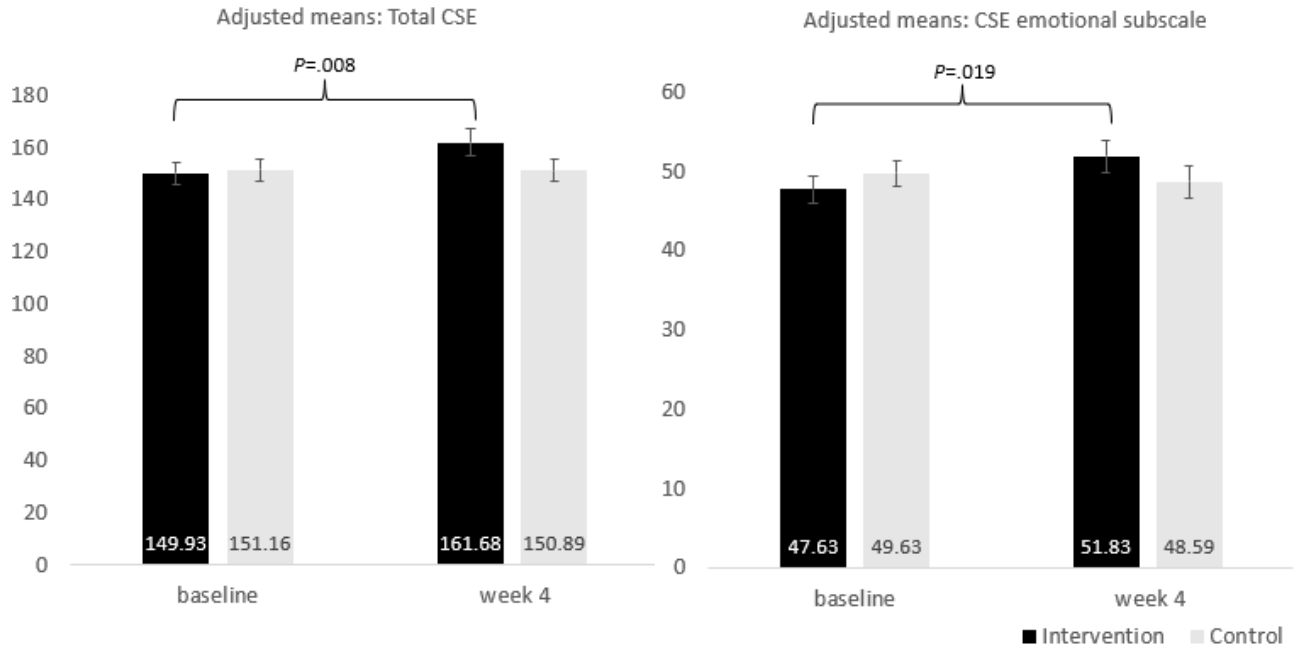
[d]$P=.05$ (approaching significance).

[e]MDMQ: Multidimensional Mood Questionnaire.

[f]On the basis of 1768 observations at baseline and 1130 observations at the end point.

[g]N/A: not applicable.

Figure 2. Adjusted means for interaction between group and time for coping self-efficacy (CSE).

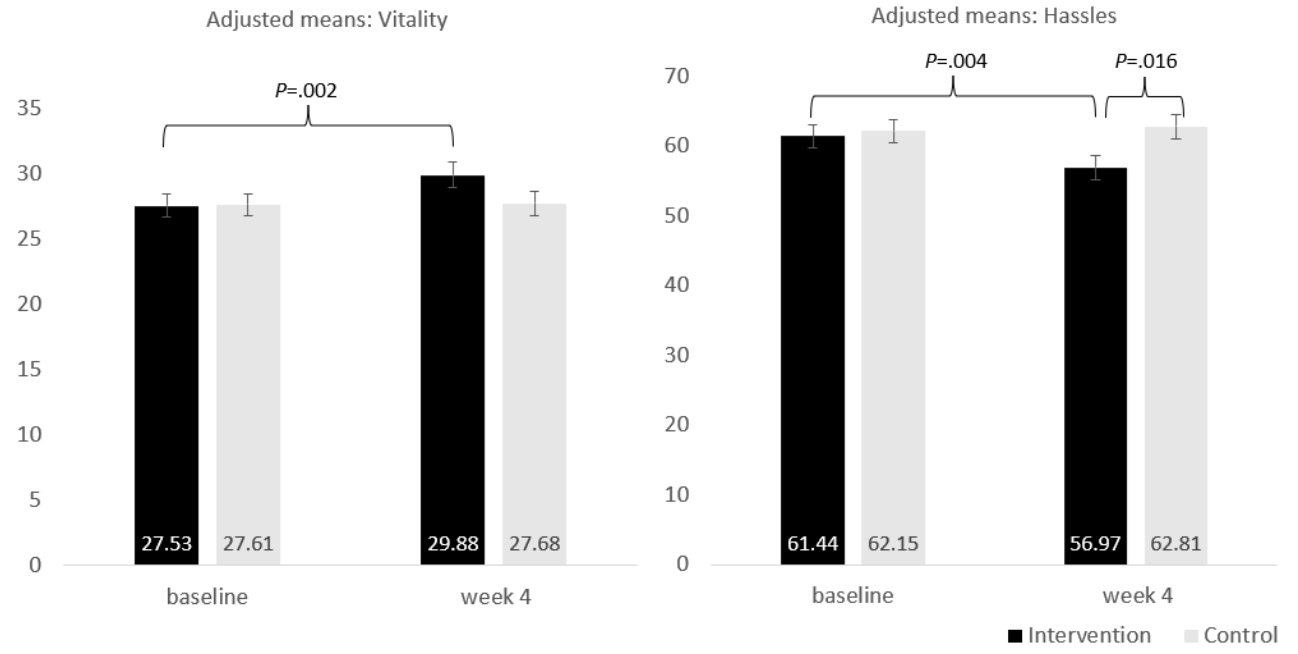


Changes in Secondary Outcomes

Vitality, Hassles or Uplifts, Affect, Perceived Stress, and Satisfaction With Life

There were significant group-by-time interaction effects for vitality (*P*=.03) and hassles (*P*=.02; Table 4). There were no significant effects for positive and negative affect, perceived stress, uplifts, and satisfaction with life. Pairwise comparisons suggested that at week 4, there was a significant difference between groups for hassles (*P*=.02), with the intervention group reporting lower hassles. The magnitude of the change in the intervention group was significant for vitality (*P*=.002) and hassles (*P*=.004), with the intervention group showing positive improvements in both instances (Figure 3).

Figure 3. Adjusted means for interaction between group and time for vitality and hassles.

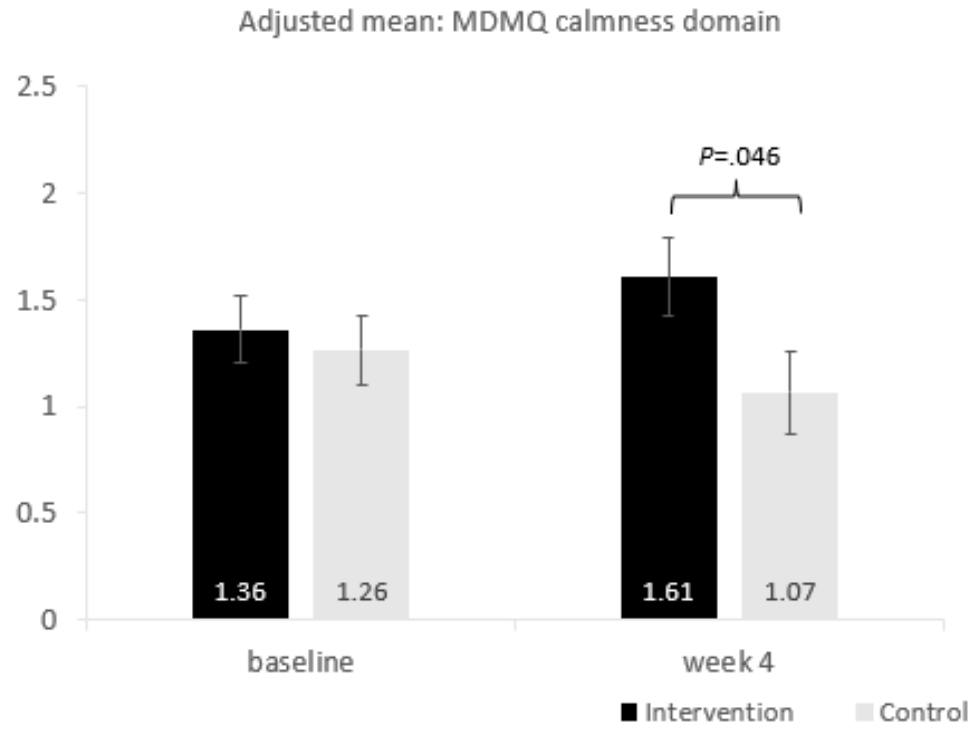


Momentary Mood

There was a significant interaction between group and time for the momentary mood outcome of the Multidimensional Mood Questionnaire in the calmness domain (*P*=.04). Pairwise comparisons suggested that at baseline, there was no significant group difference in calmness scores (*P*=.67), whereas at week 4, calmness was significantly higher in the intervention group than in the control group (*P*=.046; Figure 4). Pairwise comparisons suggested no significant changes from baseline to

week 4 in the intervention (P=.10) or control groups (P=.21). There were no significant effects for valence or energetic arousal (all P values>.10).

Figure 4. Interaction between group and time for the Multidimensional Mood Questionnaire (MDMQ) calmness domain.



App Usability

Of the 68 participants in the intervention group at week 2, a total of 39 (57%) indicated that they would recommend the app to family and friends, with a further 16 (24%) who were unsure. A total of 60% (41/68) indicated that they wanted to keep using the app for a few months or indefinitely, with 16% (11/68) not wanting to access the app any further. Pep talks and customizable voice options were selected as the favorite features of the app for most users (Table 6).

Table 6. Top features selected by users of the intervention app at week 2 (n=68).

Feature selected[a]	Values, n (%)
The Pep talks	53 (78)
Being able to select different voice companions	43 (63)
The variety of pep talks available	36 (53)
The look and visuals of the app	32 (47)
The tone and style of voices	30 (44)
The material covered by the pep talk	30 (44)
The variety of voice options available	28 (41)
The daily notifications or inspirational quotes	26 (38)
The layout of the menu	19 (28)
The ability to gift a pep	14 (21)
The Peppervescent points	5 (7)
The weekly emails	4 (6)

[a]Each user could select up to 3 options.

In terms of qualitative feedback regarding how they felt after listening to a pep, most participants (49/68, 72%) referenced positive energy with a small number of detractors (5/68, 7%; Table 7).

Table 7. Summary of qualitative feedback grouped by key themes with example quotes.

Theme	Quote 1	Quote 2	Quote 3
Positive energy			
Calmer or relaxed (n=17)	"I felt a bit more calm and assured and ready to move forward with my day with more confidence"	"Inspired, calm, armed with practical advice to get through tough times"	"Relaxed and supported."
Clearer or uplifted (n=13)	"I felt a 'pep' in my step after listening to the talks. I really enjoyed being able to have a quick pick me up from a familiar voice."	"Clear minded. Refreshed"	"Uplifted and a more positive outlook"
Generally better or supported (n=13)	"Most often I feel better after listening to a pep talk. I find I use them most when I get stuck in a thought or something happens that derails my day."	"Help me change the way I think about the issue I'm having."	"A little better"
Motivated (n=6)	"Motivated and more positive"	"After some peps, I felt a little better and more motivated"	"Some of them made me feel better, more motivated or happier about the day"
Null or negative			
Neutral (n=14)	"Sometimes I felt a little bit better other times I felt in different"	"Fine but it didn't tell me anything new"	"About the same, really. It was nice, though."
Worse (n=5)	"I couldn't get through the whole thing"	"Patronized. It felt like the app equivalent of someone saying "you'll be fine, you're great" when you disclose that you struggle with your self image. Well intended, but misses the mark"	"This isn't for me"

Discussion

Principal Findings

In this study, we aimed to evaluate the effectiveness and feasibility of a self-guided app targeting emotional well-being relative to active control of regular mood monitoring. The primary feature of the intervention app was to provide pep talks delivered by familiar and customizable voices to boost people's mood and coping resources. After 4 weeks, there was no significant change in the total CSE. Exploratory analyses suggested a significant improvement in one component, namely, the perceived ability to stop unpleasant thoughts and emotions (CSE Emotional subscale). The total app use was reduced over the 4 weeks and was not predicted by any participant characteristics. Pep talks, which are the central features of the app, were favorably rated by users. Those who entered the data immediately after listening to a pep largely indicated feeling positive changes in their momentary mood state. Of the 3 aspects of momentary mood measured daily, calmness improved significantly for the app users. Finally, there were substantial improvements in perceived daily hassles and vitality for those randomized to the app condition. Overall, the significant changes observed were aligned with qualitative feedback, which highlighted the impacts of the app on feeling calmer and more positive.

There is very little published evidence in the space of well-being apps, despite the hundreds of apps available for download [21]. Most of the 48 well-being apps reviewed in 2021 assessed mental health outcomes (n=19) and included clinical approaches suitable for these targeted outcomes [22]. Medium effect sizes were reported among the 6 that assessed emotional regulation. Almost all the studies evaluating emotional regulation [22] included full mindfulness interventions, which were largely self-guided intensive programs, including audio sessions ranging between 10 and 30 minutes. The self-guided app that we evaluated was designed to fill a unique space in targeting daily emotional well-being and to be much lighter, which was guided by counseling principles rather than formal, structured programs. Instead of delivering formal therapy, it is more likely to provide a buffer to the constant drain on resources that may lead to larger changes in well-being or increases in stress.

More recently, the Positive Activity Model has supported the idea that simple tools can improve well-being [44]. The current findings support the idea that less intensive intervention apps may be effective. CSE was selected as the primary outcome for this study, given its importance for a range of emotional well-being outcomes. Specifically, it has been flagged as an important mediator of how people manage traumatic events with the potential to mitigate the experience of possible distress [45]. CSE has also been flagged as a relevant factor for maintaining health behavior in the Health Action Process Approach model [46], consuming healthier diets [47], and in elite athlete's performance [48]. In this trial, the findings for CSE were close to being statistically significant overall, with the subscale measuring the ability to control unpleasant thoughts and emotions showing significant improvement over time in the intervention group. Bandura [49] noted that the ability or inability to control and redirect thoughts is a major aspect of anxiety and that the ability to turn them off is critical for good well-being. As a key feature of the app, it appears that pep talks could assist in developing this critical skill.

App usability and user feedback data were promising. More than 75% (53/68) of participants completing evaluation in week 2 reported wanting to continue using the app beyond the trial. Before commencing, participants were blinded to the content of the app and were told only that the study evaluated its effectiveness on daily stressors. A small number of people did not enjoy the app (ie, 5-11, depending on the metric used). Nevertheless, this sample was not specifically interested in wellness apps, as evidenced by the low number of existing wellness apps installed on the participants' devices. In combination with partial blinding, it was expected that the app would not be to everyone's taste. However, the overall qualitative feedback was also mostly positive, reflecting positive changes in calmness while also focusing on feeling positive or having an improved outlook. This appears to have translated to positive changes in momentary calmness and vitality but not overall satisfaction with life.

The hassle conceptualization of stress has been called "a minor events approach" [50] and aligns with the targeted outcome of using the app. Effects were present for this domain but not for general perceived stress. Early conceptions of well-being suggested that hassles were a large part of overall well-being [51] and that hassles were better at predicting well-being, mental health, and health status than life events [50,52]. Using a Dynamic Structural Equation Modeling approach on 14-day data from students, Tran et al [53] reported that physical health complaints were predicted by the experience of hassles the day prior. Overall, the ability to significantly reduce hassles could have health benefits and improve overall well-being. Our measure did not capture the centrality of hassles nor ongoing themes and issues [54]. Future studies could aim to improve measurement to determine if changes occurred for substantial hassles or in sheer volume, which would change the interpretation of possible outcomes.

Hassles may also be reduced through increased vitality, given that positive energy has been associated with fewer negative appraisals of personal problems [55]. Vitality is an important outcome, as it includes feelings of aliveness and high positive energy is associated with motivation [55], which reinforces its potential as a buffer against resource depletion. In recent years, there has been a documented reduction in vitality because of coronavirus-related anxieties [56], meaning that improvements in vitality may be even more timely.

The increase in vitality in the intervention group represented a 7.8% improvement on average. A previous study observed that a 5% to 10% decrease in vitality (based on the 36-Item Short Form vitality subscale) was associated with increased disease risk for conditions ranging from depression to angina and osteoarthritis [57]. Thus, the magnitude of the observed change for the intervention group could have meaningful impacts on future health if these effects can be maintained over the 4-week trial duration.

In the context of a broader theoretical model, it appears that a simple tool can have benefits for some aspects of coping. However, the CoR theory needs to consider not only equipping people with tools but also the confidence to use them [58]. Using a pain paradigm, researchers have shown that greater perceptions of control are most beneficial in the presence of confidence [59]. Self-worth is likely to be an important variable. Future interventions positioning apps as resources for mental health and well-being should account for these variables, as they are likely to have significant mediation or moderation effects on outcomes.

Strengths and Limitations

This study was a controlled, rigorously designed trial, powered appropriately to detect meaningful effects. Previous studies have used anywhere between 12 days and 12 weeks to evaluate similar apps with those of longer durations targeting clinical samples [21,22]. The 4-week period proved enough to detect changes in some well-being indicators but did not provide an indication of how sustained these changes are over time. This is particularly true for outcomes, such as vitality and hassles, which may be more transient than constructs, such as self-efficacy.

Our primary research questions focused on the effectiveness and feasibility of the app rather than mechanisms for change; therefore, we can only use the literature to guide the interpretation of which specific features of the app were associated with the changes observed. Coaching frameworks, humor [60-62], and positive charismatic people [63] may be beneficial. A total of 2 peps were listened to more than the others, and some of the voice options were also more popular. Appropriate variety in topic and delivery could be an important part of success; however, it is unclear where this tipping point lies.

The current sample was targeted toward individuals in the market for whom the app was developed and designed in a process that occurred before this effectiveness trial. The target market included a narrow age range, comprising mostly women with no major mental health challenges. Approximately one-third of the interested participants were deemed ineligible. It may be possible that healthier samples and women have stronger CSE overall, which makes it easier to build upon [64]. However, its effectiveness in other groups remains unknown. The sample also included a higher proportion of people with university degrees or greater in those aged between 25 and 50 years compared with the general population in Australia, of which 38.7% had obtained this level of education [65]. It is possible that recruitment through our institution's social media pages attracted more highly educated samples because of its reputation as a national science organization in Australia. However, analysis of other data in Australia suggests that well-being indicators may be harder to shift in samples with greater income or education levels because of higher expectations regarding life circumstances [66].

Not all feedback received regarding the Hey Lemonade app was positive. There was a notable proportion of unsure selections (16/68, 24%) regarding the recommendation of the app to others. Qualitative feedback regarding the ability to gift peps to people provided some insights into possible hesitation to recommend with users, suggesting that they needed to feel comfortable enough to forward on "self-help–type" materials. The nature of a mental well-being focus may create challenges for recommending and sharing components.

One of the largest challenges in this study was the technical issues associated with daily mood monitoring using the freely available SEMA3 app. These may have disproportionately affected the control group, for whom mood monitoring was the primary task associated with trial participation, as well as Android users who appeared to report more technical issues. Unfortunately, using a third-party app, such as SEMA3, meant that these issues were outside the control of the research team. It is reassuring that no negative changes were observed in the control group; however, it is unclear if this may have disrupted any possible positive change generated by mood monitoring for those in the control group.

Conclusions and Implications

The Hey Lemonade app is designed to be a simple, no-fail tool to assist people in managing daily stress. The intervention did not result in a significant change in the primary outcome of CSE, but it did improve momentary calmness and secondary outcomes, including positive feelings and the ability to cope with general life hassles over 4 weeks. This indicates that a digital intervention such as the Hey Lemonade app might be useful for bolstering people's resource kitty and hence make them more equipped to face daily challenges. More broadly, the findings also suggest that simple solutions may have the ability to generate meaningful outcomes for well-being [44]. These findings were witnessed in a targeted sample, excluding those inexperienced with apps or experiencing substantial mental health or life challenges. Nevertheless, the Hey Lemonade app represents a promising self-guided approach for managing daily life stress and promoting positive well-being states.

Acknowledgments

This project was cofunded by Hey Lemonade and the Commonwealth Scientific and Industrial Research Organisation KickStart program. Hey Lemonade reviewed and commented on the study design and the final manuscript.

Authors' Contributions

All authors were involved in the trial. EB designed the study and prepared the manuscript. SG and NK contributed to the design of the study. EB, SG, NK, and DB performed data analyses. MR managed the trial. All authors contributed to and reviewed the final manuscript.

Conflicts of Interest

This project was performed as part of the normal duties in relation to authors' employment at Commonwealth Scientific and Industrial Research Organisation.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials)-EHEALTH checklist (version 1.6.1).
[PDF File (Adobe PDF File), 12849 KB - mental_v10i1e44925_app1.pdf]

References

1. Brindal E, Ryan JC, Kakoschke N, Golley S, Zajac IT, Wiggins B. Individual differences and changes in lifestyle behaviours predict decreased subjective well-being during COVID-19 restrictions in an Australian sample. J Public Health (Oxf) 2022 Jun 27;44(2):450-456 [FREE Full text] [doi: 10.1093/pubmed/fdab040] [Medline: 33683320]
2. Stewart-Brown S. Emotional wellbeing and its relation to health. Physical disease may well result from emotional distress. BMJ 1998 Dec 12;317(7173):1608-1609 [FREE Full text] [doi: 10.1136/bmj.317.7173.1608] [Medline: 9848897]
3. Diener E. Subjective well-being. Psychol Bull 1984 May;95(3):542-575. [Medline: 6399758]
4. Ryff CD. Psychological well-being revisited: advances in the science and practice of eudaimonia. Psychother Psychosom 2014;83(1):10-28 [FREE Full text] [doi: 10.1159/000353263] [Medline: 24281296]
5. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
6. Smith L, Jacob L, Yakkundi A, McDermott D, Armstrong NC, Barnett Y, et al. Correlates of symptoms of anxiety and depression and mental wellbeing associated with COVID-19: a cross-sectional study of UK-based respondents. Psychiatry Res 2020 Sep;291:113138 [FREE Full text] [doi: 10.1016/j.psychres.2020.113138] [Medline: 32562931]
7. Fancourt D, Steptoe A. The longitudinal relationship between changes in wellbeing and inflammatory markers: are associations independent of depression? Brain Behav Immun 2020 Jan;83:146-152 [FREE Full text] [doi: 10.1016/j.bbi.2019.10.004] [Medline: 31604140]
8. Lamers SM, Bolier L, Westerhof GJ, Smit F, Bohlmeijer ET. The impact of emotional well-being on long-term recovery and survival in physical illness: a meta-analysis. J Behav Med 2012 Oct;35(5):538-547 [FREE Full text] [doi: 10.1007/s10865-011-9379-8] [Medline: 21918889]

9. Diener E, Pressman SD, Hunter J, Delgadillo-Chase D. If, why, and when subjective well-being influences health, and future needed research. Appl Psychol Health Well Being 2017 Jul;9(2):133-167 [FREE Full text] [doi: 10.1111/aphw.12090] [Medline: 28707767]
10. Sin NL. The protective role of positive well-being in cardiovascular disease: review of current evidence, mechanisms, and clinical implications. Curr Cardiol Rep 2016 Nov;18(11):106 [FREE Full text] [doi: 10.1007/s11886-016-0792-z] [Medline: 27612475]
11. LePine JA, LePine MA, Jackson CL. Challenge and hindrance stress: relationships with exhaustion, motivation to learn, and learning performance. J Appl Psychol 2004 Oct;89(5):883-891. [doi: 10.1037/0021-9010.89.5.883] [Medline: 15506867]
12. Folkman S. Stress: appraisal and coping. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY, USA: Springer; 2013:1913-1915.
13. Lundberg U. Stress hormones in health and illness: the roles of work and gender. Psychoneuroendocrinology 2005 Nov;30(10):1017-1021. [doi: 10.1016/j.psyneuen.2005.03.014] [Medline: 15963652]
14. Salim S. Oxidative stress: a potential link between emotional wellbeing and immune response. Curr Opin Pharmacol 2016 Aug;29:70-76. [doi: 10.1016/j.coph.2016.06.006] [Medline: 27400336]
15. Kanner AD, Coyne JC, Schaefer C, Lazarus RS. Comparison of two modes of stress measurement: daily hassles and uplifts versus major life events. J Behav Med 1981 Mar;4(1):1-39. [doi: 10.1007/BF00844845] [Medline: 7288876]
16. DeLongis A, Folkman S, Lazarus RS. The impact of daily stress on health and mood: psychological and social resources as mediators. J Pers Soc Psychol 1988 Mar;54(3):486-495. [doi: 10.1037//0022-3514.54.3.486] [Medline: 3361420]
17. Hobfoll SE. Social and psychological resources and adaptation. Rev Gen Psychol 2002 Dec 01;6(4):307-324. [doi: 10.1037/1089-2680.6.4.307]
18. Baumeister RF, Heatherton TF. Self-regulation failure: an overview. Psychol Inq 1996 Jan;7(1):1-15. [doi: 10.1207/s15327965pli0701_1]
19. IQVIA Institute for Human Data Science Study: Impact of Digital Health Grows as Innovation, Evidence and Adoption of Mobile Health Apps Accelerate. IQVIA. Durham, NC, USA: IQVIA; 2017. URL: https://www.iqvia.com/newsroom/2017/11/impact-of-digital-health-grows-as-innovation-evidence-and-adoption-of-mobile-health-apps-accelerate [accessed 2022-12-01]
20. Rickard N, Arjmand HA, Bakker D, Seabrook E. Development of a mobile phone app to support self-monitoring of emotional well-being: a mental health digital innovation. JMIR Ment Health 2016 Nov 23;3(4):e49 [FREE Full text] [doi: 10.2196/mental.6202] [Medline: 27881358]
21. Marshall JM, Dunstan D, Bartik W. Positive psychology mobile applications for increasing happiness and wellbeing - a systematic app store review. R U appy? Eur J Appl Posit Psychol 2020 Oct 1;4:12 [FREE Full text]
22. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 08;8(11):e31170 [FREE Full text] [doi: 10.2196/31170] [Medline: 34747713]
23. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
24. Chesney MA, Neilands TB, Chambers DB, Taylor JM, Folkman S. A validity and reliability study of the coping self-efficacy scale. Br J Health Psychol 2006 Sep;11(Pt 3):421-437 [FREE Full text] [doi: 10.1348/135910705X53155] [Medline: 16870053]
25. How people discover, use, and stay engaged with apps. think with Google. Mountain View, CA, USA: think with Google; 2016. URL: https://www.thinkwithgoogle.com/_qs/documents/331/how-users-discover-use-apps-google-research.pdf [accessed 2022-12-01]
26. Koval P, Hinton J, Dozo N, Gleeson J, Alvarez M, Harrison A, et al. SEMA3: Smartphone Ecological Momentary Assessment. Version 3. 2019. URL: https://sema3.com/about.html [accessed 2022-12-01]
27. Beames JR, Kikas K, Werner-Seidler A. Prevention and early intervention of depression in young people: an integrated narrative review of affective awareness and Ecological Momentary Assessment. BMC Psychol 2021 Aug 16;9(1):113 [FREE Full text] [doi: 10.1186/s40359-021-00614-6] [Medline: 34392830]
28. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985 Feb;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
29. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
30. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
31. Bostic TJ, McGartland Rubio D, Hood M. A validation of the subjective vitality scale using structural equation modeling. Soc Indic Res 2000;52(3):313-324. [doi: 10.1023/a:1007136110218]
32. Wilhelm P, Schoebi D. Assessing mood in daily life: structural validity, sensitivity to change, and reliability of a short-scale to measure three basic dimensions of mood. Eur J Psychol Assess 2007 Jan;23(4):258-267. [doi: 10.1027/1015-5759.23.4.258]
33. Hey Lemonade. URL: https://heylemonade.com.au/download/ [accessed 2023-03-04]

34. Mayfield J, Mayfield M. Motivating Language Theory: Effective Leader Talk in the Workplace. Cham, Switzerland: Springer; 2018.
35. Grant AM. Making positive change: a randomized study comparing solution-focused vs. problem-focused coaching questions. J Syst Ther 2012 Jun;31(2):21-35. [doi: 10.1521/jsyt.2012.31.2.21]
36. Grant AM, Gerrard B. Comparing problem-focused, solution-focused and combined problem-focused/solution-focused coaching approach: solution-focused coaching questions mitigate the negative impact of dysfunctional attitudes. Coaching 2020;13(1):61-77. [doi: 10.1080/17521882.2019.1599030]
37. Grant AM, O'Connor SA. The differential effects of solution‐focused and problem‐focused coaching questions: a pilot study with implications for practice. Ind Commer Train 2010;42(2):102-111. [doi: 10.1108/00197851011026090]
38. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
39. West BT, Welch KB, Galecki AT. Linear Mixed Models: A Practical Guide Using Statistical Software. 2nd edition. Boca Raton, FL, USA: Taylor and Francis; 2015.
40. Lu K, Mehrotra DV. Specification of covariance structure in longitudinal data analysis for randomized clinical trials. Stat Med 2010 Feb 20;29(4):474-488. [doi: 10.1002/sim.3820] [Medline: 20020423]
41. Mehrotra DV, Li X, Liu J, Lu K. Analysis of longitudinal clinical trials with missing data using multiple imputation in conjunction with robust regression. Biometrics 2012 Dec;68(4):1250-1259. [doi: 10.1111/j.1541-0420.2012.01780.x] [Medline: 22994905]
42. Xi W, Pennell ML, Andridge RR, Paskett ED. Comparison of intent-to-treat analysis strategies for pre-post studies with loss to follow-up. Contemp Clin Trials Commun 2018 May 9;11:20-29 [FREE Full text] [doi: 10.1016/j.conctc.2018.05.008] [Medline: 30023456]
43. Oleson JJ, Jones MA, Jorgensen EJ, Wu YH. Statistical considerations for analyzing Ecological Momentary Assessment data. J Speech Lang Hear Res 2022 Jan 12;65(1):344-360 [FREE Full text] [doi: 10.1044/2021_JSLHR-21-00081] [Medline: 34910571]
44. Lyubomirsky S, Layous K. How do simple positive activities increase well-being? Curr Dir Psychol Sci 2013 Feb 01;22(1):57-62. [doi: 10.1177/0963721412469809]
45. Sumer N, Karanci AN, Berument SK, Gunes H. Personal resources, coping self-efficacy, and quake exposure as predictors of psychological distress following the 1999 earthquake in Turkey. J Trauma Stress 2005 Aug;18(4):331-342. [doi: 10.1002/jts.20032] [Medline: 16281230]
46. Schwarzer R, Renner B. Social-cognitive predictors of health behavior: action self-efficacy and coping self-efficacy. Health Psychol 2000 Sep;19(5):487-495. [Medline: 11007157]
47. Matthews JI, Doerr L, Dworatzek PD. University students intend to eat better but lack coping self-efficacy and knowledge of dietary recommendations. J Nutr Educ Behav 2016 Jan;48(1):12-9.e1. [doi: 10.1016/j.jneb.2015.08.005] [Medline: 26424531]
48. Nicholls AR, Polman R, Levy AR. Coping self-efficacy, pre-competitive anxiety, and subjective performance among athletes. Eur J Sport Sci 2010 Mar;10(2):97-102. [doi: 10.1080/17461390903271592]
49. Bandura A. Self-efficacy conception of anxiety. Anxiety Res 1988;1(2):77-98. [doi: 10.1080/10615808808248222]
50. Chamberlain K, Zika S. The minor events approach to stress: support for the use of daily hassles. Br J Psychol 1990 Nov;81 (Pt 4):469-481. [doi: 10.1111/j.2044-8295.1990.tb02373.x] [Medline: 2279232]
51. Zika S, Chamberlain K. Relation of hassles and personality to subjective well-being. J Pers Soc Psychol 1987 Jul;53(1):155-162. [doi: 10.1037/0022-3514.53.1.155]
52. Weinberger M, Hiner SL, Tierney WM. In support of hassles as a measure of stress in predicting health outcomes. J Behav Med 1987 Feb;10(1):19-31. [doi: 10.1007/BF00845125] [Medline: 3585999]
53. Tran ST, Grotkowski K, Miller SA, Reed BW, Koven ML, Buscemi J, et al. Hassles predict physical health complaints in undergraduate students: a dynamic structural equation model analysis of daily diary data. Psychol Health 2021 Jul;36(7):828-846. [doi: 10.1080/08870446.2020.1800010] [Medline: 32779488]
54. Gruen RJ, Folkman S, Lazarus RS. Centrality and individual differences in the meaning of daily hassles. J Pers 1988 Dec;56(4):743-762. [doi: 10.1111/j.1467-6494.1988.tb00475.x] [Medline: 3230505]
55. Ryan RM, Frederick C. On energy, personality, and health: subjective vitality as a dynamic reflection of well-being. J Pers 1997 Sep;65(3):529-565. [doi: 10.1111/j.1467-6494.1997.tb00326.x] [Medline: 9327588]
56. Arslan G, Yıldırım M, Aytaç M. Subjective vitality and loneliness explain how coronavirus anxiety increases rumination among college students. Death Stud 2022;46(5):1042-1051. [doi: 10.1080/07481187.2020.1824204] [Medline: 32998663]
57. Bjorner JB, Wallenstein GV, Martin MC, Lin P, Blaisdell-Gross B, Tak Piech C, et al. Interpreting score differences in the SF-36 Vitality scale: using clinical conditions and functional outcomes to define the minimally important difference. Curr Med Res Opin 2007 Apr;23(4):731-739. [doi: 10.1185/030079907x178757] [Medline: 17407629]
58. Benight CC, Ironson G, Klebe K, Carver CS, Wynings C, Burnett K, et al. Conservation of resources and coping self-efficacy predicting distress following a natural disaster: a causal model analysis where the environment meets the mind. Anxiety Stress Coping 1999 Jan;12(2):107-126. [doi: 10.1080/10615809908248325]

59. Litt MD. Self-efficacy and perceived control: cognitive mediators of pain tolerance. J Pers Soc Psychol 1988 Jan;54(1):149-160. [doi: 10.1037//0022-3514.54.1.149] [Medline: 3346804]
60. Wellenzohn S, Proyer RT, Ruch W. Humor-based online positive psychology interventions: a randomized placebo-controlled long-term trial. J Posit Psychol 2016 Jan 28;11(6):584-594. [doi: 10.1080/17439760.2015.1137624]
61. Wellenzohn S, Proyer RT, Ruch W. How do positive psychology interventions work? A short-term placebo-controlled humor-based study on the role of the time focus. Pers Individ Dif 2016 Jul;96:1-6. [doi: 10.1016/j.paid.2016.02.056]
62. Wellenzohn S, Proyer RT, Ruch W. Who benefits from humor-based positive psychology interventions? The moderating effects of personality traits and sense of humor. Front Psychol 2018 May 28;9:821 [FREE Full text] [doi: 10.3389/fpsyg.2018.00821] [Medline: 29892252]
63. Bono JE, Ilies R. Charisma, positive emotions and mood contagion. Leadersh Q 2006 Aug;17(4):317-334. [doi: 10.1016/j.leaqua.2006.04.008]
64. Colodro H, Godoy-Izquierdo D, Godoy J. Coping self-efficacy in a community-based sample of women and men from the United Kingdom: the impact of sex and health status. Behav Med 2010;36(1):12-23. [doi: 10.1080/08964280903521362] [Medline: 20185397]
65. Education and training: census. Information on qualifications, educational attendance and type of educational institution. Australian Bureau of Statistics. Canberra, Australia: Australian Bureau of Statistics; 2021. URL: https://www.abs.gov.au/statistics/people/education/education-and-training-census/latest-release#data-downloads [accessed 2022-02-23]
66. Kristoffersen I. Great expectations: education and subjective wellbeing. J Econ Psychol 2018 Jun;66:64-78 [FREE Full text] [doi: 10.1016/j.joep.2018.04.005]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CoR: Conservation of Resource
CSE: coping self-efficacy

Edited by J Torous, G Eysenbach; submitted 11.12.22; peer-reviewed by L Bonvicini, F Venturelli; comments to author 11.01.23; revised version received 19.01.23; accepted 12.02.23; published 03.04.23.

Please cite as:
Brindal E, Kakoschke N, Golley S, Rebuli M, Baird D
Effectiveness and Feasibility of a Self-guided Mobile App Targeting Emotional Well-being in Healthy Adults: 4-Week Randomized Controlled Trial
JMIR Ment Health 2023;10:e44925
URL: https://mental.jmir.org/2023/1/e44925
doi:10.2196/44925
PMID:37010898

Original Paper

Efficacy of a Brief Blended Cognitive Behavioral Therapy Program for the Treatment of Depression and Anxiety in University Students: Uncontrolled Intervention Study

Ece Atik[1], MSc; Johannes Stricker[2], PhD; Magnus Schückes[3], MSc; Andre Pittig[1], PhD

[1]Translational Psychotherapy, Institute of Psychology, University of Goettingen, Goettingen, Germany

[2]Clinical Psychology Research Group, Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[3]University of Mannheim, Mannheim, Germany

Corresponding Author:
Johannes Stricker, PhD
Clinical Psychology Research Group
Department of Experimental Psychology
Heinrich Heine University Düsseldorf
Universitätsstr. 1
Düsseldorf, 40225
Germany
Phone: 49 211 81 10850
Email: johannes.stricker@hhu.de

Abstract

Background: Blended cognitive behavioral therapy (bCBT)—the combination of cognitive behavioral therapy and digital mental health applications—has been increasingly used to treat depression and anxiety disorders. As a resource-efficient treatment approach, bCBT appears promising for addressing the growing need for mental health care services, for example, as an early intervention before the chronification of symptoms. However, further research on the efficacy and feasibility of integrated bCBT interventions is needed.

Objective: This study aimed to evaluate the efficacy of a novel bCBT program comprising short (25 min), weekly face-to-face therapy sessions combined with a smartphone-based digital health app for treating mild to moderate symptoms of depression or anxiety.

Methods: This prospective uncontrolled trial comprised 2 measurement points (before and after treatment) and 2 intervention groups. We recruited university students with mild to moderate symptoms of depression or anxiety. On the basis of the primary symptoms, participants were assigned to either a depression intervention group (n=67 completers) or an anxiety intervention group (n=33 completers). Participants in each group received 6 weekly individual psychotherapy sessions via videoconference and completed modules tailored to their respective symptoms in the smartphone-based digital health app.

Results: The depression group displayed medium to large improvements in the symptoms of depression (Cohen d=−0.70 to −0.90; P<.001). The anxiety group experienced significant improvements in the symptoms of generalized anxiety assessed with the Generalized Anxiety Disorder-7 scale with a large effect size (Cohen d=−0.80; P<.001) but not in symptoms of anxiety assessed with the Beck Anxiety Inventory (Cohen d=−0.35; P=.06). In addition, both groups experienced significant improvements in their perceived self-efficacy (Cohen d=0.50; P<.001 in the depression group and Cohen d=0.71; P<.001 in the anxiety group) and quality of life related to psychological health (Cohen d=0.87; P<.001 in the depression group and Cohen d=0.40; P=.03 in the anxiety group). Work and social adjustment of patients improved significantly in the depression group (Cohen d=−0.49; P<.001) but not in the anxiety group (Cohen d=−0.06; P=.72). Patients' mental health literacy improved in the anxiety group (Cohen d=0.45; P=.02) but not in the depression group (Cohen d=0.21; P=.10). Patient satisfaction with the bCBT program and ratings of the usability of the digital app were high in both treatment groups.

Conclusions: This study provides preliminary evidence for the feasibility and efficacy of a novel brief bCBT intervention. The intervention effects were generalized across a broad spectrum of patient-reported outcomes. Hence, the newly developed bCBT intervention appears promising for treating mild to moderate depression and anxiety in young adults.

(JMIR Ment Health 2023;10:e44742) doi:10.2196/44742

KEYWORDS

blended cognitive behavioral therapy; bCBT; digital mental health; e–mental health; depression; anxiety disorder; video psychotherapy; mobile phone

Introduction

Background

Depression and anxiety disorders are widespread [1,2]. Both disorders have severe consequences for those affected, impairing their social and occupational lives and physical health [3-5]. According to the World Health Organization, depression is the most disabling disease in high- and middle-income countries, topping heart and cerebrovascular diseases [6]. Similarly, anxiety disorders are linked to disability, low educational attainment, and low income [7,8].

University students are particularly at risk for developing depression and anxiety disorders [9]. Several studies suggest that up to 30% of university students display elevated and distressing symptoms of depression or anxiety [10-12]. Some groups of students, such as female students [13,14], first-year students [15,16], and health students [10,17], are at an even greater risk for depression or anxiety. Many university students are in a major transitory phase of their life, maturing from adolescence into adulthood [18]. Changing life habits, family dynamics, academic challenges, and concerns about occupational life after graduation contribute to the elevated risk of developing symptoms of depression or anxiety in university students [19,20]. Elevated symptoms of depression and anxiety induce high costs for university students. For example, depression and anxiety are linked to poorer academic success [21-23] and related withdrawals from university [24].

In up to 75% of the cases, both disorders start in mid-20 years of age [25]. Early interventions for depression and anxiety, implemented before the disorders have reached more severe and chronic states, are frequently recommended [26-28]. Thus, interventions targeting mild to moderate symptoms of depression or anxiety in university students constitute a promising approach for easing the burden of mental illness and preventing severe and chronic trajectories of both disorders. As university students are typically tech-savvy [29], using digital treatment tools may be a promising approach for early interventions in this population.

Commonly recommended treatment options for depression and anxiety include psychotherapy and pharmacotherapy [30]. Regarding psychotherapeutic interventions, cognitive behavioral therapy (CBT) is a well-researched and highly effective approach for treating depression and anxiety [31]. In recent years, computerized and internet-based CBT (iCBT) have increasingly been used [32,33]. However, many studies suggest that stand-alone digital treatments without personal contact with a health care provider may suffer from several shortcomings. For example, adherence to digitalized stand-alone treatment is typically low [34], and the rates of premature treatment termination are high [35,36]. iCBT is rarely personalized and often falls short of meeting patients' individual needs [37]. Thus, its efficacy and effectiveness are limited when delivered in a stand-alone format [38,39]. The problems of stand-alone digital treatments are often also attributed to the lack of psychotherapist contact providing support and guidance throughout the treatment [35,39,40]. Thus, blended CBT (bCBT) programs, combining face-to-face CBT with digital tools, have been increasingly advocated [41,42].

Previous studies support the efficacy and feasibility of bCBT programs for treating depression and anxiety [43-45]. Some of these studies found stronger symptom reduction in patients receiving bCBT compared with CBT [46-48]. Other studies have compared bCBT programs with CBT programs that comprised more or longer face-to-face sessions than bCBT programs [44,49-51]. Importantly, in these studies, the reduced face-to-face time in bCBT compared with CBT did not lead to the reduced efficacy of bCBT programs compared with more extensive CBT programs [44,49-51]. As resources for mental health care are scarce, even in high-income countries [52], bCBT may be an effective and efficient alternative to face-to-face CBT only. In particular, as an early intervention for mild to moderate symptoms of depression and anxiety in young adults, bCBT may represent a promising treatment option. However, there is little evidence on the efficacy and acceptability of bCBT in this domain. Furthermore, bCBT has seldom been investigated in the vulnerable population of university students.

Further arguments for the use of bCBT can be found in the acceptability of bCBT treatments. Acceptance ratings by psychotherapists and patients have been found to be higher for bCBT than for stand-alone iCBT programs [53,54]. For example, a recent study investigating the attitudes of licensed psychotherapists in Germany found that most practitioners prefer bCBT over stand-alone iCBT programs [53]. During our study period (late 2021 and early 2022), the COVID-19 pandemic caused increased rates of mental disorders and psychological distress in Germany [55,56], further straining the mental health care system and leading to a growing need for mental health services worldwide [57,58]. As the prevalence rates of depression and anxiety have increased over the last few years, also outside of pandemic phases [19,59], the need for effective and resource-efficient programs, such as bCBT, is becoming increasingly pressing.

Objective

This study aimed to evaluate the efficacy and feasibility of a novel smartphone-based digital health app, *elona therapy*, for use in bCBT. This app was designed to overcome the limitations of iCBT and the existing bCBT programs. Previous bCBT interventions typically use face-to-face sessions and a smartphone-based digital health app as distinct therapeutic elements [46,60]. In contrast, *elona therapy* offers an integrated synthesis of digital and face-to-face elements where therapists can adapt the individually relevant therapeutic content that patients can access on their smartphones. Specifically, digital content incorporated into the treatment is customized by the therapist according to the individual symptoms, personal needs of the patient, and the current psychotherapeutic focus.

University students with either mild to moderate symptoms of depression or anxiety received a 6-week bCBT intervention, including face-to-face individual CBT sessions (comprising six 25-min sessions) via videoconference combined with the depression or anxiety module of *elona therapy*. Despite the briefness of the face-to-face sessions, we expected that using the bCBT program with *elona therapy* is associated with significant improvements in mental health outcomes.

Methods

Study Design

This study evaluated the efficacy and feasibility of 2 novel bCBT programs, one targeting mild to moderate symptoms of depression and one targeting mild to moderate symptoms of anxiety. We conducted a 6-week trial comprising 2 assessment points (pretreatment [T0] and posttreatment [T1]) and 2 intervention groups (depression and anxiety). During the 6-week intervention, participants received weekly individual CBT sessions (25 min each) over the web via videoconference and used the depression or anxiety module of *elona therapy* on their smartphones. At both assessment points, participants completed a broad set of self-report measures capturing symptoms of depression, symptoms of anxiety, daily functioning level, competencies, and satisfaction with the treatment. All participants provided their informed consent for study participation and data processing for research purposes before the study.

Ethics Approval

Ethics approval for this study was obtained from the Ethical Board of the University of Mannheim (EK Mannheim 27/2021).

Participants

This study comprised 2 university student samples, one with mild to moderate symptoms of depression and one with mild to moderate symptoms of anxiety. A total of 107 participants were initially recruited in the study. Depending on their symptomatology, 71 participants were assigned to the depression group and 36 participants were assigned to the anxiety intervention group. Between the pre- and postintervention points, 3 participants dropped out from the depression sample and 3 participants dropped out from the anxiety sample. The data of 1 participant from the depression sample were lost because of a technical problem. In the depression sample, 67 participants (58/67, 87% female; 9/67, 13% male; age: mean 23.61, SD 3.72; range 18-36 years) completed the study by taking part in both assessment points. In the anxiety sample, 33 participants (30/33, 91% female; 3/33, 9% male; age: mean 24.7, SD 3.5; range 19-38 years) completed the study by taking part in both assessment points. Figure 1 displays the study flowchart. Table 1 shows the demographic characteristics of study completers.

Post hoc power analyses (Cronbach α=.05) showed that the study had a high power to detect a large (99.99% for Cohen d=0.80), a high power to detect a medium (98.09% for Cohen d=0.50), and a small power to detect a small (36.46% for Cohen d=0.20) within-group effect in the depression group. Similarly, post hoc power analyses (α=.05) showed that the study had a high power to detect a large (99.37% for Cohen d=0.8), a moderate power to detect a medium (79.54% for Cohen d=0.5), and a small power to detect a small (20% for Cohen d=0.2) within-group effect in the anxiety group.

Figure 1. Study flowchart. GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9; PTSD: posttraumatic stress disorder.

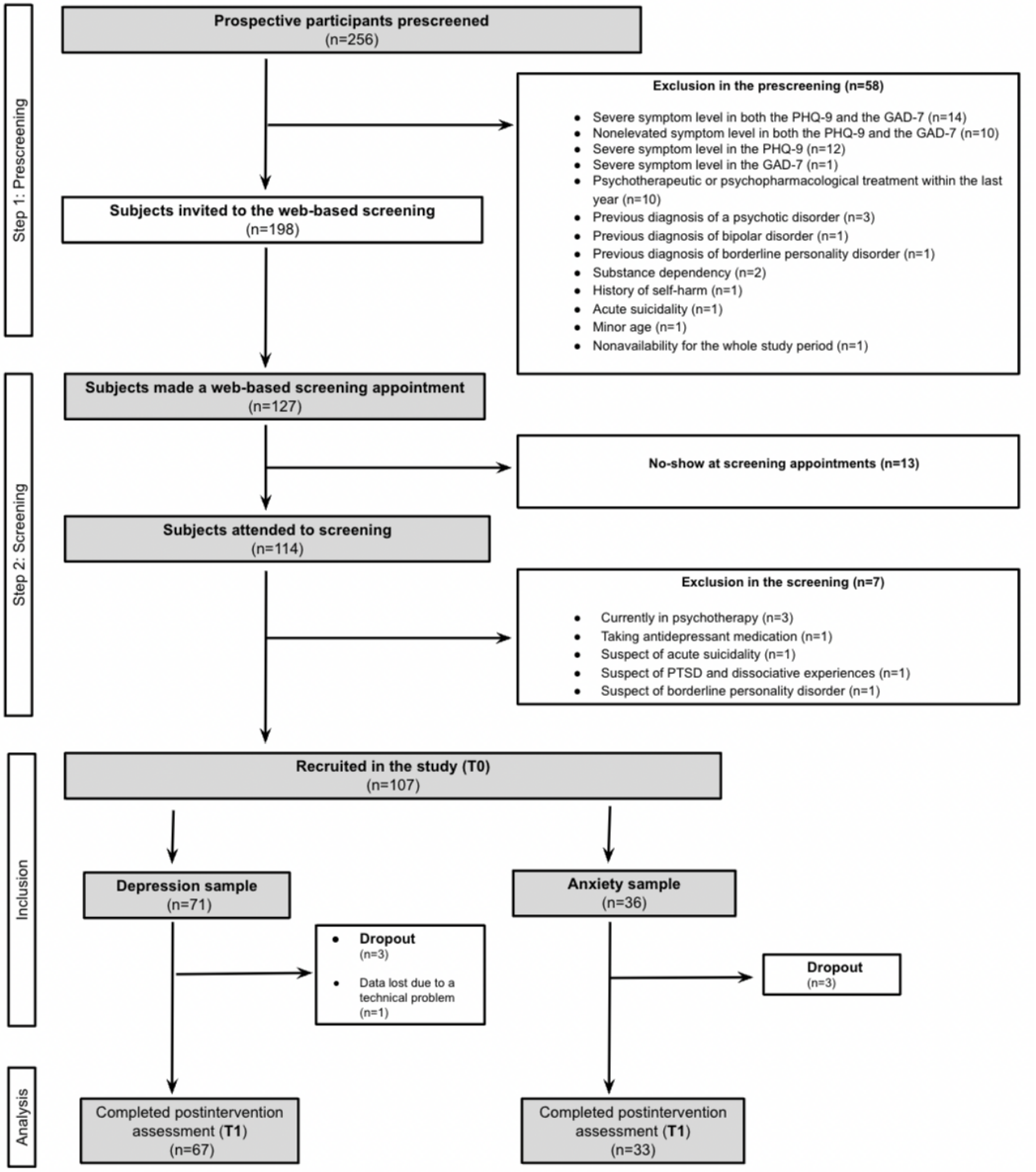


Table 1. Sample characteristics (n=67).

Characteristics	Depression sample (n=67)	Anxiety sample (n=33)
Age (years), mean (SD)	23.61 (3.72)	24.70 (3.50)
Female, n (%)	58 (87)	30 (91)
University major, n (%)		
Sports	1 (1)	0 (0)
Psychology	11 (16)	4 (12)
Medicine and related fields (eg, dentistry and pharmacy)	13 (19)	5 (15)
Social sciences and humanities	22 (33)	15 (45)
Science and engineering	12 (18)	3 (9)
Management	3 (4)	4 (12)
Recent graduates (nonuniversity students)	5 (7)	2 (6)

Recruitment

University students with mild to moderate symptoms of depression or anxiety were recruited from October 2021 to February 2022 through flyers distributed on campuses and student email groups of several universities in North Rhine-Westphalia, Germany. The study comprised a 2-step inclusion process. In the first step (prescreening), participants completed a web-based survey consisting of the Patient Health Questionnaire-9 (PHQ-9) [61,62], the Generalized Anxiety Disorder-7 (GAD-7) scale [63,64], demographic questions, and further questions examining specific exclusion criteria. Participants who met all the inclusion criteria and none of the exclusion criteria addressed in the web-based assessment were invited to the second step (screening) where they were interviewed by a clinical psychologist (Master of Science).

The inclusion criteria used in the study were as follows: (1) mild to moderate symptoms of depression or anxiety (ie, PHQ-9 or GAD-7 scores between 5 and 15 [64,65]), (2) aged between 18 and 65 years, (3) adequate proficiency in German for engaging in psychotherapy and using the *elona therapy* app, and (4) having a smartphone with iOS or Android operating systems and an internet connection. The exclusion criteria used in the study were as follows: (1) acute suicidality (a score of 2 or 3 on the item 9 of the PHQ-9), (2) current or prior International Classification of Diseases, 10th Revision (ICD-10) bipolar disorder, (3) current or prior ICD-10 borderline personality disorder, (4) current or prior ICD-10 psychotic disorder, (5) symptoms of posttraumatic stress disorder or dissociative experiences, (6) current or prior self-harm (addressed only in the screening with a therapist), (7) current or prior ICD-10 alcohol or substance dependency (except nicotine), and (8) psychotherapeutic or psychopharmacological treatment in the previous year.

The second step, the screening process, comprised a video-based interview with a clinical psychologist currently in advanced postgraduate psychotherapeutic training (interview duration: approximately 25 min). The 2-step inclusion process was used to cross-validate self-report responses with clinical assessments and to assess sensitive inclusion criteria (eg, suicidality) thoroughly and empathically. In the screening interview, inclusion and exclusion criteria were revisited by psychologists, and participants with nonelevated or severe symptoms of depression or anxiety using ICD-10 criteria and those presenting suicidality or other exclusion criteria were excluded. Those with severe symptom levels were advised to consult appropriate health care services and were provided with contact information of emergency services. Participants meeting all the inclusion criteria and none of the exclusion criteria after the second screening step were recruited to the study and assigned to either the depression or the anxiety intervention group, depending on their respective symptomatology. For persons who displayed both elevated symptoms of depression and anxiety, assignment to the depression or anxiety group was jointly decided by the participant and interviewing clinical psychologist (shared decision-making). All interviewing clinical psychologists received interview training from an experienced licensed psychotherapist before the study and supervision throughout the recruitment phase.

bCBT Intervention Program

The participants received bCBT for depression or anxiety over 6 weeks. The weekly 25-minute video-based individual CBT sessions comprised psychoeducation, internet-based therapeutic exercises (eg, creation of a behavioral activation plan and creation of a personal toolbox), and joint reflections on those tasks. Table 2 summarizes the standardized content of the face-to-face sessions in this study. The treatment manual included 18 digital tasks for the depression group and 16 digital tasks for the anxiety group. The face-to-face sessions were developed based on well-established cognitive behavioral treatment manuals for depression and anxiety [66-69]. CBT sessions were provided by clinical psychologists who currently underwent postgraduate psychotherapeutic training, were regularly seeing outpatients, and received supervision by licensed psychotherapists.

Table 2. The standardized content of study sessions for persons with increased symptoms of depression or anxiety.

Session	Manual—depression	Manual—anxiety
1	• Establishing a therapeutic relationship • Psychoeducation on the characteristics of depression regarding cognitive, affective, behavioral, and physical symptoms	• Establishing a therapeutic relationship • Psychoeducation on the characteristics of anxiety regarding cognitive, affective, behavioral, and physical symptoms
2	• Discussion of homework on personal stress factors and vulnerabilities • Psychoeducation on the vulnerability-stress model for explaining the development of depression	• Discussion of homework on personal stress factors and vulnerabilities • Psychoeducation on the vulnerability-stress model for explaining the development of anxiety
3	• Joint reflection on homework on the effects of the interaction among cognitions, affect, and behavior • Psychoeducation on downward and upward spirals	• Joint reflection on homework on individual anxiety-inducing situations for patients • Psychoeducation on the vicious circle of anxiety
4	• Discussion of activity planning homework • Creation of a concrete plan for an enjoyable activity for the week (behavioral activation)	• Psychoeducation and joint reflection on anticipatory anxiety and short- and long-term effects of avoidance and safety behaviors
5	• Discussion of the patient's experiences in the previous week's planned activity • Psychoeducation on cognitive distortions in depression, the cognitive triad of negative views by Beck [70] about oneself, the world, and the future	• Summary of previously learned relations between safety behavior and the maintenance of anxiety • Psychoeducation on the link between thoughts and bodily symptoms of anxiety and possibilities for restructuring one's catastrophic thoughts
6	• Discussion of homework on cognitive restructuring • Summary of all points learned in the bCBT[a] program. The creation of a personalized "toolbox" with previously learned and helpful content, contributing to relapse prevention	• Joint reflection on completed activities on the relation between thoughts and physical reactions • Psychoeducation on catastrophic thoughts and interactive practice of questioning them and formulating helpful alternative thoughts

[a]bCBT: blended cognitive behavioral therapy.

To improve their progress between the weekly sessions, patients engaged in additional between-session homework using *elona therapy* on their smartphones. *elona therapy* provides patients with relevant interventions, helpful activities, exercises, and psychoeducational resources related to their respective symptoms. The app content can be individualized according to the patient's needs, for example, based on their diagnosis, therapy progress, and joint therapeutic decisions. *elona therapy* has been developed based on current treatment recommendations and evidence-based CBT interventions [66-69]. The app is designed to strengthen the active participation of patients in outpatient psychotherapy and integrates therapeutic content into their daily lives. *elona therapy* provides a 2-sided interface for patients (smartphone app) and psychotherapists (web application). Through the web application, psychotherapists can track individual progress, adapt resources and interventions displayed in the patient's app, and assign tasks based on the content of the face-to-face CBT sessions. Psychotherapists can choose from over 400 available content pieces. Through the smartphone app, patients can access psychoeducational content (text, video, or audio materials), complete interactive homework tasks, record their mood in a diary format, write notes or questions relevant to their therapy sessions, and plan or schedule new activities. In addition, participants can create their personal toolkit of preferred content and tasks within the app. Although the intervention program for this study was manualized to some degree, the participating therapists were encouraged to further individualize the digital content in the *elona therapy* app appropriate for the patient's needs. Patients were encouraged to use the app approximately 25 minutes per week, mirroring a session duration. Table 3 shows the available content in the *elona therapy* app for patients with symptoms of depression or anxiety. The participating clinical psychologists received training on the use and functions of *elona therapy* by a senior licensed psychotherapist before the study.

Table 3. Content of the elona therapy app for persons with increased symptoms of depression or anxiety.

Module and chapter	Content
Depression	
About depression	Psychoeducational content and disorder-specific knowledge about psychotherapy, the symptoms of depression, individual symptoms, comorbid conditions, and factors that promote the development and maintenance of depression.
My behavior	Relationship between activities and depression, behavior analysis, and activity building and planning. Values-based work.
My thoughts	Relationship between thoughts and depression. Methods of cognitive restructuring and introduction of the concept of detached mindfulness.
My emotions	Relationship between emotions and depression. Psychoeducational content to perceive emotions in a more differentiated way, to understand them as indications of needs and support in dealing with emotions.
My relationships	Depression in a social context. Acquisition of social skills and competencies.
Relapse prevention for depression	Methods of general relapse prevention such as recognizing individual early warning signs and building a toolbox for difficult situations.
Anxiety	
About anxiety	Psychoeducational content and disorder-specific knowledge about psychotherapy, the symptoms of anxiety disorders, individual symptoms, comorbidities, and the vicious cycle of anxiety.
Emergence and maintenance of anxiety	Psychoeducational content on the development and maintenance conditions of anxiety disorders and individual application of the acquired knowledge.
Exposure methods	Information on exposure procedures. More specifically, the mechanisms of exposure are taught and the patient is given the opportunity to prepare for exposure interventions.
Thoughts and anxiety	Support in identifying and dealing with anxiety-related thoughts. Practicing methods to examine and change anxiety-related thoughts.
Emotions and anxiety	Psychoeducational content on differentiating emotions and interpreting them as indicators of psychological needs. Developing strategies for dealing with emotions.
ACT[a] for anxiety	Learning and practicing methods of ACT to deal with symptoms of anxiety
Relapse prevention for anxiety	Methods of general relapse prevention. Motivating patients to continuously practice what they have learned.
Relaxation[b]	
About relaxation	Psychoeducational content on relaxation techniques, focus on physical processes
PMR[c]	Mechanisms of action and goals of PMR according to Jacobson and related exercises
Imagination	Psychoeducation on imagination exercises and related exercises
Mindfulness and meditation	Psychoeducation on awareness and meditation techniques and related exercises

[a]ACT: Acceptance and Commitment Therapy.

[b]The Relaxation module was available to both intervention groups.

[c]PMR: progressive muscle relaxation.

Outcome Measures

This study comprised a broad set of outcome measures, covering symptoms of depression and anxiety, perceived functioning, well-being, satisfaction with the treatment program, and competencies [71-75]. We assessed the symptoms of depression and anxiety in both intervention groups as these symptoms regularly co-occur [76,77].

Symptoms of Depression

The symptoms of depression were the primary outcome for the depression group and a secondary outcome for the anxiety group. We assessed the symptoms of depression with 2 well-established self-report instruments: the PHQ-9 [61,62] and the Beck Depression Inventory-II (BDI-II) [78,79]. The PHQ-9 assesses the frequency of symptoms of depression over the past 2 weeks with 9 items (eg, feeling tired or having little energy) using a 4-point Likert scale ranging from 0 (“not at all”) to 3 (“nearly every day”). The BDI-II assesses the severity of symptoms of depression over the past 2 weeks with 21 items (eg, loss of interest) using a 4-point Likert scale ranging from 0 (eg, “I have not lost interest in other people or activities”) to 3 (eg, “It’s hard to get interested in anything”). Many studies have demonstrated the reliability and validity of the German PHQ-9 and BDI-II instruments [79-81]. The Cronbach α value in this study was .85 at T0 and .88 at T1 for BDI-II and .70 at T0 and .78 at T1 for PHQ-9.

Symptoms of Anxiety

The symptoms of anxiety were the primary outcome for the anxiety group and a secondary outcome for the depression

group. We assessed the symptoms of anxiety with 2 well-established self-report instruments: the GAD-7 [63,64] and the Beck Anxiety Inventory (BAI) [82,83]. The GAD-7 contains 7 items assessing the severity of symptoms of generalized anxiety (eg, trouble relaxing) over the past 2 weeks using a 4-point Likert scale ranging from 0 ("not at all") to 3 ("nearly every day"). The BAI comprises 21 items assessing the severity of common symptoms of anxiety (eg, numbness or tingling) over the past month on a 4-point Likert scale ranging from 0 ("not at all") to 3 ("severely—it bothered me a lot"). The German GAD-7 and BAI instruments display satisfactory reliability and validity [63,84]. The Cronbach α value in this study was .79 at T0 and .84 at T1 for GAD-7 and .90 at T0 and .91 at T1 for BAI.

Quality of Life

Quality of life was assessed with the WHOQOL-BREF scale [85,86]. This standardized 26-item self-report instrument comprises 5 subscales: overall quality of life and general health (2 items), physical health (7 items), psychological health (6 items), quality of life related to social relationships (3 items), and quality of life related to one's environment (8 items). Participants rated each item on a 5-point Likert scale (eg, 1="not at all" to 5="completely"). Previous studies have demonstrated the reliability and validity of the German WHOQOL-BREF [87]. The Cronbach α value in this study was .72 at T0 and .72 at T1 for the overall quality of life and general health domain, .70 at T0 and .77 at T1 for the physical health domain, .78 at T0 and .81 at T1 for the psychological health domain, .47 at T0 and .61 at T1 for the social relationships domain, and .72 at T0 and .73 at T1 for the environment domain.

Functioning in Work and Social Life

Patients' adjustment to work and social life was assessed with the Work and Social Adjustment Scale (WSAS) [88,89]. This instrument captures impairments in relevant life domains beyond the symptom level [88,89]. The WSAS contains 5 items (eg, "Because of my depression/anxiety, my ability to work is impaired") rated on a 9-point Likert scale ranging from 0 ("not at all impaired") to 8 ("very severely impaired to the point I can't work"). Higher scores on this scale indicate higher maladjustment. The German WSAS has been previously shown to be a reliable and valid measurement instrument [88]. The Cronbach α value in this study was .77 at T0 and .82 at T1.

Perceived Self-Efficacy

We assessed the participants' perceived self-efficacy with the General Perceived Self-Efficacy Scale (GSE) [90]. This scale measures beliefs about one's capacity to cope with new, difficult, or stressful situations. The GSE contains 10 items (eg, "It is easy for me to stick to my aims and accomplish my goals") on a 4-point Likert scale ranging from 1 *strongly disagree* to 4 *strongly agree*. Previous work demonstrated the reliability and validity of the GSE [91,92]. The Cronbach α value in this study was .88 at T0 and .90 at T1.

Mental Health Literacy

We assessed mental health literacy (ie, patients' knowledge about mental disorders and their treatments) with a 20-item version of the Mental Health Literacy Scale (MHLS) [93,94]. The 20-item MHLS includes items assessing general mental health literacy (eg, "I am confident that I know where to seek information about mental illness") using a 5-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). The MHLS items used in our study were the items numbered 16 to 35 in the original 35-item MHLS. The items not used in this study focused on identifying different specific mental disorder diagnoses, which is beyond the scope of this study. The original MHLS demonstrated adequate internal consistency and validity [94]. The Cronbach α value for the MHLS in this study was .83 at T0 and .82 at T1.

Satisfaction With Treatment

We assessed patient satisfaction with their treatment at T1 using the German version of the Client Satisfaction Questionnaire (CSQ-8) [95,96]. The CSQ-8 captures patient satisfaction using 8 items (eg, "Have the services you received helped you to deal more effectively with your problems?") with a 4-point Likert scale ranging from 1 (eg, "No, they seemed to make things worse") to 4 (eg, "Yes, they helped a great deal"). Higher CSQ-8 scores (range 8-32) indicate higher patient satisfaction. The German CSQ-8 has been extensively validated for assessing patient satisfaction in mental health treatment [96,97]. The Cronbach α value in this study was .93.

Usability of the elona therapy App

The usability of the *elona therapy* app was assessed with the German version of the System Usability Scale (SUS) [98,99]. The SUS is a self-report instrument comprising 10 items with a 5-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). The SUS raw scores are summed and then multiplied by 2.5 to facilitate interpretation. According to the SUS interpretation guidelines, a transformed score >68 indicates an above-average usability [100,101]. The German version of the SUS has shown satisfactory reliability and validity in various studies [99]. The Cronbach α value in this study was .89.

Serious Adverse Events

Participating therapists were instructed to report in case their patients experienced any serious adverse events. Serious adverse events include negative events that are harmful or threatening beyond transient distress and are life-threatening or require hospitalization. Self-harm, suicidal ideation, or attempts are potentially serious adverse events.

Statistical Analyses

We evaluated the changes from before to after the treatment using 2-tailed paired sample *t* tests. Statistical analyses were conducted based on completers, including only participants who have completed at least 1 questionnaire after treatment. In addition, we evaluated our measurement instruments' reliability (Cronbach α), and we computed the pre-post effect sizes (Cohen *d*) for each outcome. For the depression group, we treated the depression measures (PHQ-9 and BDI-II) as primary outcomes and all other measures as secondary outcomes. For the anxiety group, we treated the anxiety measures (BAI and GAD-7) as primary outcomes and all other measures as secondary outcomes.

Results

Engagement With the Intervention

On average, participants in the depression sample completed 21.6 tasks and spent a total of 88.12 minutes using the *elona therapy* app during the intervention period. Participants in the anxiety group completed on average 24.1 tasks and spent 106.9 minutes using the *elona therapy* app during the intervention period. The mean number of completed face-to-face sessions in the depression group was 5.64 (SD 0.64; range 2-6), and the mean number of completed face-to-face sessions in the anxiety group was 5.55 (SD 0.62; range 4-6). Patient engagement with *elona therapy* was high; only 1 patient in the depression group was a minimal user, which is defined as using *elona therapy* for <15 minutes for the entire study period, with only 7 minutes of total use time.

Depression Group

Table 4 displays the means and SDs for the pre- and postintervention points, the results of the 2-tailed paired samples *t* tests, and the corresponding effect sizes (Cohen *d*) for the depression group. Regarding symptoms of depression (primary outcome), we observed significant improvements in the PHQ-9 score (t_{63}=−7.23, *P*<.001; Cohen *d*=−0.90) and the BDI-II score (t_{65}=−5.66, *P*<.001; Cohen *d*=−0.70). Regarding the secondary outcomes, the symptoms of anxiety assessed with the GAD-7 (t_{60}=−6.71, *P*<.001; Cohen *d*=−0.86) and BAI (t_{35}=−4.55, *P*<.001; Cohen *d*=−0.76) and maladjustment assessed with the WSAS (t_{58}=−3.74, *P*<.001; Cohen *d*=−0.49) significantly decreased. In addition, perceived self-efficacy assessed with the GSE (t_{60}=3.93, *P*<.001; Cohen *d*=0.50) and quality of life assessed with the WHOQOL-BREF (t_{61}=6.48, *P*<.001; Cohen *d*=0.82 for overall quality of life and general health; t_{61}=6.31, *P*<.001; Cohen *d*=0.80 for physical health; t_{61}=6.85, *P*<.001; Cohen *d*=0.87 for psychological health; t_{61}=4.27, *P*<.001; Cohen *d*=0.54 for quality of life related to social relationships; and t_{63}=4.79, *P*<.001; Cohen *d*=0.61 for quality of life related to one's environment) significantly improved. One secondary outcome, mental health literacy in the MHLS, did not significantly improve during the intervention period (t_{59}=1.66, *P*=.10; Cohen *d*=0.21). At the postintervention point, participants rated the usability of the smartphone-based digital health app for treating depression as high (mean usability score of 74.15 out of 100 on the SUS, corresponding to above-average usability [98]). Participants also rated their satisfaction with the bCBT intervention program as high (mean 26.14 out of 32 on the CSQ-8). There were no serious adverse events reported.

Table 4. Means and SDs for the pre- and postintervention points, results of the 2-tailed paired samples *t* tests, corresponding effect sizes (Cohen *d*), and sample size for each comparison in the depression group.

Outcome	Preintervention point, mean (SD)	Postintervention point, mean (SD)	*t* test (*df*)[a]	*P* value	Effect size[b], Cohen *d*	Sample size, n[c]
Symptoms of depression (PHQ-9[d])	10.82 (3.72)	6.69 (3.47)	−7.23 (63)	<.001	−0.90	64
Symptoms of depression (BDI-II[e])	19.40 (7.46)	13.33 (7.54)	−5.66 (65)	<.001	−0.70	66
Symptoms of anxiety (GAD-7[f])	8.12 (3.83)	5.23 (2.88)	−6.71 (60)	<.001	−0.86	61
Symptoms of anxiety (BAI[g])	12.83 (8.25)	8.61 (5.99)	−4.55 (35)	<.001	−0.76	36
Perceived self-efficacy (GSE[h])	25.03 (5.37)	27.07 (5.34)	3.93 (60)	<.001	0.5	61
Work and social adjustment (WSAS[i])	18.09 (7.68)	14.42 (7.71)	−3.74 (58)	<.001	−0.49	59
Mental health literacy (MHLS[j])	84.91 (8.10)	85.80 (8.3)	1.66 (59)	.10	0.21	60
Quality of life and general health (WHOQOL-BREF[k])	6.63 (1.66)	7.65 (1.51)	6.48 (61)	<.001	0.82	62
Physical health (WHOQOL-BREF)	25.10 (3.83)	27.98 (3.62)	6.31 (61)	<.001	0.8	62
Psychological health (WHOQOL-BREF)	17.51 (3.58)	20.26 (3.78)	6.85 (61)	<.001	0.87	62
Quality of life—social (WHOQOL-BREF)	9.85 (2.2)	10.97 (2.27)	4.27 (61)	<.001	0.54	62
Quality of life—environmental (WHOQOL-BREF)	31.34 (4.48)	33.24 (3.85)	4.79 (61)	<.001	0.61	62

[a]The *t* test value associated with the 2-tailed paired sample *t* test assessing pre-post differences.

[b]Effect size=Cohen *d* associated with the respective pre-post difference.

[c]Some participants failed to complete specific postintervention measures, leading to slight deviations in the sample size between the comparisons.

[d]PHQ-9: Patient Health Questionnaire-9 [61,62].

[e]BDI-II: Beck Depression Inventory-II [78,79].

[f]GAD-7: Generalized Anxiety Disorder-7 [63,64].

[g]BAI: Beck Anxiety Inventory [82,83].

[h]GSE: General Self-Efficacy Scale [90].

[i]WSAS: Work and Social Adjustment Scale [88,89].

[j]MHLS: Mental Health Literacy Scale [93,94].

[k]WHOQOL−BREF: World Health Organization Quality of Life Scale (brief version) [85,86].

Anxiety Group

Table 5 displays the means and SDs for the pre- and postintervention points, the results of the 2-tailed paired samples *t* tests, and the corresponding effect sizes (Cohen *d*) for the anxiety group. Regarding symptoms of anxiety (primary outcome), we observed significant improvements in the GAD-7 score (t_{30}=−4.45, *P*<.001; Cohen *d*=−0.80) but not in the BAI score (t_{31}=−1.96, *P*=.06; Cohen *d*=−0.35). Regarding the secondary outcomes, the symptoms of depression assessed with the PHQ-9 (t_{32}=−4.58, *P*<.001; Cohen *d*=−0.8) and the BDI-II (t_{21}=−3.53, *P*=.002; Cohen *d*=−0.75) significantly decreased. In addition, perceived self-efficacy (t_{30}=3.97, *P*<.001; Cohen *d*=0.71), mental health literacy (t_{28}=2.07, *P*=.048; Cohen *d*=0.38), overall quality of life and general health (t_{30}=2.53, *P*=.02; Cohen *d*=0.45), and psychological health (t_{30}=2.21, *P*=.03; Cohen *d*=0.40) assessed with the WHOQOL-BREF significantly improved. Maladjustment assessed with the WSAS (t_{31}=−0.36, *P*=.72; Cohen *d*=−0.06), physical health (t_{30}=1.75, *P*=.09; Cohen *d*=0.31), quality of life related to social relationships (t_{30}=0.18, *P*=.86; Cohen *d*=0.03), and quality of life related to one's environment (t_{30}=1.38, *P*=.18; Cohen *d*=0.25) assessed with the WHOQOL-BREF did not significantly improve in the intervention period. At the postintervention point, participants rated the usability of the smartphone-based digital health app for treating symptoms of anxiety as high (mean usability score of 81.68 [out of 100] on the SUS, corresponding to above-average usability [98]). Participants also rated their satisfaction with the bCBT intervention program as high (mean 28.21 out of 32 on the CSQ-8). There were no serious adverse events reported.

Table 5. Means and SDs for the pre- and postintervention points, results of the 2-tailed paired samples *t* tests, corresponding effect sizes (Cohen *d*), and sample size for each comparison in the anxiety group.

Outcome	Preintervention point, mean (SD)	Postintervention point, mean (SD)	*t* test (*df*)[a]	*P* value	Effect size[b], Cohen *d*	Sample size, n[c]
Symptoms of anxiety (GAD-7[d])	11.33 (3.93)	8.23 (4.91)	−4.45 (30)	<.001	−0.80	31
Symptoms of anxiety (BAI[e])	21 (10.48)	17.69 (10.47)	−1.96 (31)	.06	−0.35	32
Symptoms of depression (PHQ-9[f])	10.7 (4.28)	7.3 (4.33)	−4.58 (32)	<.001	−0.80	33
Symptoms of depression (BDI-II[g])	18.06 (9.92)	12.95 (9.22)	−3.53 (21)	.002	−0.75	22
Perceived self-efficacy (GSE[h])	24.94 (6.43)	27.19 (5.95)	3.97 (30)	<.001	0.71	31
Work and social adjustment (WSAS[i])	16.61 (8.21)	16.03 (8.15)	−0.36 (31)	.72	−0.06	32
Mental health literacy (MHLS[j])	86.45 (8.17)	87.34 (7.18)	2.07 (28)	.048	.38	29
Quality of life and general health (WHOQOL-BREF[k])	6.7 (1.74)	7.16 (1.66)	2.53 (30)	.02	0.45	31
Physical health (WHOQOL-BREF)	24.79 (4.31)	25.9 (4.77)	1.75 (30)	.09	0.31	31
Psychological health (WHOQOL-BREF)	18.39 (4.59)	19.13 (4.64)	2.21 (30)	.03	0.4	31
Quality of life—social (WHOQOL-BREF)	10.33 (2.29)	10.32 (2.23)	0.18 (30)	.86	0.03	31
Quality of life—environmental (WHOQOL-BREF)	30.09 (4.33)	30.68 (3.78)	1.38 (30)	.18	0.25	31

[a]The *t* test value associated with the 2-tailed paired sample *t* test assessing pre-post differences.

[b]Effect size=Cohen *d* associated with the respective pre-post difference.

[c]Some participants failed to complete some postintervention measures, leading to slight deviations in the sample size between the comparisons.

[d]GAD-7: Generalized Anxiety Disorder-7 [63,64].

[e]BAI: Beck Anxiety Inventory [82,83].

[f]PHQ-9: Patient Health Questionnaire-9 [61,62].

[g]BDI-II: Beck Depression Inventory-II [78,79].

[h]GSE: General Perceived Self-Efficacy Scale [90].

[i]WSAS: Work and Social Adjustment Scale [88,89].

[j]MHLS: Mental Health Literacy Scale [93,94].

[k]WHOQOL-BREF: World Health Organization Quality of Life Scale (brief version) [85,86].

Discussion

Principal Findings

This study investigated the efficacy and feasibility of a 6-week bCBT program for mild to moderate depression or anxiety in university students. Overall, participants showed improvements in their primary symptoms, which generalized to other symptom domains, quality of life, and self-efficacy. In addition, participants reported high satisfaction with the intervention program and the *elona therapy* app. The intervention evaluated in this study produced within-group effect sizes that were considerably larger than those observed in waitlist control groups over similar periods [102,103] suggesting that the brief bCBT program may be efficacious.

In addition, a previous meta-analysis on the effectiveness of standard evidence-based psychological treatments in practices revealed similarly large pre-post treatment effect sizes for depression (Cohen *d*=0.87) and anxiety (Cohen *d*=0.88), as observed in our study [104]. Importantly, the studies included in this meta-analysis comprised, on average, more face-to-face time with a therapist [104]. Similarly, the effect sizes obtained in this study were similar to the effect sizes observed in previous, more extensive bCBT programs for patients with depression or anxiety disorders, ranging from 12 to 15 weeks [46,105]. Various factors such as measurement instruments and sample characteristics may influence effect sizes, so treatment effects should only be cautiously compared between studies. However, the current data indicated that compared with previous approaches, the evaluated bCBT program is at least an equally effective and resource-efficient intervention for reducing the symptoms of depression and anxiety in university students.

Regarding within-group differences between intervention outcomes, we observed larger effects for the symptoms of depression or anxiety than for secondary outcomes such as work and social adjustment or self-efficacy. This pattern of findings is in line with a previous meta-analysis [104] and may be explained by the explicit focus of our intervention on symptom reduction. In the anxiety group, significant symptom reduction was observed in the GAD-7 but not in the BAI (*P*=.06). The GAD-7 primarily assesses symptoms of generalized (cognitive) anxiety, whereas the BAI comprises various items capturing physical symptoms of anxiety (eg, heat sensations, accelerated heart rate, and shakiness). Our intervention's focus on the cognitive aspects of anxiety and psychoeducation rather than in vivo exposure therapy may explain this finding.

Although work and social adjustment improved significantly in the depression group, this was not the case for the anxiety group. This may be because of worse work and social adjustment before the intervention in the individuals with the symptoms of depression compared with those with symptoms of anxiety, as it has been found earlier in the literature [106,107]. Similarly, all domains of quality of life significantly improved in the depression group and only some improved in the anxiety group. Follow-up studies with similar sample sizes across symptom groups (ie, depression or anxiety) are required to understand whether the observed differences may be because of the lower sample size of the anxiety group compared with the depression group in this study. Moreover, the intervention program we offered in the anxiety group appeared significantly effective in improving mental health literacy, whereas an improvement in this domain was not significant in the depression group. It remains unclear whether this is because of different patient characteristics in groups or different content provided for those groups, either within sessions or in the *elona therapy* app (Tables 2 and 3).

Regarding the feasibility of the evaluated program, participants in both study groups rated their satisfaction with the intervention as high. Satisfaction ratings were similar to those provided in the study by Romijn et al [105], who found that participants rated their satisfaction with the bCBT program to be 25.61 (SD 4.21) on average or higher than that reported by Kooistra et al [108], who found the mean satisfaction to be 22.71 (SD 4.82). Notably, these previous bCBT programs are more extensive than those we provided, with the one in the study by Romijn et al [105] lasting 15 weeks and the one in the study by Kooistra et al [108] lasting 10 weeks. Participants also indicated high usability of the *elona therapy* modules, corresponding to acceptable usability according to guidelines [101]. High feasibility was also evident from the remarkably low attrition rates. Approximately 94% of the participants in the depression group and 92% of the participants in the anxiety group completed the 6-week bCBT program as intended. Moreover, the participants' good engagement with the app during the study period further supports the feasibility of the *elona therapy* app. The blended treatment program introduced in this study appears to be a safe treatment option, as no adverse events were reported in this study.

Limitations and Future Research Directions

As this study was planned as an uncontrolled trial, several limitations point toward future research directions. First, this study comprised only 2 measurement points (before and after the assessment). Future research with more measurement points is needed to establish the ideal duration of the evaluated bCBT intervention (eg, to establish whether incremental symptom reduction occurs after the sixth week). Further work should also test the stability of the obtained improvements through follow-up assessments. Second, as a feasibility study focusing on within-group improvements, this study did not include control groups. Thus, future research using appropriate active (eg, standard CBT treatment) or passive (eg, waitlist control) groups is needed. In addition, dismantling studies could be useful for evaluating the effects of different elements of the *elona therapy* modules. Third, we only performed a completer analysis in this trial, which may have overestimated the intervention's benefits. However, given the small number of dropouts in both groups, the differences in completer and intent-to-treat analyses were small. Fourth, this study focused on a convenience sample, including a vulnerable population (ie, university students) with mild to moderate symptoms. The findings may have limited generalizability to populations with different characteristics and to other student samples in different cultures or different regions of Germany. Fifth, we only used patient-reported outcomes in this study. The use of only self-report may provide a limited depiction of patients' situations owing to patients' possible inaccurate evaluation of their situation or response biases such as social desirability bias. Further work can complement some of the outcome measures (eg, symptoms of depression and anxiety) by a clinical rating scale such as therapist rating questionnaires.

This trial used only a quantitative study design, which may inform researchers only limitedly about the investigated program. To explore patients' in-depth experiences with this novel bCBT program, an additional qualitative study was needed. As a follow-up to this trial, we invited all participants who had completed the last face-to-face session with their therapist to participate in a follow-up interview conducted by an interviewer who was not involved in the therapeutic process. The results of this qualitative investigation were published in a separate article [109].

Further work is needed to assess the efficacy and feasibility of the evaluated bCBT program in other groups and individuals with more severe symptomatology.

In summary, this feasibility study provided preliminary evidence based on an uncontrolled pre-post trial design for the efficacy of a novel bCBT program for treating mild to moderate depression and anxiety. Promisingly, substantial symptom improvements were generalized across various outcomes and were accompanied by high patient satisfaction and adherence. Future research is needed to evaluate the efficacy of the evaluated program in different populations and over longer time intervals.

Acknowledgments

The authors would like thank all therapists and patients who participated in the study. This project was funded by the Robert-Enke Foundation (Robert-Enke-Stiftung), a nonprofit organization that supports projects, initiatives, and institutions that educate people about depressive disorders and facilitates their research or treatment. The funding source did not play any role in the study design, analysis, or interpretation of the results. The authors would like to thank the Heinrich Heine University Düsseldorf for funding the publication fees (University Library, Open-Access-Funds of the Heinrich Heine University Düsseldorf).

Conflicts of Interest

EA declares that she is employed by Elona Health, the manufacturer of the elona therapy app used in this study. JS and AP are consultants to Elona Health. MS is a shareholder of Elona Health.

References

1. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 2022 Apr 01;17(3):327-335 [FREE Full text] [doi: 10.31887/dcns.2015.17.3/bbandelow]
2. Gelenberg AJ. The prevalence and impact of depression. J Clin Psychiatry 2010 Mar 15;71(3):e06 [FREE Full text] [doi: 10.4088/JCP.8001tx17c] [Medline: 20331925]
3. Haller H, Cramer H, Lauche R, Gass F, Dobos GJ. The prevalence and burden of subthreshold generalized anxiety disorder: a systematic review. BMC Psychiatry 2014 May 01;14(1):128 [FREE Full text] [doi: 10.1186/1471-244x-14-128]
4. Hansson L. Quality of life in depression and anxiety. Int Rev Psychiatry 2009 Jul 11;14(3):185-189 [FREE Full text] [doi: 10.1080/09540260220144966]
5. Saarijärvi S, Salminen JK, Toikka T, Raitasalo R. Health-related quality of life among patients with major depression. Nord J Psychiatry 2002;56(4):261-264 [FREE Full text] [doi: 10.1080/08039480260242741] [Medline: 12470316]
6. The global burden of disease : 2004 update. World Health Organization. 2008. URL: https://apps.who.int/iris/handle/10665/43942 [accessed 2022-04-25]
7. Hendriks SM, Spijker J, Licht CM, Beekman AT, Hardeveld F, de Graaf R, et al. Disability in anxiety disorders. J Affect Disord 2014 Oct;166:227-233 [FREE Full text] [doi: 10.1016/j.jad.2014.05.006] [Medline: 25012435]
8. Kessler RC, Ruscio AM, Shear K, Wittchen HU. Epidemiology of anxiety disorders. In: Stein M, Steckler T, editors. Behavioral Neurobiology of Anxiety and Its Treatment. Berlin, Germany: Springer; 2009.
9. Mofatteh M. Risk factors associated with stress, anxiety, and depression among university undergraduate students. AIMS Public Health 2021;8(1):36-65 [FREE Full text] [doi: 10.3934/publichealth.2021004] [Medline: 33575406]
10. de Paula W, Breguez GS, Machado EL, Meireles AL. Prevalence of anxiety, depression, and suicidal ideation symptoms among university students: a systematic review. Braz J Health Rev 2020;3(4):8739-8756 [FREE Full text] [doi: 10.34119/bjhrv3n4-119]
11. Ozen NS, Ercan I, Irgil E, Sigirli D. Anxiety prevalence and affecting factors among university students. Asia Pac J Public Health 2010 Jan;22(1):127-133 [FREE Full text] [doi: 10.1177/1010539509352803] [Medline: 20032042]
12. Shamsuddin K, Fadzil F, Ismail WS, Shah SA, Omar K, Muhammad NA, et al. Correlates of depression, anxiety and stress among Malaysian university students. Asian J Psychiatr 2013 Aug;6(4):318-323 [FREE Full text] [doi: 10.1016/j.ajp.2013.01.014] [Medline: 23810140]
13. Bahrami F, Yousefi N. Females are more anxious than males: a metacognitive perspective. Iran J Psychiatry Behav Sci 2011;5(2):83-90 [FREE Full text] [Medline: 24644451]
14. Boggiano AK, Barrett M. Gender differences in depression in college students. Sex Roles 1991 Dec;25(11-12):595-605 [FREE Full text] [doi: 10.1007/bf00289566]
15. Bayram N, Bilgel N. The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. Soc Psychiatry Psychiatr Epidemiol 2008 Aug;43(8):667-672 [FREE Full text] [doi: 10.1007/s00127-008-0345-x] [Medline: 18398558]
16. Mahmoud JS, Staten R, Hall LA, Lennie TA. The relationship among young adult college students' depression, anxiety, stress, demographics, life satisfaction, and coping styles. Issues Ment Health Nurs 2012 Mar;33(3):149-156 [FREE Full text] [doi: 10.3109/01612840.2011.632708] [Medline: 22364426]
17. Moir F, Yielder J, Sanson J, Chen Y. Depression in medical students: current insights. Adv Med Educ Pract 2018 May;Volume 9:323-333 [FREE Full text] [doi: 10.2147/amep.s137384]
18. Sarokhani D, Delpisheh A, Veisani Y, Sarokhani MT, Manesh RE, Sayehmiri K. Prevalence of depression among university students: a systematic review and meta-analysis study. Depress Res Treat 2013;2013:373857 [FREE Full text] [doi: 10.1155/2013/373857] [Medline: 24187615]
19. Beiter R, Nash R, McCrady M, Rhoades D, Linscomb M, Clarahan M, et al. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. J Affect Disord 2015 Mar 01;173:90-96 [FREE Full text] [doi: 10.1016/j.jad.2014.10.054] [Medline: 25462401]
20. Depression and college students. National Institute of Mental Health. 2012. URL: https://www.pugetsound.edu/sites/default/files/file/1784_depression-and-college-students_0.pdf [accessed 2022-04-25]
21. Andrews B, Wilding JM. The relation of depression and anxiety to life-stress and achievement in students. Br J Psychol 2004 Nov;95(Pt 4):509-521 [FREE Full text] [doi: 10.1348/0007126042369802] [Medline: 15527535]
22. Eisenberg D, Gollust SE, Golberstein E, Hefner JL. Prevalence and correlates of depression, anxiety, and suicidality among university students. Am J Orthopsychiatry 2007 Oct;77(4):534-542 [FREE Full text] [doi: 10.1037/0002-9432.77.4.534] [Medline: 18194033]
23. Hysenbegasi A, Hass SL, Rowland CR. The impact of depression on the academic productivity of university students. J Ment Health Policy Econ 2005 Sep;8(3):145-151. [Medline: 16278502]

24. Meilman P, Manley C, Gaylor MS, Turco JH. Medical withdrawals from college for mental health reasons and their relation to academic performance. J Am Coll Health 1992 Mar;40(5):217-223 [FREE Full text] [doi: 10.1080/07448481.1992.9936283] [Medline: 1573130]
25. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Üstün TB. Age of onset of mental disorders: a review of recent literature. Curr Opinion Psychiatry 2007 Jul;20(4):359-364 [FREE Full text] [doi: 10.1097/yco.0b013e32816ebc8c]
26. Cuijpers P, van Straten A, Smit F, Mihalopoulos C, Beekman A. Preventing the onset of depressive disorders: a meta-analytic review of psychological interventions. Am J Psychiatry 2008 Oct;165(10):1272-1280 [FREE Full text] [doi: 10.1176/appi.ajp.2008.07091422] [Medline: 18765483]
27. Feldner MT, Zvolensky MJ, Schmidt NB. Prevention of anxiety psychopathology: a critical review of the empirical literature. Clin Psychol Sci Pract 2004;11(4):405-424 [FREE Full text] [doi: 10.1093/clipsy.bph098]
28. van't Veer-Tazelaar PJ, van Marwijk HW, van Oppen P, van der Horst HE, Smit F, Cuijpers P, et al. Prevention of late-life anxiety and depression has sustained effects over 24 months: a pragmatic randomized trial. Am J Geriatr Psychiatry 2011 Mar;19(3):230-239 [FREE Full text] [doi: 10.1097/jgp.0b013e3181faee4d] [Medline: 21425519]
29. Prensky M. Digital natives, digital immigrants part 1. On Horizon 2001 Sep;9(5):1-6 [FREE Full text] [doi: 10.1108/10748120110424816]
30. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry 2013 Jun;12(2):137-148 [FREE Full text] [doi: 10.1002/wps.20038] [Medline: 23737423]
31. David D, Cristea I, Hofmann SG. Why cognitive behavioral therapy is the current gold standard of psychotherapy. Front Psychiatry 2018;9:4 [FREE Full text] [doi: 10.3389/fpsyt.2018.00004] [Medline: 29434552]
32. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
33. Cuijpers P, Kleiboer A, Karyotaki E, Riper H. Internet and mobile interventions for depression: opportunities and challenges. Depress Anxiety 2017 Jul;34(7):596-602 [FREE Full text] [doi: 10.1002/da.22641] [Medline: 28471479]
34. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468 [FREE Full text] [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]
35. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342 [FREE Full text] [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
36. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2008 Sep 24;39(5):705-712 [FREE Full text] [doi: 10.1017/s0033291708004224]
37. Andrews G, Williams AD. Internet psychotherapy and the future of personalized treatment. Depress Anxiety 2014 Nov;31(11):912-915 [FREE Full text] [doi: 10.1002/da.22302] [Medline: 25407580]
38. Weisel K, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
39. Wright J, Mishkind M, Eells TD, Chan SR. Computer-assisted cognitive-behavior therapy and mobile apps for depression and anxiety. Curr Psychiatry Rep 2019 Jun 27;21(7):62 [FREE Full text] [doi: 10.1007/s11920-019-1031-2] [Medline: 31250242]
40. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):89-103 [FREE Full text] [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
41. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
42. Newby J, Mason E, Kladnistki N, Murphy M, Millard M, Haskelberg H, et al. Integrating internet CBT into clinical practice: a practical guide for clinicians. Clin Psychol 2021 Mar 15;25(2):164-178 [FREE Full text] [doi: 10.1080/13284207.2020.1843968]
43. Bruinsma A, Kampman M, Exterkate CC, Hendriks GJ. [An exploratory study of 'blended' cognitive behavioural therapy (CBT) for patients with a panic disorder: results and patients' experiences]. Tijdschr Psychiatr 2016;58(5):361-370 [FREE Full text] [Medline: 27213635]
44. Kooistra LC, Wiersma JE, Ruwaard J, Neijenhuijs K, Lokkerbol J, van Oppen P, et al. Cost and effectiveness of blended versus standard cognitive behavioral therapy for outpatients with depression in routine specialized mental health care: pilot randomized controlled trial. J Med Internet Res 2019 Oct 29;21(10):e14261 [FREE Full text] [doi: 10.2196/14261] [Medline: 31663855]
45. Lungu A, Jun JJ, Azarmanesh O, Leykin Y, Chen CE. Blended care-cognitive behavioral therapy for depression and anxiety in real-world settings: pragmatic retrospective study. J Med Internet Res 2020 Jul 06;22(7):e18723 [FREE Full text] [doi: 10.2196/18723] [Medline: 32628120]

46. Berger T, Krieger T, Sude K, Meyer B, Maercker A. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: results of a pragmatic randomized controlled trial. J Affect Disord 2018 Feb;227:455-462 [FREE Full text] [doi: 10.1016/j.jad.2017.11.021] [Medline: 29154168]
47. Klein JP, Berger T, Schröder J, Späth C, Meyer B, Caspar F, et al. Effects of a psychological internet intervention in the treatment of mild to moderate depressive symptoms: results of the EVIDENT study, a randomized controlled trial. Psychother Psychosom 2016;85(4):218-228 [FREE Full text] [doi: 10.1159/000445355] [Medline: 27230863]
48. Schuster R, Laireiter AR, Berger T, Moritz S, Meyer B, Hohagen F, et al. Immediate and long-term effectiveness of adding an internet intervention for depression to routine outpatient psychotherapy: subgroup analysis of the EVIDENT trial. J Affect Disord 2020 Sep 01;274:643-651 [FREE Full text] [doi: 10.1016/j.jad.2020.05.122] [Medline: 32663998]
49. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
50. Thase ME, Wright JH, Eells TD, Barrett MS, Wisniewski SR, Balasubramani GK, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. Am J Psychiatry 2018 Mar 01;175(3):242-250 [FREE Full text] [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
51. Thase ME, McCrone P, Barrett MS, Eells TD, Wisniewski SR, Balasubramani GK, et al. Improving cost-effectiveness and access to cognitive behavior therapy for depression: providing remote-ready, computer-assisted psychotherapy in times of crisis and beyond. Psychother Psychosom 2020 May 12;89(5):307-313 [FREE Full text] [doi: 10.1159/000508143] [Medline: 32396917]
52. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. Lancet 2007 Sep;370(9590):878-889 [FREE Full text] [doi: 10.1016/s0140-6736(07)61239-2]
53. Phillips EA, Himmler S, Schreyögg J. Preferences of psychotherapists for blended care in Germany: a discrete choice experiment. BMC Psychiatry 2022 Feb 12;22(1):112 [FREE Full text] [doi: 10.1186/s12888-022-03765-x] [Medline: 35151294]
54. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, E-COMPARED consortium. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017 Jun;8:1-9 [FREE Full text] [doi: 10.1016/j.invent.2017.01.001] [Medline: 30135823]
55. Kohls E, Baldofski S, Moeller R, Klemm SL, Rummel-Kluge C. Mental health, social and emotional well-being, and perceived burdens of university students during COVID-19 pandemic lockdown in Germany. Front Psychiatry 2021;12:643957 [FREE Full text] [doi: 10.3389/fpsyt.2021.643957] [Medline: 33889102]
56. Mata J, Wenz A, Rettig T, Reifenscheid M, Möhring K, Krieger U, et al. Health behaviors and mental health during the COVID-19 pandemic: a longitudinal population-based survey in Germany. Soc Sci Med 2021 Oct;287:114333 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114333] [Medline: 34455337]
57. Byrne A, Barber R, Lim CH. Impact of the COVID-19 pandemic – a mental health service perspective. Prog Neurol Psychiatry 2021;25(2):27-33b [FREE Full text] [doi: 10.1002/pnp.708]
58. The impact of COVID-19 on mental, neurological and substance use services. World Health Organization. 2020 Oct 5. URL: https://www.who.int/publications/i/item/978924012455 [accessed 2022-04-25]
59. Duffy ME, Twenge JM, Joiner TE. Trends in mood and anxiety symptoms and suicide-related outcomes among U.S. undergraduates, 2007-2018: evidence from two national surveys. J Adolesc Health 2019 Nov;65(5):590-598 [FREE Full text] [doi: 10.1016/j.jadohealth.2019.04.033] [Medline: 31279724]
60. Etzelmueller A, Radkovsky A, Hannig W, Berking M, Ebert DD. Patient's experience with blended video- and internet based cognitive behavioural therapy service in routine care. Internet Interv 2018 Jun;12:165-175 [FREE Full text] [doi: 10.1016/j.invent.2018.01.003] [Medline: 30135780]
61. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
62. Löwe B, Spitzer RL, Zipfel S, Herzog W. Gesundheitsfragebogen für Patienten (PHQ-D). Manual und Testunterlagen. (2. Ed.). Universitätsklinikum Heidelberg. New York, NY: Pfizer; 2002. URL: https://www.klinikum.uni-heidelberg.de/fileadmin/Psychosomatische_Klinik/download/PHQ_Manual1.pdf [accessed 2022-03-01]
63. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274 [FREE Full text] [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
64. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
65. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002 Sep;32(9):509-515 [FREE Full text] [doi: 10.3928/0048-5713-20020901-06]
66. Faßbinder E, Klein JP, Sipos V, Schweiger U. Therapie-Tools Depression: Mit E-Book inside und Arbeitsmaterial. Weinheim, Germany: Beltz; Oct 5, 2015.
67. Hautzinger M. Kognitive Verhaltenstherapie bei Depressionen: Mit Online-Materialien. Weinheim, Germany: Beltz; Apr 14, 2021.

68. Neudeck P, Mühlig S. Therapie-Tools Verhaltenstherapie: Therapieplanung, Probatorik, Verhaltensanalyse. Mit E-Book inside und Arbeitsmaterial (Beltz Therapie-Tools). Weinheim, Germany: Beltz; Mar 11, 2020.
69. Schaub A, Roth E, Goldmann U. Kognitiv-psychoedukative Therapie zur Bewältigung von Depressionen: Ein Therapiemanual. Göttingen, Germany: Hogrefe Verlag; Aug 1, 2006.
70. Beck AT. Cognitive models of depression. J Cognit Psychother 1987;1(1):5-37 [FREE Full text]
71. Ankuta GY, Abeles N. Client satisfaction, clinical significance, and meaningful change in psychotherapy. Prof Psychol Res Pr 1993 Feb;24(1):70-74 [FREE Full text] [doi: 10.1037/0735-7028.24.1.70]
72. Jerusalem M. Persönliche Ressourcen. Vulnerabilität und Streßerleben. Göttingen, Germany: Hogrefe Verlag; Jan 1, 1990.
73. Kolovos S, Kleiboer A, Cuijpers P. Effect of psychotherapy for depression on quality of life: meta-analysis. Br J Psychiatry 2016 Dec;209(6):460-468 [FREE Full text] [doi: 10.1192/bjp.bp.115.175059] [Medline: 27539296]
74. Rytsälä HJ, Melartin TK, Leskelä US, Lestelä-Mielonen PS, Sokero TP, Isometsä ET. Determinants of functional disability and social adjustment in major depressive disorder: a prospective study. J Nerv Ment Dis 2006 Aug;194(8):570-576 [FREE Full text] [doi: 10.1097/01.nmd.0000230394.21345.c4] [Medline: 16909064]
75. Zahra D, Qureshi A, Henley W, Taylor R, Quinn C, Pooler J, et al. The work and social adjustment scale: reliability, sensitivity and value. Int J Psychiatry Clin Pract 2014 Jun;18(2):131-138 [FREE Full text] [doi: 10.3109/13651501.2014.894072] [Medline: 24527886]
76. Aina Y, Susman JL. Understanding comorbidity with depression and anxiety disorders. J Am Osteopath Assoc 2006 May;106(5 Suppl 2):S9-14. [Medline: 16738013]
77. Hirschfeld RM. The comorbidity of major depression and anxiety disorders: recognition and management in primary care. Prim Care Companion J Clin Psychiatry 2001 Dec;3(6):244-254 [FREE Full text] [doi: 10.4088/pcc.v03n0609] [Medline: 15014592]
78. Beck AT, Steer RA, Brown G. Manual for the Beck Depression Inventory-II. Psychological Corporation. 1996. URL: https://www.nctsn.org/measures/beck-depression-inventory-second-edition [accessed 2022-03-01]
79. Kühner C, Bürger C, Keller F, Hautzinger M. [Reliability and validity of the Revised Beck Depression Inventory (BDI-II). Results from German samples]. Nervenarzt 2007 Jun;78(6):651-656 [FREE Full text] [doi: 10.1007/s00115-006-2098-7] [Medline: 16832698]
80. Dolle K, Schulte-Körne G, O'Leary AM, von Hofacker N, Izat Y, Allgaier AK. The Beck depression inventory-II in adolescent mental health patients: cut-off scores for detecting depression and rating severity. Psychiatry Res 2012 Dec 30;200(2-3):843-848 [FREE Full text] [doi: 10.1016/j.psychres.2012.05.011] [Medline: 22657953]
81. Hinz A, Mehnert A, Kocalevent RD, Brähler E, Forkmann T, Singer S, et al. Assessment of depression severity with the PHQ-9 in cancer patients and in the general population. BMC Psychiatry 2016 Feb 02;16:22 [FREE Full text] [doi: 10.1186/s12888-016-0728-6] [Medline: 26831145]
82. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988;56(6):893-897 [FREE Full text] [doi: 10.1037/0022-006x.56.6.893]
83. Hoyer J, Margraf J. Fragebögen und Ratingskalen zur globalen Messung von Angst, Ängstlichkeit und Angststörungen. In: Angstdiagnostik. Berlin, Germany: Springer; 2003.
84. Geissner E, Huetteroth A. [beck anxiety inventory German version - a reliable, valid, patientfriendly instrument for measuring clinical anxiety]. Psychother Psychosom Med Psychol 2018 Mar;68(3-4):118-125 [FREE Full text] [doi: 10.1055/s-0043-122941] [Medline: 29351711]
85. Angermeyer MC, Kilian R, Matschinger H. WHOQOL - 100 und WHOQOL - BREF: Handbuch für die deutschsprachige Version der WHO Instrumente zur Erfassung von Lebensqualität. Göttingen, Germany: Hogrefe; 2000.
86. World Health Organization. Division of Mental Health. WHOQOL-BREF : introduction, administration, scoring and generic version of the assessment : field trial version, December 1996. World Health Organization. 1996. URL: https://apps.who.int/iris/handle/10665/63529 [accessed 2022-03-01]
87. Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A Report from the WHOQOL Group. Qual Life Res 2004 Mar;13(2):299-310 [FREE Full text] [doi: 10.1023/b:qure.0000018486.91360.00]
88. Heissel A, Bollmann J, Kangas M, Abula K, Rapp M, Sanchez A. Validation of the German version of the work and social adjustment scale in a sample of depressed patients. Research Square. Preprint posted online January 29, 2021 2023 [FREE Full text] [doi: 10.21203/rs.3.rs-151437/v1]
89. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry 2002 May;180:461-464 [FREE Full text] [doi: 10.1192/bjp.180.5.461] [Medline: 11983645]
90. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. In: Weinman J, Wright S, Johnston M, editors. Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs. Windsor, UK: NFER-Nelson; 1995.
91. Schwarzer R, Jerusalem M. Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin, Germany: Freie Universität Berlin; Jan 1999.
92. Schwarzer R, Mueller J, Greenglass E. Assessment of perceived general self-efficacy on the internet: data collection in cyberspace. Anxiety Stress Coping 1999 Jan;12(2):145-161 [FREE Full text] [doi: 10.1080/10615809908248327]

93. Lalk C. Online-based self-help intervention for GAD. OSF Home. 2021 Dec 15. URL: https://osf.io/tj4re/ [accessed 2023-08-03]
94. O'Connor M, Casey L. The Mental Health Literacy Scale (MHLS): a new scale-based measure of mental health literacy. Psychiatry Res 2015 Sep 30;229(1-2):511-516 [FREE Full text] [doi: 10.1016/j.psychres.2015.05.064] [Medline: 26228163]
95. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plan 1982;5(3):233-237 [FREE Full text] [doi: 10.1016/0149-7189(82)90074-x] [Medline: 10259963]
96. Schmidt J, Lamprecht F, Wittmann WW. [Satisfaction with inpatient management. Development of a questionnaire and initial validity studies]. Psychother Psychosom Med Psychol 1989 Jul;39(7):248-255. [Medline: 2762479]
97. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and validity of assessing user satisfaction with web-based health interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
98. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
99. Gao M, Kortum P, Oswald FL. Multi-language toolkit for the system usability scale. Int J Human Comput Interact 2020 Aug 19;36(20):1883-1901 [FREE Full text] [doi: 10.1080/10447318.2020.1801173]
100. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. In J Human Comput Interact 2008 Jul 30;24(6):574-594 [FREE Full text] [doi: 10.1080/10447310802205776]
101. Brooke J. SUS: a retrospective. J Usab Stud 2013 Jan;8(2):40.
102. Posternak MA, Miller I. Untreated short-term course of major depression: a meta-analysis of outcomes from studies using wait-list control groups. J Affect Disord 2001 Oct;66(2-3):139-146 [FREE Full text] [doi: 10.1016/s0165-0327(00)00304-9] [Medline: 11578666]
103. Steinert C, Stadter K, Stark R, Leichsenring F. The effects of waiting for treatment: a meta-analysis of waitlist control groups in randomized controlled trials for social anxiety disorder. Clin Psychol Psychother 2017 May;24(3):649-660 [FREE Full text] [doi: 10.1002/cpp.2032] [Medline: 27445199]
104. Wakefield S, Kellett S, Simmonds-Buckley M, Stockton D, Bradbury A, Delgadillo J. Improving Access to Psychological Therapies (IAPT) in the United Kingdom: a systematic review and meta-analysis of 10-years of practice-based evidence. Br J Clin Psychol 2021 Mar 23;60(1):1-37 [FREE Full text] [doi: 10.1111/bjc.12259] [Medline: 32578231]
105. Romijn G, Batelaan N, Koning J, van Balkom A, de Leeuw A, Benning F, et al. Acceptability, effectiveness and cost-effectiveness of blended cognitive-behavioural therapy (bCBT) versus face-to-face CBT (ftfCBT) for anxiety disorders in specialised mental health care: a 15-week randomised controlled trial with 1-year follow-up. PLoS One 2021 Nov 12;16(11):e0259493 [FREE Full text] [doi: 10.1371/journal.pone.0259493] [Medline: 34767575]
106. Hendriks SM, Spijker J, Licht CM, Hardeveld F, de Graaf R, Batelaan NM, et al. Long-term work disability and absenteeism in anxiety and depressive disorders. J Affect Disord 2015 Jun 01;178:121-130 [FREE Full text] [doi: 10.1016/j.jad.2015.03.004] [Medline: 25805404]
107. Kennedy BL, Lin Y, Schwab JJ. Work, social, and family disabilities of subjects with anxiety and depression. South Med J 2002;95(12):1424-1427 [FREE Full text] [doi: 10.1097/00007611-200295120-00014]
108. Kooistra LC, Ruwaard J, Wiersma JE, van Oppen P, van der Vaart R, van Gemert-Pijnen JE, et al. Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. Internet Interv 2016 May;4:61-71 [FREE Full text] [doi: 10.1016/j.invent.2016.01.003] [Medline: 30135791]
109. Braun P, Atik E, Guthardt L, Apolinário-Hagen J, Schückes M. Barriers to and facilitators of a blended cognitive behavioral therapy program for depression and anxiety based on experiences of university students: qualitative interview study. JMIR Form Res 2023 Apr 12;7:e45970 [FREE Full text] [doi: 10.2196/45970] [Medline: 37043272]

Abbreviations

BAI: Beck Anxiety Inventory
bCBT: blended cognitive behavioral therapy
BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
CSQ-8: Client Satisfaction Questionnaire
GAD-7: Generalized Anxiety Disorder-7
GSE: General Self-Efficacy Scale
iCBT: internet-based cognitive behavioral therapy
ICD-10: International Classification of Diseases, 10th Revision
MHLS: Mental Health Literacy Scale
PHQ-9: Patient Health Questionnaire-9
SUS: System Usability Scale
WSAS: Work and Social Adjustment Scale

Edited by J Torous; submitted 01.12.22; peer-reviewed by I Choi, M Subotic-Kerry, A AL-Asadi; comments to author 26.04.23; revised version received 25.05.23; accepted 25.06.23; published 25.08.23.

Please cite as:
Atik E, Stricker J, Schückes M, Pittig A
Efficacy of a Brief Blended Cognitive Behavioral Therapy Program for the Treatment of Depression and Anxiety in University Students: Uncontrolled Intervention Study
JMIR Ment Health 2023;10:e44742
URL: https://mental.jmir.org/2023/1/e44742
doi:10.2196/44742
PMID:37624631

Original Paper

Evaluating the Modified Patient Health Questionnaire-2 and Insomnia Severity Index-2 for Daily Digital Screening of Depression and Insomnia: Validation Study

Jae Won Oh[1], PhD; Sun Mi Kim[2], PhD; Deokjong Lee[1,3], MD, PhD; Nak-Hoon Son[4], PhD; Jinsun Uh[5], BA; Ju Hong Yoon[6], PhD; Yukyung Choi[7], PhD; San Lee[1,3], MD, MPH

[1]Department of Psychiatry, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Republic of Korea

[2]Department of Nursing, Andong Science College, Andong, Republic of Korea

[3]Department of Psychiatry and the Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[4]Department of Statistics, Keimyung University, Daegu, Republic of Korea

[5]Mobigen Co, Ltd, Seongnam, Republic of Korea

[6]Korea Electronics Technology Institute, Seongnam, Republic of Korea

[7]Department of Intelligent Mechatronics Engineering, Sejong University, Seoul, Republic of Korea

Corresponding Author:
San Lee, MD, MPH
Department of Psychiatry
Yongin Severance Hospital, Yonsei University College of Medicine
363 Dongbaekjukjeondaero, Jung-dong, Giheung-gu
Yongin, 16995
Republic of Korea
Phone: 82 031 5189 8531
Fax: 82 031 5189 8565
Email: sanlee@yonsei.ac.kr

Abstract

Background: The Patient Health Questionnaire-2 (PHQ-2) and Insomnia Severity Index-2 (ISI-2) are screening assessments that reflect the past 2-week experience of depression and insomnia, respectively. Retrospective assessment has been associated with reduced accuracy owing to recall bias.

Objective: This study aimed to increase the reliability of responses by validating the use of the PHQ-2 and ISI-2 for daily screening.

Methods: A total of 167 outpatients from the psychiatric department at the Yongin Severance Hospital participated in this study, of which 63 (37.7%) were male and 104 (62.3%) were female with a mean age of 35.1 (SD 12.1) years. Participants used a mobile app (“Mental Protector”) for 4 weeks and rated their depressive and insomnia symptoms daily on the modified PHQ-2 and ISI-2 scales. The validation assessments were conducted in 2 blocks, each with a fortnight response from the participants. The modified version of the PHQ-2 was evaluated against the conventional scales of the Patient Health Questionnaire-9 and the Korean version of the Center for Epidemiologic Studies Depression Scale–Revised.

Results: According to the sensitivity and specificity analyses, an average score of 3.29 on the modified PHQ-2 was considered valid for screening for depressive symptoms. Similarly, the ISI-2 was evaluated against the conventional scale, Insomnia Severity Index, and a mean score of 3.50 was determined to be a valid threshold for insomnia symptoms when rated daily.

Conclusions: This study is one of the first to propose a daily digital screening measure for depression and insomnia delivered through a mobile app. The modified PHQ-2 and ISI-2 were strong candidates for daily screening of depression and insomnia, respectively.

(JMIR Ment Health 2023;10:e45543) doi:10.2196/45543

KEYWORDS

Patient Health Questionnaire-2; PHQ-2; Insomnia Severity Index; ISI-2; depression; insomnia; mobile health; mobile phone

Introduction

Background

Depression is the leading cause of mental illness in the general population and in primary care, instigating a substantial global health-related burden [1]. It is characterized by sadness, the loss of interest, tiredness, poor concentration, feelings of guilt, and disturbed sleep or appetite. In severe cases, depression can be associated with chronic diseases, high use of health care services, suicide, shortened life expectancy, and reduced quality of life [2].

Insomnia (a common comorbid condition with depression) is associated with difficulties in initiating or maintaining sleep, and its prevalence related to a reduction in physical and mental health [3]. Clinical and epidemiological studies have highlighted that sleep disturbances are closely linked to major depression [4]. In addition, studies have reported a comparable prevalence of insomnia among 41% of patients diagnosed with depression, with additional research suggesting a potential bidirectional link between the 2 disorders [5].

Various methods have been used to assess and screen for the symptoms of depression and insomnia. Screening assessment does not confirm diagnosis; however, depending on the results, they might be used to determine the risk of the respondent and refer accordingly [6]. Therefore, it is essential that such measures are reliable and valid to ensure that the generated outcomes are clinically accurate [7]. Brief self-report measures have been developed and used to screen for common mental disorders, depression, and insomnia. A self-report measure of depression, Patient Health Questionnaire-9 (PHQ-9), which matches the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria of major depressive disorder, has been evaluated as the most reliable screening tool for depression through various systemic reviews and meta-analyses [6,8,9]. Likewise, the Insomnia Severity Index (ISI), a self-report measure of insomnia, has been recognized to effectively detect sleep problems [10,11].

However, based on the interest in fewer screening questions and the efficacy of such measures of depression and insomnia, further shortened versions of the PHQ-9 and ISI have been presented and validated in the literature. The Patient Health Questionnaire-2 (PHQ-2) uses the first 2 questions of the PHQ-9 concerning the frequency of depressed mood and anhedonia over the past 2 weeks [12]. This has demonstrated good sensitivity for detecting major depression [13,14]. Similarly, the Insomnia Severity Index-2 (ISI-2) is a 2-item version of the ISI that grades one's satisfaction with current sleep patterns and assesses the interference with daily functioning because of sleep patterns. This has been proposed as a strong candidate for a brief scale measuring insomnia, with good psychometric properties [15].

Despite the reliable and valid quality of brief self-report measures for screening, depression and insomnia have often been underdetected and undertreated within primary care settings and the general community [16,17]. Various factors, including low prioritization, stigma, and barriers to assessment, have been suggested as reasons for underreporting [18]. With the increase in the availability and adoption of smartphones, enhanced methods of using new digital technologies to access and engage health care treatment have been established globally [19-21]. This has introduced assessment methods and intervention delivery, as apps can be designed for self-assessment and assist patients in assessing and monitoring their symptoms [22]. The use of smartphone apps also allows for ecological momentary assessment (EMA), which increases the accuracy of self-report responses by reducing recall bias and preventing aggregation by measuring the key variables in natural environments and in real time [23,24]. Existing systematic reviews have identified that EMA methods for collecting data are effective and reasonable for measuring momentary mood and stress owing to the ubiquity and easy use of smartphones [25].

Objective

In accordance with the growth in the ability of EMA, this study aimed to address the existing barriers to assessment. As existing PHQ-2 and ISI-2 screening assessments are reflective of the experience from the previous 2 weeks, this study aimed to enhance the EMA by implementing the PHQ-2 and ISI-2 daily via a smartphone app. In addition, a psychometric assessment of the daily PHQ-2 and ISI-2 scales was conducted to evaluate the validity of smartphone-based daily screening measures. In particular, the assessment of PHQ-2 has been conducted twice daily, in the morning and afternoon, to control for the potential diurnal variations of depressive symptoms, which can fluctuate depending on the time [26].

Methods

Participants

Patients from the psychiatric outpatient clinic of the Yongin Severance Hospital in South Korea were recruited for this study. An offline advertisement was posted on bulletin boards near the psychiatry department. To be eligible, the participants must (1) be outpatients of psychiatric department (patients do not necessarily require a clinical diagnosis of depression or insomnia), (2) be aged ≥18 years, (3) provide consent for voluntary research participation, and (4) have an Android smartphone with access to Wi-Fi or 4G or 5G internet connectivity.

Ethics Approval

This study was approved by the Institutional Review Board of the Yongin Severance Hospital (institutional review board number 9-2020-0160) and conducted in accordance with the Declaration of Helsinki. All participants provided written informed consent before participation.

Materials

PHQ-9 Cutoff Score

The PHQ-9 is a 9-item questionnaire that assesses the severity of depression. The scores range from 0 to 27, with each question scored between 0 (not at all) and 3 (nearly every day). This scale is a reliable and valid measure for screening depression severity. A validation study of the Korean version of the PHQ-9, which reported a cutoff score of 10 for screening depression,

demonstrated 81.8% sensitivity and 89.9% specificity [27]. Henceforth, in this study, the Korean version of the PHQ-9, with a cutoff score of 10, was used to assess the level of depression within the participants.

Modified PHQ-2

The PHQ-2 is a reduced version of the existing PHQ-9 scale, which evaluates the frequency of depressive symptoms and anhedonia over the past fortnight, with each item scored between 0 (not at all) and 3 (nearly every day). The use of the PHQ-2 for screening depression has been previously validated with a threshold PHQ-2 score of ≥3, reporting a sensitivity of 83% and a specificity of 92% in detecting major depression [12]. A recent validation of the PHQ-2 has reported a more conservative approach of considering 2 as the threshold value, which has been presented in various studies [13,28]. This study used the same PHQ-2 scale items, which were "How often have you been bothered by the following? 1) Little interest or pleasure in doing things; 2) Feeling down, depressed or hopeless," and a modified response options for participants to rate their symptoms daily, with scores ranging from 0 to 4 (never, rarely, sometimes, often, and always). The number of PHQ-2 response options was modified to be in line with other daily scales and enhance participants' usability, considering that this was a daily measure over a period of 28 days.

Center for Epidemiologic Studies Depression Scale–Revised

The Center for Epidemiologic Studies Depression Scale–Revised is a revised version of the original Center for Epidemiologic Studies Depression, which is used to screen for depressive disorders. It is a 20-item measure that adequately assesses the fundamental symptoms of depression as described in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [29]. In this study, the validated Korean version of the Center for Epidemiologic Studies Depression Scale–Revised (K-CESD-R) was implemented, with a threshold of 13 to identify the level of depression [30].

ISI Scores

The ISI is a 7-item self-report questionnaire that assesses the nature, severity, and impact of insomnia. The factors evaluated were sleep onset, sleep maintenance and early morning awakening problems, sleep dissatisfaction, interference of sleep difficulties with daytime functioning, noticeability of sleep problems by others, and distress caused by sleep difficulties. The items are scored on a 5-point Likert scale ranging from 0 (no problem) to 4 (very severe problem). The total score is interpreted in 4 different categories: absence of insomnia (0-7), subthreshold insomnia (8-14), moderate insomnia (15-21), and severe insomnia (22-28). Scores >15 were considered to indicate insomnia symptoms. Studies have reported adequate psychometric properties of the current scale [16].

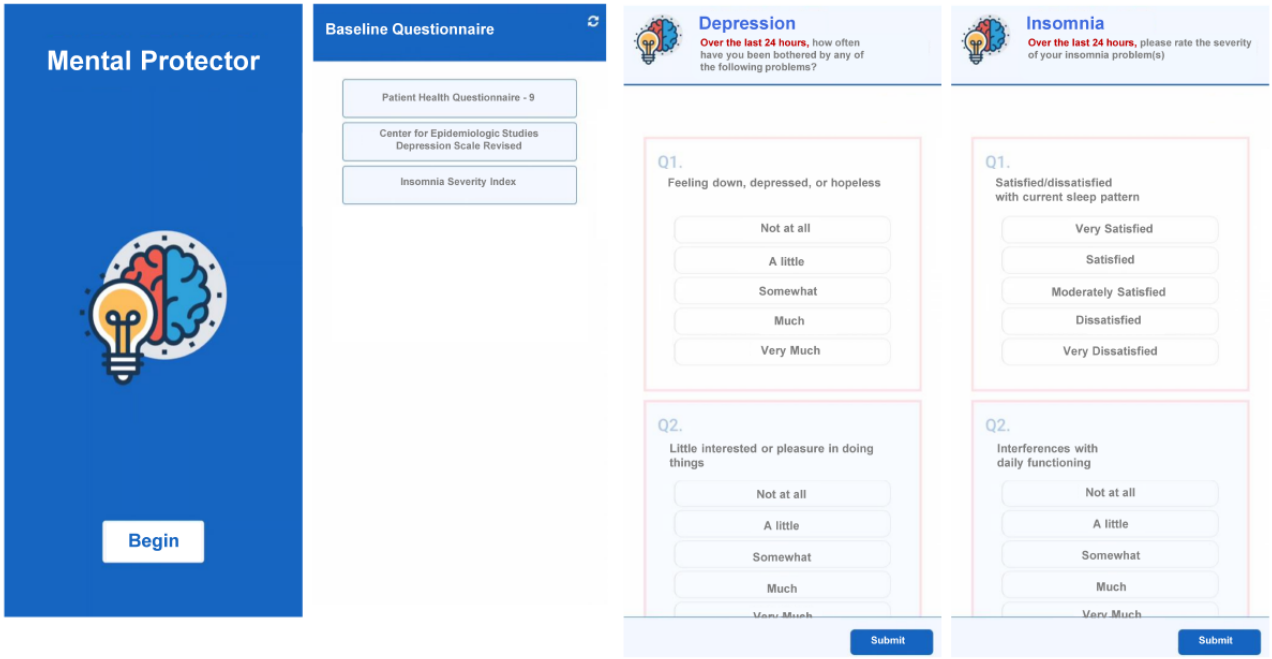
ISI-2 Scores

The ISI-2 is a 2-item version of the ISI scale, and 1 item asks participants' satisfaction with their current sleep pattern and another item represents the amount of interference with daily functioning because of their current sleep pattern. The total scores of the 2 items ranged from 0 to 8, and the scale has been validated with a cutoff of 6 points, which is considered appropriate to detect insomnia disorder with a sensitivity score of 84% and a specificity score of 76% [15]. This study implemented the ISI-2 scale items asking the participants to rate their symptoms daily, with scores ranging from 0 (not at all) to 4 (all the time).

Mental Protector Mobile App

The Mental Protector is a mobile app that comprises self-reporting depression- and insomnia-related scales designed to obtain data via an EMA [31]. Full scales of the K-CESD-R, PHQ-9, and ISI were administered at the baseline, midpoint (2 weeks after use), and end point (4 weeks after use). Modified PHQ-2 items were delivered daily to the participants through a push-alert method, once in the morning and once in the afternoon. Similarly, the ISI-2 scales were prompted to the participants, but only in the morning. The app was developed by a multidisciplinary team comprising a psychiatrist, researchers with expertise in eHealth delivery of interventions, and software programmers. The psychiatrist and researchers with experience in digital therapeutics provided advice on the study design and the protocol. On the basis of these guidelines, an IT firm named Mobigen Co developed the actual app for research. The initial language used was created in Korean, using the validated Korean version of the scales included. The screenshot image of the app used in this research has been translated into English and is presented in Figure 1.

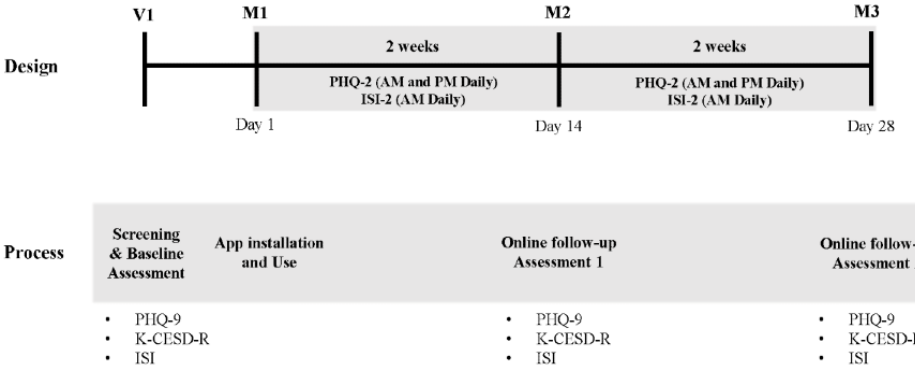
Figure 1. Screenshots of the Mental Protector app in the order of the main title page, baseline questionnaire page and pages for daily depression and insomnia rating, respectively.



Procedures

Participants were recruited from the Department of Psychiatry Outpatient Clinic at Yongin Severance Hospital. Eligible participants were introduced to the study by the research team, and informed consent was obtained. The participants were introduced to the mobile app (“Mental Protector”) that was installed on Android devices. In addition, participants’ basic sociodemographic information was obtained, and baseline assessments of the PHQ-9, K-CESD-R, and ISI were performed. These assessments were conducted under the supervision of the study’s researchers who provided appropriate instructions on site at the outpatient clinic. Once completed, the participants were explained about the daily prompt alerts delivered by the push function. A push alert for the modified PHQ-2 scale was sent twice daily: once in the morning and once in the late afternoon. Considering the differences in one’s depressive levels depending on the time of the day, the option to assess twice daily was decided [26]. Regarding ISI-2, the push alert was sent once in the morning, questioning the participants’ sleep satisfaction from the previous night. Furthermore, in cases where the participants did not respond to the daily screening items from the initial push alert, additional automated push alerts were sent up to 3 times in total to prompt participation. The study was conducted for 28 days from the day of the app installation. It has been designed in 2 blocks of 14-day assessments, as the conventional method of depression- and insomnia-screening assessments are performed with a reflection of the previous 14 days [32,33]. Hence, daily assessment is required over this period to determine the mean score for both the modified PHQ-2 and ISI-2. Daily assessments with 4 questions in the morning and 2 questions in the afternoon are estimated to take approximately 1 minute to complete, and follow-up assessments on days 14 and 28 are estimated to take 15 minutes. The details of the study design are presented in Figure 2.

Figure 2. Study protocol outlining the design and process involved in this study. ISI: Insomnia Severity Index; K-CESD-R: Korean version of Center for Epidemiologic Studies Depression Scale–Revised; M1: mobile use time point 1; M2: mobile use time point 2; M3: mobile use time point 3; PHQ-9: Patient Health Questionnaire-9; V1: visit 1.



Statistical Analysis

The validation assessments were conducted in 2 blocks, each with a fortnightly response from the participants. On the basis of the PHQ-9 and K-CESD-R, the sensitivity and specificity of the modified PHQ-2 were calculated, and receiver operating characteristic (ROC) curve analysis was performed to identify the most appropriate threshold score for the modified PHQ-2. The mean score of the modified PHQ-2 was calculated from the assessment conducted in the morning and afternoon over a period of 2 weeks. Similarly, the sensitivity and specificity of ISI-2 were calculated based on the initial ISI response, and ROC curve analysis was used to identify the most appropriate threshold score for ISI-2. The mean score of the modified ISI-2 was calculated from the scores of the assessments conducted in the morning for a fortnight. Participants required 100% completion rates for the key biweekly measures—PHQ-9, K-CESD-R, and ISI—and >60% completion rates for daily assessments of the modified PHQ-2 and ISI-2 to be considered for analysis. SAS software (version 9.4; SAS Institute) was used for the analyses, and statistical significance was set at $P<.05$.

Results

Baseline Characteristics

A total of 201 outpatients participated in this study. Of the 201 outpatients, participant withdrawals and missing responses were removed, leaving 167 (83.1%) outpatients eligible for analysis in this study, 63 (37.7%) male patients and 104 (62.3%) female patients, aged between 18 and 69 years (mean 35.1, SD 12.1 years). This included the respondents with >60% of the daily assessment completion rates. There were no missing cases in the first 14 days of the daily assessment, and missing cases occurred during the 14- to 28-day period. On average, participants had 14.7 (SD 2.4) years of education. Regarding living arrangement status, 63.5% (106/167) of the patients were living alone and 36.5% (61/167) of the patients lived with cohabitants in their households. A total of 53.3% (89/167) of participants were currently employed, whereas 46.7% (78/167) were unemployed. Of the 167 participants, 86 (51.5%) participants had primary diagnosis of depression and 81 (48.1%) participants had other diagnoses, which included 33 (41%) patients with anxiety disorders, 19 (23%) patients with trauma and stress-related disorders, 10 (12%) patients with bipolar disorder, 6 (7%) patients with obsessive and compulsive disorder, 3 (4%) patients with psychosis, 8 (10%) patients with insomnia, 1 (1%) patient with attention-deficit/hyperactivity disorder, and 1 (1%) patient with bulimia nervosa. The basic sociodemographic details of the participants and the baseline measures of depression and insomnia are presented in Table 1, and the detailed breakdown of those with other than depression as primary diagnosis is presented in Multimedia Appendix 1.

Table 1. Demographic characteristics of participants (N=167).

Characteristics	Male participants (n=63)	Female participants (n=104)
Primary diagnosis, n (%)		
Depression	30 (47.6)	56 (53.9)
Other	33 (52.4)	48 (46.1)
Age (years), n (%)		
18-25	33 (52.5)	19 (18.3)
26-35	12 (19)	23 (22.1)
36-45	12 (19)	32 (30.8)
>45	6 (9.5)	30 (28.8)
Education, n (%)		
High school and below	20 (31.7)	22 (21.1)
University	26 (41.3)	63 (60.6)
Graduate and above	17 (27)	19 (18.3)
Current employment status, n (%)		
Unemployed	31 (49.2)	47 (45.2)
Employed	32 (50.8)	57 (54.8)
Socioeconomic status, n (%)		
High	4 (6.3)	5 (4.8)
Mid	34 (54)	67 (65.1)
Low	25 (39.7)	31 (30.1)
Alcohol consumption, n (%)		
Nondrinker	37 (58.7)	67 (64.4)
Drinker	26 (41.3)	37 (35.6)
Smoking, n (%)		
Nonsmoker	38 (60.3)	71 (68.3)
Past smoker	8 (12.7)	14 (13.4)
Current smoker	16 (27)	19 (18.3)
Living arrangement, n (%)		
Alone	50 (79.4)	56 (53.8)
Together	13 (20.6)	48 (46.2)
Baseline measures, mean (SD)		
PHQ-9[a]	12.4 (6.4)	12.2 (7.0)
K-CESD-R[b]	30.6 (17.7)	32.8 (20.2)
ISI[c]	12.6 (6.6)	13.4 (6.2)

[a]PHQ-9: Patient Health Questionnaire-9.

[b]K-CESD-R: Korean version of the Center for Epidemiologic Studies Depression Scale–Revised.

[c]ISI: Insomnia Severity Index.

Reliability Analysis of Modified PHQ-2

Internal consistency assessment of the modified PHQ-2 scale was conducted at 2 time points: after week 2 and at the end of week 4. Correlational analyses were conducted with the PHQ-2 responses and the responses of the PHQ-9 and K-CESD-R assessed at the midpoint and end of week 2. This process was repeated to assess the internal consistency of the modified PHQ-2 morning and evening scores at weeks 3 and 4, with the PHQ-9 and K-CESD-R scores at the end of week 4. The correlation matrix with Pearson *r* values is presented in Table 2.

Table 2. Correlation matrix of depression with Patient Health Questionnaire-2 (PHQ-2).

	AM PHQ-2[a]	PM PHQ-2[b]	Mid–PHQ-9[c,d]	Mid–K-CESD-R[e,f]	End PHQ-9[g]	End K-CESD-R[h]
Weeks 1-2 (*r* value)						
AM PHQ-2						
r	1.00	0.95	0.79	0.78	—[i]	—
P value	—	<.001	<.001	<.001	—	—
PM PHQ- 2						
r	0.95	1.00	0.80	0.78	—	—
P value	<.001	—	<.001	<.001	—	—
Mid–PHQ-9						
r	0.79	0.80	1.00	0.91	—	—
P value	<.001	<.001	—	<.001	—	—
Mid–K-CESD-R						
r	0.78	0.78	0.91	1.00	—	—
P value	<.001	<.001	<.001	—	—	—
Weeks 3-4 (*r* value)						
AM PHQ-2						
r	1.00	0.97	0.81	0.79	—	—
P value	—	<.001	<.001	<.001	—	—
PM PHQ-2						
r	0.97	1.00	0.79	0.76	—	—
P value	<.001	—	<.001	<.001	—	—
End PHQ-9[h]						
r	0.81	0.79	—	—	1.00	0.93
P value	<.001	<.001	—	—	—	<.001
End K-CESD-R[i]						
r	0.79	0.76	—	—	0.93	1.00
P value	<.001	<.001	—	—	<.001	—

[a]Morning PHQ-2 assessment.

[b]Afternoon PHQ-2.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]PHQ-9 assessment at week 2.

[e]K-CESD-R: Korean version of the Center for Epidemiologic Studies Depression Scale–Revised.

[f]K-CESD-R assessment at week 2.

[g]PHQ-9 assessment at week 4.

[h]K-CESD-R assessment at week 4.

[i]Not applicable.

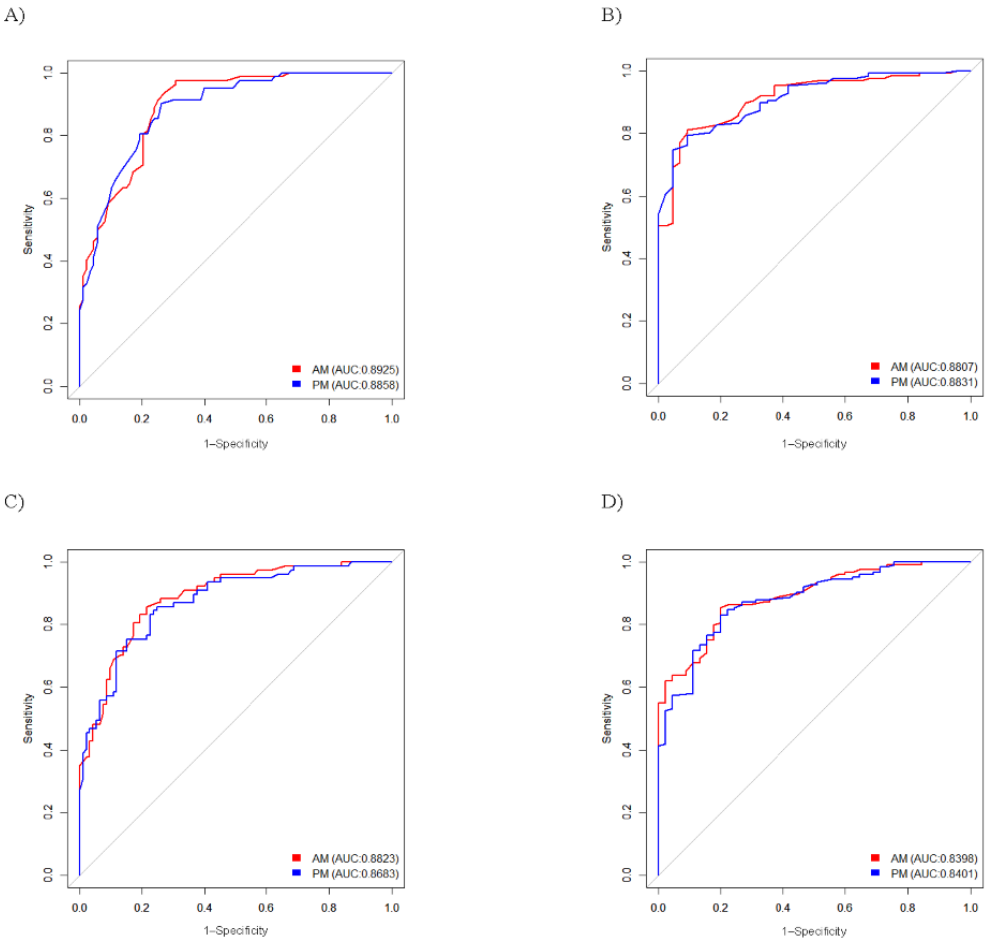
Cronbach α assessment of the modified PHQ-2 at week 2 presented a value of .94 and .94 for PHQ-9 and K-CESD-R, respectively, at midpoint (both *P*<.001) and .94 and .95 for PHQ-9 and K-CESD-R, respectively, at end point of week 4 (both *P*<.001), both identifying a high reliability for modified PHQ-2 measures in comparison with PHQ-9 and K-CESD-R. The tests showed a significant correlation between each depression scale score.

Test-retest reliability of the modified PHQ-2 scale was assessed by comparing the mean morning scores of the modified PHQ-2 between the first 2 weeks and the latter 2 weeks. This was assessed using the evening scores of the modified PHQ-2. The Pearson coefficient scores reported good test-retest reliability for reporting 0.89 and 0.90, respectively, with significant *P* values (both *P*<.001).

Area Under the Curve and Cutoff Point of the Modified PHQ-2

The ROC curve evaluated the sensitivity and specificity of the modified PHQ-2 rating scale. The modified PHQ-2 daily rating scales were assessed twice with the PHQ-9 and twice with the K-CESD-R, each at 2 time points: once at the midpoint of week 2 and at the end of week 4. According to the area under the curve (AUC) values, there was no significant difference between the modified PHQ-2 screening conducted in the morning and afternoon. Thus, the ability to detect depression did not differ depending on the time of the PHQ-2 assessment. The details of the ROC assessment, including the AUC values, are shown in Figure 3.

Figure 3. Receiver operating characteristics (ROC) curve for the modified Patient Health Questionnaire-2 (PHQ-2) daily rating scale, with the comparison of modified PHQ-2 morning (AM) and afternoon (PM) responses. They are displayed as follows: A) PHQ-9 at week 2; B) Korean version of Center for Epidemiologic Studies Depression Scale–Revised (K-CESD-R) at week 2; C) PHQ-9 at week 4; D) K-CESD-R at week 4. AUC: area under the curve.



The sensitivity and specificity values corresponding to the modified PHQ-2 cutoff scores are presented in Table 3. Here, a score of ≥10 for the PHQ-9 and 13 for the K-CESD-R detected depression. On the basis of ordering of cutoff values of all 8 assessments, including morning and afternoon at week 2 and week 4, and in comparison with the PHQ-9 and K-CESD-R scales, Youden J (YJ) statistics were used to determine the optimal threshold value. The YJ index summarizes the ROC curve by assessing the effectiveness of the diagnostic mark and enabling the selection of an optimal cutoff value [34]. Table 3 shows that 3.29 was most frequently identified cutoff value that had the highest YJ value and, therefore, was considered the acceptable cutoff value for daily screening of depression. Initially, 15 participants with low adherence rates, below 60%, were removed. The overall completion rate of the included 167 participants was 94.8%. Those with missing responses were from the second fortnight of the study; therefore, it is anticipated that the accuracy of the data from weeks 3 and 4 might have implications. Consequently, when establishing the optimal cutoff value, the preference for the results from the first 2 weeks was considered.

Table 3. Cutoff value of Patient Health Questionnaire-2 (PHQ-2) compared with Patient Health Questionnaire-9 (PHQ-9) and the Korean version of Center for Epidemiologic Studies Depression Scale–Revised (K-CESD-R).

Scale and cutoff value	Sensitivity	Specificity	YJ[a]
Week 2			
PHQ-2, AM[b], compared with PHQ-9			
3.14	0.98	0.69	0.67
3.29	0.94	0.73	0.67
3.36	0.91	0.75	0.66
3.21	0.96	0.69	0.65
3.43	0.89	0.76	0.65
PHQ-2, PM[c], compared with PHQ-9			
3.29	0.90	0.75	0.65
3.21	0.91	0.71	0.63
3.43	0.85	0.77	0.63
3.50	0.84	0.79	0.63
3.71	0.80	0.82	0.63
PHQ-2, AM, compared with K-CESD-R			
3.29	0.80	0.83	0.63
3.14	0.83	0.79	0.62
3.36	0.77	0.85	0.62
3.21	0.82	0.79	0.61
3.07	0.84	0.77	0.61
PHQ-2, PM , compared with K-CESD-R			
3.29	0.79	0.90	0.69
3.36	0.75	0.90	0.65
3.21	0.80	0.85	0.64
3.43	0.74	0.90	0.64
3.50	0.72	0.90	0.62
Week 4			
PHQ-2, AM , compared with PHQ-9			
3.67	0.86	0.78	0.63
3.79	0.83	0.80	0.63
3.91	0.80	0.82	0.62
3.67	0.84	0.78	0.62
3.78	0.83	0.79	0.62
PHQ-2, PM , compared with PHQ-9			
3.71	0.86	0.74	0.60
3.78	0.84	0.76	0.60
3.86	0.83	0.77	0.60
4.08	0.75	0.84	0.59
3.68	0.86	0.73	0.59
PHQ-2, AM , compared with K-CESD-R			
3.86	0.68	0.98	0.66

Scale and cutoff value	Sensitivity	Specificity	YJ[a]
3.79	0.70	0.96	0.66
3.09	0.85	0.81	0.66
3.00	0.87	0.79	0.66
3.91	0.67	0.98	0.65
PHQ-2, PM , compared with K-CESD-R			
3.45	0.78	0.91	0.68
3.41	0.80	0.89	0.68
3.57	0.77	0.91	0.68
3.43	0.79	0.89	0.67
3.21	0.82	0.85	0.67

[a]YJ: Youden J statistic value.

[b]PHQ-2 assessed in the morning.

[c]PHQ-2 assessed in the afternoon.

Reliability Assessment of ISI-2

In addition to the modified PHQ-2, an internal consistency assessment of the ISI-2 scale was also conducted, first after 2 weeks of use and then again at the end of week 4. Correlational analyses of the ISI-2 responses for the first 2 weeks were compared, with the ISI responses assessed at the midpoint and at the end of week 2. This process was repeated to assess the internal consistency of the ISI-2 at weeks 3 and 4 with the ISI scores at the end of week 4, the end point. Pearson *r* for the first 2 weeks was 0.78 and 0.81 for the last 2 weeks; Cronbach α was .88 and .89, respectively.

The test-retest reliability of the ISI-2 scale was also assessed by comparing the mean scores of the ISI-2 in the first 2 weeks with the latter 2 weeks. The Pearson coefficient scores reported good test-retest reliability, reporting 0.88 with a significant *P* value ($P<.001$).

AUC and Cutoff Point of ISI-2

The ROC curve was used to evaluate the sensitivity and specificity of the daily ISI-2 rating scale. Each scale evaluation was performed twice, each at 2 time points, once at the midpoint of week 2 and the other at the end point of week 4. The AUC value was 0.85 and 0.88, respectively, for midpoint and end point assessment. The sensitivity and specificity values corresponding to various ISI-2 cutoff scores are presented in Multimedia Appendix 2. On the basis of assessments and use of the YJ index for the PHQ-2, a threshold value of 3.50 was considered acceptable to screen for insomnia when using the ISI-2. This cutoff value was consistent at the midpoint and end point, with the highest overall YJ score based on sensitivity and specificity analyses.

Discussion

Principal Findings

Traditional clinical assessments have relied heavily on retrospective questionnaires, asking respondents to recall their experiences and summarizing their responses to questions [35]. Previous research has shown that individuals with clinical depression show greater inaccuracy in their recall, which can have considerable clinical implications for the patient and their treatment experience [36]. Thus, this study proposed a method of digital daily screening to reduce any recall bias and observe the severity and frequency of conditions over time.

Using brief measures of depression and insomnia, namely, the modified PHQ-2 and ISI-2, respectively, this study presented a daily screening process to increase the accuracy of the EMA and reduce any recall bias. In addition, to justify that these digital daily screening scales are comparable with the existing measures, a psychometric validation of daily rating scales of the modified PHQ-2 and ISI-2 was conducted. Considering the ubiquitous nature of smartphone devices [37], these screening measures were delivered via smartphone apps. To control the possible variations in depressed mood depending on the time of day [26], screening assessments were conducted once daily in the morning and again in the afternoon for the depression-screening scale of the modified PHQ-2. Sleep quality was assessed once daily. The study was conducted with outpatient psychiatric patients, thus comprising a clinical population with depressive symptoms as their primary diagnosis or with other mental health conditions.

As per the study design, daily assessments were considered more accurate for screening of depressive symptoms than asking participants to reflect on the previous fortnight. From the psychometric evaluation of the modified PHQ-2 scale, the ROC curve findings reported no difference between the morning and afternoon responses of the modified PHQ-2 when assessed, in contrast to the traditional scales of the PHQ-9 and K-CESD-R. Unlike the anticipated diurnal variation, there was no significant disparity, suggesting that the use of the modified PHQ-2 was a viable measure for screening at any time of the day. Moreover, based on the reliability evaluations, the modified PHQ-2 scale showed good psychometric properties, demonstrating that it is a reliable screening measure for depression.

With the adaptation of the PHQ-2 as a daily screening measure, this study identified the mean cutoff score of the modified

PHQ-2 when assessed for a period of 2 weeks. When assessed in comparison with both the PHQ-9 and K-CESD-R scales, for both morning and afternoon assessments of the modified PHQ-2, a mean value of 3.29 over a fortnight was considered appropriate to determine users with depressive symptoms. Existing literature has considered that PHQ-2 score of 2 or higher out of a total score of 6 to potentially screen for depression [13,28]. As this study validated a modified version and the modified scale had an additional response option, a cutoff value of 3.29 from a total score of 8 can be considered comparable with the scores derived from the initial validation studies.

The ISI, an insomnia-screening scale, has been developed into a shorter version of the ISI-2 in which users reflect on their sleep-related experiences from the previous 2 weeks. Similar to the PHQ-2 scale, this study proposed an alternative of rating the ISI-2 daily for a fortnight to screen for insomnia severity. This would assist in determining the severity and frequency of the symptoms across a 14-day period, providing a more precise clinical judgment for patients. Therefore, a psychometric assessment of the daily ISI-2 measures was conducted to evaluate its appropriateness for screening insomnia. Overall, the results demonstrated good internal consistency and test-retest reliability. When implementing ISI-2 as a daily screening measure for insomnia, an optimal cutoff score of 3.50 was presented according to sensitivity and specificity values. Unlike the PHQ-2 validation, the ISI-2 had a considerable difference in its cutoff value with the existing validation research. A previous study demonstrated that individuals with insomnia have sleep-related attention bias, which can influence the misperception of their sleep and the extent of their impairment in daytime activities [38]. Hence, owing to the risk of attention and recall bias, assessing insomnia and daytime impairment symptoms daily might create the possibility of a conservative approach, even with lower scores potentially indicating the risk of insomnia, unlike prior assessments that rely on a weekly or fortnightly recall.

Strengths and Limitations

This study had important strengths. Before this study, there had been no consideration of using the PHQ-2 and ISI-2 to screen for depression and insomnia daily. This was the first study to propose the daily use of these scales with detailed psychometric evaluations. On the basis of this study's findings, both the PHQ-2 and ISI-2 can be used daily with the use of smartphone technology to screen for depression and insomnia, implementing the mean cutoff values provided. This will benefit clinicians and researchers by improving the reliability of patient responses and reducing recall bias. In addition, the use of daily scoring of the PHQ-2 and ISI-2 will provide the benefit of being able to observe daily changes in users' ratings of their depressive and insomnia symptoms using smartphone apps.

Regardless of the strengths identified, this study had some limitations. The validation process of the depression-screening scale, PHQ-2, involved the assessment and evaluation of 2 robust scales: PHQ-9 and K-CESD-R. Nonetheless, the PHQ-2 is a shortened version of the PHQ-9 and thus includes items that overlap with the scales. This means that an objective comparison was only performed using the K-CESD-R. Additional comparisons with other depression-screening scales could have enforced the validity of the new scale. Similarly, the ISI-2 is a shortened version of the original ISI. Including additional well-renowned scales such as the Pittsburgh Sleep Quality Index in the validation process would have strengthened the process. Furthermore, this study focused only on the recruitment of a clinical population. Further research in the general population may be necessary to validate these scales for general use. In addition, this study validated the daily rating scale by comparing daily measures with those of the conventional retrospective scale. Further use of daily assessments may provide opportunities for diverse validation studies on daily rating measures. Finally, although there were no missing cases in the first fortnight of the research period, the study included cases with missing responses between days 14 and 28. This would have influenced the validity and robustness of the process, which was addressed by identifying the cutoff values primarily from the data of the first fortnight. Nonetheless, there were no missing data during the first 14 days of the study, and the data collected during this period were considered more reliable and accurate. As a result, the preference for the results regarding the modified PHQ-2 cutoff values was determined from the first 2 weeks of assessment. Having no missing data in the first 14 days potentially demonstrates the importance of brief and easy user interface to retrain user participation.

Conclusions

This study implemented the use of the modified PHQ-2 and ISI-2 and successfully validated these as daily digital screening scales, providing cutoff scores to be used to screen for depression and insomnia symptoms measured daily over a period of 14 days. This is in accordance with the existing assessment of depression and insomnia symptoms, which is conducted as a reflection of the previous fortnight [32,33]. Therefore, this study provides a new alternative for such screening measures and demonstrates the efficacy of using smartphone apps to obtain such momentary data. Medical practitioners may implement the proposed modified PHQ-2 and ISI-2 daily screening assessments for depression and insomnia to accurately assess the patients daily and use the mean values of both measures to present clinical findings that are equivalent to the conventional screening measures used to assess depressive and insomnia symptoms.

Acknowledgments

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation grant funded by the Korean government (Ministry of Science and Information and Communication Technology; 2020-0-01969).

In addition, this study was supported by a faculty research grant from the Yonsei University College of Medicine (6-2021-0067).

Data Availability

The data collected and used in this meta-analysis can be requested from the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participants with other diagnosis.
[DOCX File , 16 KB - mental_v10i1e45543_app1.docx]

Multimedia Appendix 2
Cutoff scores of Insomnia Severity Index-2.
[DOCX File , 16 KB - mental_v10i1e45543_app2.docx]

References

1. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of depression in the community from 30 countries between 1994 and 2014. Sci Rep 2018 Feb 12;8(1):2861 [FREE Full text] [doi: 10.1038/s41598-018-21243-x] [Medline: 29434331]
2. Li H, Ge S, Greene B, Dunbar-Jacob J. Depression in the context of chronic diseases in the United States and China. Int J Nurs Sci 2019 Jan 10;6(1):117-122 [FREE Full text] [doi: 10.1016/j.ijnss.2018.11.007] [Medline: 31406877]
3. Li L, Wu C, Gan Y, Qu X, Lu Z. Insomnia and the risk of depression: a meta-analysis of prospective cohort studies. BMC Psychiatry 2016 Nov 05;16(1):375 [FREE Full text] [doi: 10.1186/s12888-016-1075-3] [Medline: 27816065]
4. Staner L. Comorbidity of insomnia and depression. Sleep Med Rev 2010 Feb;14(1):35-46. [doi: 10.1016/j.smrv.2009.09.003] [Medline: 19939713]
5. Fang H, Tu S, Sheng J, Shao A. Depression in sleep disturbance: a review on a bidirectional relationship, mechanisms and treatment. J Cell Mol Med 2019 Apr;23(4):2324-2332 [FREE Full text] [doi: 10.1111/jcmm.14170] [Medline: 30734486]
6. El-Den S, Chen TF, Gan YL, Wong E, O'Reilly CL. The psychometric properties of depression screening tools in primary healthcare settings: a systematic review. J Affect Disord 2018 Jan 01;225:503-522. [doi: 10.1016/j.jad.2017.08.060] [Medline: 28866295]
7. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010 May;19(4):539-549 [FREE Full text] [doi: 10.1007/s11136-010-9606-8] [Medline: 20169472]
8. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, et al. Screening for depression in primary care with patient health questionnaire-9 (PHQ-9): a systematic review. J Affect Disord 2021 Jan 15;279:473-483. [doi: 10.1016/j.jad.2020.09.131] [Medline: 33126078]
9. Levis B, Benedetti A, Thombs BD, DEPRESSion Screening Data (DEPRESSD) Collaboration. Accuracy of patient health questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019 Apr 09;365:l1476 [FREE Full text] [doi: 10.1136/bmj.l1476] [Medline: 30967483]
10. Bastien CH, Vallières A, Morin CM. Validation of the insomnia severity index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
11. Schulte T, Hofmeister D, Mehnert-Theuerkauf A, Hartung T, Hinz A. Assessment of sleep problems with the insomnia severity index (ISI) and the sleep item of the patient health questionnaire (PHQ-9) in cancer patients. Support Care Cancer 2021 Dec;29(12):7377-7384 [FREE Full text] [doi: 10.1007/s00520-021-06282-x] [Medline: 34050799]
12. Kroenke K, Spitzer RL, Williams JB. The patient health questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
13. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010 Jul;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]
14. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Feb;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
15. Kraepelien M, Blom K, Forsell E, Hentati Isacsson N, Bjurner P, Morin CM, et al. A very brief self-report scale for measuring insomnia severity using two items from the insomnia severity index - development and validation in a clinical population. Sleep Med 2021 May;81:365-374 [FREE Full text] [doi: 10.1016/j.sleep.2021.03.003] [Medline: 33813233]
16. Morin CM, Belleville G, Bélanger L, Ivers H. The insomnia severity index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [Medline: 21532953]

17. Voros V, Fekete S, Tenyi T, Rihmer Z, Szili I, Osvath P. Untreated depressive symptoms significantly worsen quality of life in old age and may lead to the misdiagnosis of dementia: a cross-sectional study. Ann Gen Psychiatry 2020 Sep 15;19:52 [FREE Full text] [doi: 10.1186/s12991-020-00302-6] [Medline: 32944058]
18. Fekadu A, Demissie M, Birhane R, Medhin G, Bitew T, Hailemariam M, et al. Under detection of depression in primary care settings in low and middle-income countries: a systematic review and meta-analysis. Syst Rev 2022 Feb 05;11(1):21 [FREE Full text] [doi: 10.1186/s13643-022-01893-9] [Medline: 35123556]
19. Colombo D, Fernández-Álvarez J, Patané A, Semonella M, Kwiatkowska M, García-Palacios A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. J Clin Med 2019 Apr 05;8(4):465 [FREE Full text] [doi: 10.3390/jcm8040465] [Medline: 30959828]
20. Kim S, Kim JA, Lee JY. International trend of non-contact healthcare and related changes due to COVID-19 pandemic. Yonsei Med J 2022 Jan;63(Suppl):S22-S33 [FREE Full text] [doi: 10.3349/ymj.2022.63.S22] [Medline: 35040603]
21. Yoon D. Preparing for a new world: making friends with digital health. Yonsei Med J 2022 Jan;63(Suppl):S108-S111 [FREE Full text] [doi: 10.3349/ymj.2022.63.S108] [Medline: 35040611]
22. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol Res Pr 2011;42(6):505-512 [FREE Full text] [doi: 10.1037/a0024485]
23. Yang YS, Ryu GW, Han I, Oh S, Choi M. Ecological momentary assessment using smartphone-based mobile application for affect and stress assessment. Healthc Inform Res 2018 Oct;24(4):381-386 [FREE Full text] [doi: 10.4258/hir.2018.24.4.381] [Medline: 30443427]
24. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
25. Yang YS, Ryu GW, Choi M. Methodological strategies for ecological momentary assessment to evaluate mood and stress in adult patients using mobile phones: systematic review. JMIR Mhealth Uhealth 2019 Apr 01;7(4):e11215 [FREE Full text] [doi: 10.2196/11215] [Medline: 30932866]
26. Wirz-Justice A. Diurnal variation of depressive symptoms. Dialogues Clin Neurosci 2008;10(3):337-343 [FREE Full text] [doi: 10.31887/DCNS.2008.10.3/awjustice] [Medline: 18979947]
27. Han C, Jo SA, Kwak JH, Pae CU, Steffens D, Jo I, et al. Validation of the Patient Health Questionnaire-9 Korean version in the elderly population: the Ansan Geriatric study. Compr Psychiatry 2008;49(2):218-223. [doi: 10.1016/j.comppsych.2007.08.006] [Medline: 18243897]
28. Jo M, Koo HY, Cho IY, Lee Y, Yoon S, Yang Y, et al. Usefulness of the patient health questionnaire-2 in screening for depression. Korean J Fam Pract 2019 Aug 20;9(4):336-340 [FREE Full text] [doi: 10.21215/kjfp.2019.9.4.336]
29. Eaton WW, Smith C, Ybarra M, Muntaner C, Tien A. Center for epidemiologic studies depression scale: review and revision (CESD and CESD-R). In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults. Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers; 2004:363-377.
30. Lee S, Oh ST, Ryu SY, Jun JY, Lee K, Lee E, et al. Validation of the Korean version of Center for Epidemiologic Studies Depression Scale-Revised (K-CESD-R). Korean J Psychosom Med 2016;24(1):83-93 [FREE Full text] [doi: 10.1093/ijnp/pyw043.223]
31. Hong J, Kim J, Kim S, Oh J, Lee D, Lee S, et al. Depressive symptoms feature-based machine learning approach to predicting depression using smartphone. Healthcare (Basel) 2022 Jun 24;10(7):1189 [FREE Full text] [doi: 10.3390/healthcare10071189] [Medline: 35885716]
32. Anderson KN. Insomnia and cognitive behavioural therapy-how to assess your patient and why it should be a standard part of care. J Thorac Dis 2018 Jan;10(Suppl 1):S94-102 [FREE Full text] [doi: 10.21037/jtd.2018.01.35] [Medline: 29445533]
33. Burchert S, Kerber A, Zimmermann J, Knaevelsrud C. Screening accuracy of a 14-day smartphone ambulatory assessment of depression symptoms and mood dynamics in a general population sample: comparison with the PHQ-9 depression screening. PLoS One 2021 Jan 06;16(1):e0244955 [FREE Full text] [doi: 10.1371/journal.pone.0244955] [Medline: 33406120]
34. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. Biom J 2005 Aug;47(4):458-472. [doi: 10.1002/bimj.200410135] [Medline: 16161804]
35. Colombo D, Suso-Ribera C, Fernandez-Álvarez J, Felipe IF, Cipresso P, Palacios AG, et al. Exploring affect recall bias and the impact of mild depressive symptoms: an ecological momentary study. In: Proceedings of the 9th International Conference on Pervasive Computing Paradigms for Mental Health. 2019 Presented at: MindCare '19; April 23–24, 2019; Buenos Aires, Argentina URL: https://link.springer.com/chapter/10.1007/978-3-030-25872-6_17 [doi: 10.1007/978-3-030-25872-6_17]
36. Ben-Zeev D, Young MA, Madsen JW. Retrospective recall of affect in clinically depressed individuals and controls. Cogn Emot 2009 Aug;23(5):1021-1040 [FREE Full text] [doi: 10.1080/02699930802607937]
37. Jejdling F. Ericsson mobility report 2018. Ericsson. 2018. URL: https://www.key4biz.it/wp-content/uploads/2018/06/ericsson-mobility-report-june-2018.pdf [accessed 2022-10-15]
38. Akram U, Barclay NL, Milkins B. Sleep-related attentional bias in insomnia: time to examine moderating factors? Front Psychol 2018 Dec 14;9:2573 [FREE Full text] [doi: 10.3389/fpsyg.2018.02573] [Medline: 30618989]

Abbreviations

AUC: area under the curve
EMA: ecological momentary assessment
ISI: Insomnia Severity Index
ISI-2: Insomnia Severity Index-2
K-CESD-R: Korean version of the Center for Epidemiologic Studies Depression Scale–Revised
PHQ-2: Patient Health Questionnaire-2
PHQ-9: Patient Health Questionnaire-9
ROC: receiver operating characteristic
YJ: Youden J

Edited by J Torous; submitted 17.01.23; peer-reviewed by T Martino, C Ho; comments to author 10.03.23; revised version received 30.03.23; accepted 11.04.23; published 22.05.23.

Please cite as:
Oh JW, Kim SM, Lee D, Son NH, Uh J, Yoon JH, Choi Y, Lee S
Evaluating the Modified Patient Health Questionnaire-2 and Insomnia Severity Index-2 for Daily Digital Screening of Depression and Insomnia: Validation Study
JMIR Ment Health 2023;10:e45543
URL: https://mental.jmir.org/2023/1/e45543
doi:10.2196/45543
PMID:37213186

Original Paper

Evaluating the Psychometric Properties and Clinical Utility of a Digital Psychosocial Self-Screening Tool (HEARTSMAP-U) for Postsecondary Students: Prospective Cohort Study

Punit Virk[1,2,3], PhD; Ravia Arora[2], BSc; Heather Burt[2], MSc; Caitlin Finnamore[2], MSc, MD; Anne Gadermann[1,4], PhD; Skye Barbic[4,5], PhD; Quynh Doan[1,2,6], MDCM, PhD

[1]School of Population and Public Health, The University of British Columbia, Vancouver, BC, Canada
[2]BC Children's Hospital Research Institute, Vancouver, BC, Canada
[3]Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[4]Centre for Health Evaluation and Outcome Sciences, Providence Health, Vancouver, BC, Canada
[5]Department of Occupational Science and Occupational Therapy, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada
[6]Department of Paediatrics, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Punit Virk, PhD
School of Population and Public Health
The University of British Columbia
2206 East Mall
Vancouver, BC, V6T 1Z3
Canada
Phone: 1 604 822 2772
Email: punit.virk@ucalgary.ca

Abstract

Background: Existing screening tools for mental health issues among postsecondary students have several challenges, including a lack of standardization and codevelopment by students. HEARTSMAP-U was adapted to address these issues.

Objective: This study aimed to evaluate the suitability of HEARTSMAP-U as a self-screening tool for psychosocial issues among postsecondary students by evaluating its validity evidence and clinical utility.

Methods: A prospective cohort study was conducted with University of British Columbia Vancouver students to evaluate HEARTSMAP-U's predictive validity and convergent validity. Participating students completed baseline and 3-month follow-up assessments via HEARTSMAP-U and a clinician-administered interview.

Results: In a diverse student sample (n=100), HEARTSMAP-U demonstrated high sensitivity (95%-100%) in identifying any psychiatric concerns that were flagged by a research clinician, with lower specificity (21%-25%). Strong convergent validity (r=0.54-0.68) was demonstrated when relevant domains and sections of HEARTSMAP-U were compared with those of other conceptually similar instruments.

Conclusions: This preliminary evaluation suggests that HEARTSMAP-U may be suitable for screening in the postsecondary educational setting. However, a larger-scale evaluation is necessary to confirm and expand on these findings.

(JMIR Ment Health 2023;10:e48709) doi:10.2196/48709

KEYWORDS

mental health; screening; validity; postsecondary students; clinical utility

Introduction

Background

The prevalence of diagnosable mental health challenges and generalized psychological distress is rising in the postsecondary student population. In an American student sample from 196 postsecondary institutions (n=155,026), lifetime mental health diagnoses increased from 22% (2007) to 36% (2017), and the percentage of treatment seeking increased from 19% (2007) to 34% (2017) [1]. Similarly, in the Canadian context, between 2013 (n=22,995) and 2019 (n=38,127), the proportion of

students who self-reported a diagnosis of anxiety, depression, an addiction, an eating disorder, bipolar disorder, and schizophrenia from a medical provider increased [2]. Past-year diagnosis of any mental health disorder and formal treatment seeking rose from 22% (2013) to 37% (2019) and 31% (2013) to 37% (2019), respectively [2]. Unfortunately, poor access to care, long wait times due to service saturation, complex referral requirements, and uncoordinated care pathways have been longstanding barriers for students accessing care within postsecondary educational environments [3-6].

The university context is an important venue to consider for detecting untreated mental health issues. In fact, Canada's *National Standard for Post-Secondary Student Mental Health and Well-being* calls for postsecondary institutions to assess opportunities for early detection within their health systems [7]. Universal screening and resource navigational interventions can provide institutions with the necessary infrastructure to promote early detection and intervention of psychological (eg, depression and anxiety) and social challenges (eg, relationships and housing), which we hereafter refer to as psychosocial challenge [8-11]. Psychosocial health is considered a multifaceted term denoting the mental, emotional, and social dimensions that shape and make up individuals' well-being [12]. Moreover, screening tools may support students in reflecting on their mental health status as well as building self-awareness and mental health literacy [13,14]. This is especially important given the fact that low perceived need is a common help-seeking barrier [5]. In addition, universal screening is an approach by which all students may have an equal opportunity for early identification and service acquisition [15] and may offer an effective strategy for reaching equity-seeking groups (eg, racialized, disabled, gender, and sexual minority groups) [8,16-18]. In a campus-wide universal screening campaign open to all students, Kodish et al [16] observed substantial participation of a racially diverse student population (73.3%). Although racialized students were less likely to have received prior mental health treatment compared with non-Hispanic White students, they were just as likely to initiate treatment after completing screening. However, the implementation and integration of screening tools within campus health systems remains variable across institutions [19].

Several challenges that hinder effective screening integration in postsecondary health systems persist. First, at the health system level, positive screens have serious implications for system capacity and adequate service provision. Second, most instruments lack codevelopment with students, neglecting to reflect what students consider important to their psychological and social well-being [8,17,20,21]. Among studies evaluating student-specific psychosocial screening instruments, there is insufficient reporting on how these tools reflect what students perceive to be important to their psychosocial health and well-being [8,17,20,21]. For example, the Counseling Center Assessment of Psychological Symptoms 62-item scale was developed to facilitate multidomain psychosocial self-assessment among postsecondary students across 8 domains: depression, generalized anxiety, social anxiety, academic distress, eating concerns, frustration or anger, alcohol use, and distress [22]. However, there is a paucity of published literature reporting on the tool's content validity, specifically on students' perception of the tool's relevance to their lived experiences and the perceived acceptability of its psychosocial coverage. To address the need for a student-centered, multidomain psychosocial screening tool, we adapted existing pediatric, clinician-administered (HEARTSMAP) and patient-administered (MyHEARTSMAP) instruments into HEARTSMAP-U, a version suitable for postsecondary students. We adopted a student-centered approach, which addressed the aforementioned challenges with existing tools and helped us ensure that HEARTSMAP-U's content was relevant, applicable, and acceptable to students [23].

Third, many existing instruments are based on diagnostic criteria and focus on specific mental health conditions (eg, depression and anxiety) [24,25]. However, multidomain screening tools, such as HEARTSMAP-U, allow for a broader and more holistic assessment of psychosocial stressors experienced by students and may identify nonspecific and subthreshold issues that may not get captured by stringent diagnostic criteria [26].

Fourth, most available instruments consist exclusively of a screening component, often evaluating symptoms or functional impairments on a Likert-style scale. However, systematic reviews and meta-analyses have found that screening alone may have limited impact on patients' health outcomes and health-related quality of life (HRQOL) [24,25]. A critical issue with only screening is the need for adequate service provisions for positive screens. Therefore, the US Preventative Services Task Force recommends routine depression screening only when adequate treatment and follow-up systems are in place [27]. Without these provisions or resource information, students may feel ill-equipped to navigate through campus and community-based mental health services [11]. However, the help-seeking impact of combining psychosocial screening with supportive interventions (eg, personalized feedback, psychoeducation, and resource navigation) remains a largely understudied area. Working within the structure of existing HEARTSMAP instruments allowed us to address these challenges by adapting an instrument with both assessment and resource navigational components. A final challenge with existing instruments is the variable degree of validity evidence supporting their measurement properties for use among postsecondary students.

This Study

We iteratively adapted HEARTSMAP-U through extensive and ongoing engagement with postsecondary students and clinical experts. Campus-based mental health professionals played a vital role in ensuring that HEARTSMAP-U effectively captured a diverse range of clinically relevant psychosocial stressors, varying in severity, and that its content adequately addressed critical safety concerns, such as suicidality, homicidality, and abuse. The student consultation process played a crucial role in establishing a "common language" that facilitated mutual understanding of the tool's content between student users and researchers. This involved implementing helpful features such as hover overs for technical terms, eliminating jargon, and ensuring clarity. In addition, students provided valuable

feedback to ensure that the scoring descriptors were realistic and easily distinguishable [23].

Before implementing a screening instrument in the postsecondary educational setting, evidence demonstrating the instrument's fitness for purpose as a self-screening tool for postsecondary students is needed. In the patient-reported outcome measure context, fitness for purpose refers to how suited an instrument is for a particular setting and depends on the demonstration of validity and reliability evidence and usefulness or "clinical utility" in its context of use [28,29]. Multifaceted evidence of HEARTSMAP-U's validity and clinical utility remains to be established.

Predictive validity assesses how well an instrument can predict gold-standard performance. Clinical evaluation is the gold standard for psychosocial assessment but can be time-consuming and requires specialized training. Demonstrating HEARTSMAP-U's predictive validity is important for determining its suitability as a screener. Convergent validity measures the correlation between scores on instruments measuring similar constructs [30]. By demonstrating HEARTSMAP-U's convergence with similar instruments, we can assess how accurately it measures what it is intended to measure. Clinical utility refers to how well an instrument facilitates treatment planning, clinician-patient interactions, and collaboration [31]. To assess HEARTSMAP-U's clinical utility, we examined students' short-term help-seeking experiences and access to recommended care after self-screening.

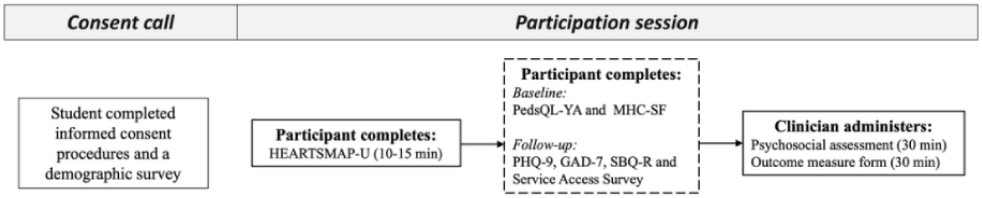
This study aimed to determine whether HEARTSMAP-U is fit for purpose as a psychosocial self-screening tool for students in the postsecondary educational setting. The primary objective of this study was to estimate the preliminary evidence of HEARTSMAP-U's convergent and predictive validity. The secondary objective was to estimate HEARTSMAP-U's clinical utility or the relationship between psychosocial screening and students' use of psychosocial resources at 3-month follow-up.

Methods

Design

We conducted a prospective cohort study with 2 time points, baseline and 3-month follow-up. Baseline and follow-up sessions took place between December 2020 and April 2021 and April 2021 and July 2021, respectively. Students provided informed e-consent before participation. All study procedures were conducted remotely and individually with each participant over the Zoom videoconferencing platform (Zoom Video Communications, Inc), which involved a combination of self-report surveys and a clinician-administered interview. Participants received a CAD $60.00 (US $45.40) gift card or check upon successful completion of the baseline and follow-up sessions. All the procedures are illustrated in Figure 1.

Figure 1. Study procedures chronologically outlined from informed consent to study participation. GAD-7: 7-item Generalized Anxiety Disorder; MHC-SF: Mental Health Continuum-Short Form; PedsQL-YA: Pediatric Quality of Life-Young Adult; PHQ-9: 9-item Patient Health Questionnaire; SBQ-R: Suicide Behaviors Questionnaire–Revised.



Ethics Approval

Approval was obtained from the University of British Columbia (UBC) Behavioural Research Ethics Board before data collection (H20-02556).

Participants

All students aged ≥17 years enrolled at UBC Vancouver at the time of recruitment were eligible to participate. We excluded students who did not have a laptop or desktop computing device to complete study participation, were unavailable for a 3-month follow-up session, or were physically residing outside British Columbia at the time of study participation. Recruitment efforts were entirely web-based (eg, social media, e-newsletters, and institutional listserves). Study promotion efforts were supported by a student-led mental health advocacy organization.

Sampling was conducted in 2 stages. In stage 1, a consecutive, convenience-based sample of 25 eligible students was included. In stage 2, quota-based purposive sampling was conducted to ensure proportional representation and diversity across several demographic variables: gender identity, racial identity, student type (undergraduate, graduate, or professional), and self-rated mental health status (excellent, very good, good, fair, or poor). All quota proportions (Multimedia Appendix 1) were based on a combination of existing UBC institutional demographic data [32], Canadian census data [33], and the broader epidemiological literature on postsecondary students [34]. To facilitate quota-based sampling, all prospective participants completed a brief 1-minute web-based expression of interest form. Select students were invited into the study on a rolling basis.

Measures

HEARTSMAP-U

HEARTSMAP-U is a self-assessment tool designed for postsecondary students to evaluate their psychosocial situation across 10 sections, including *Housing & Material Security*; *Education & Activities*; *Relationships*; *Thoughts & Anxiety*; *Substances & Behavioural Dependencies*; *Safety*; *Sexual Wellness*; *Mood*; *Abuse*; and *Professionals & Resources* [23]. HEARTSMAP-U was adapted for postsecondary students based on previously validated clinician- and patient-administered pediatric versions, HEARTSMAP and MyHEARTSMAP, respectively. The adaptation process has been thoroughly documented elsewhere [23]. Key modifications focused on

ensuring that the tool's content and language were age and developmentally appropriate to the postsecondary context of use. We worked with students and clinicians to ensure that the tool's content was relevant (ie, student specific), comprehensible, and clinically meaningful (ie, captures a broad spectrum of concern severity). For example, the *Home* section on HEARTSMAP or MyHEARTSMAP was adapted to *Housing & Material Security*, recognizing that postsecondary students may not attribute their housing situation as what is socially understood as a "home" and that financial security and self-management is critical for this population.

Each psychosocial section consists of a single item with a 4-point Likert-type scale ranging from 0 (no concern) to 3 (severe concern) to assess for challenges specific to the psychosocial area in question. Brief descriptors accompany each scoring option to help users select the score most appropriate for their situation. In addition, in each section, if the user reports any concerns (score 1-3), they are prompted to indicate whether they have already accessed resources to address those specific concerns (yes or no). Each tool section also has an open-ended textbox that allows students to qualitatively describe their situation as it pertains to their scoring.

After users score all 10 sections, the tool's decision-making algorithm sorts and aggregates sectional scores into broad domains (*Social*, *Functional*, *Student Health*, and *Psychiatry*), each characterized by a unique combination of services and resources, including social services, resources to support daily functioning, frontline resources (for nonpsychiatric, psychosocial needs), psychiatric resources, and services. Each domain maps to 3 sections; multiple sections map to >1 domain. The social domain consists of the *Housing & Material Security*, *Substances & Behavioural Dependencies*, and *Abuse* sections. The function domain includes the *Substances & Behavioural Dependencies*, *Relationships*, and *Sexual Wellness* sections. The psychiatry domain is composed of the *Mood*, *Thoughts & Anxiety*, and *Safety* sections. The domain score (0-9) is the composite or sum of each mapping section's concern severity score (0-3). Each domain score is categorized based on severity, which determines the intensity of tool-generated support recommendations. Score cutoffs are based on clinical judgment and extensive validation in community and clinical samples of youths and adolescents: "none" (0), "mild" (1-3), "moderate" (4-6), or "severe" (7-9) concerns [35-41].

On the basis of students' concern severity and urgency, HEARTSMAP-U may recommend a range of psychiatric (eg, crisis response, psychiatric consultation, counseling services, peer support services, and self-directed resources) and broader psychosocial services (eg, academic counseling, financial advising, housing services, primary care, and peer support). Recommendations may be firm (ie, urgent and time sensitive) or soft (ie, considerations and less urgent) depending on concern severity and acuity. Currently, HEARTSMAP-U does not offer direct referral to any campus- or community-based services but offers students contact and service information (eg, cost, accessibility, and hours of operation) to facilitate help seeking. A paper version of the HEARTSMAP-U instrument is provided in Multimedia Appendix 2.

Clinician-Administered Psychosocial Interview

The clinician-administered psychosocial interview served as a practical, real-world criterion standard against which we measured the evidence of HEARTSMAP-U's predictive validity. Research clinicians were asked to mimic procedures (eg, questioning, probes, and rapport) they typically used when performing intake-style assessments. Although the interview was intended to be open ended and flexible to best reflect real-world assessment content, several parameters were outlined to support the clinicians. First, the interview was not intended to be diagnostic; rather, clinicians were asked to identify broad concern areas in which students demonstrated support needs, ensuring that symptoms that were not clearly defined or were subclinical were not missed. Second, clinicians were provided with a standardized data collection tool to document the various psychological and social (psychosocial) aspects shaping students' mental health. The form was developed with 6 research clinicians to ensure a mutual understanding of reporting expectations.

Before participant recruitment, all research clinicians participated in a web-based 2-hour training session. After reviewing all the study procedures, clinicians reviewed and modified the preliminary form version developed by the study team. Clinicians proposed modifications to minimize potential reporting bias and ensure a consistent understanding and application of the form across all clinicians. To this effect, clinicians engaged in collaborative discussions to identify opportunities to improve content accuracy, clarity, and distinctions between severity levels. At the end of the session, clinicians independently applied the updated data collection form to 2 fictional vignettes, each describing students with differing mental health presentations (eg, concern type and severity). Across both cases, all clinicians consistently classified the presence or absence of concerns, concern severity, and resource needs.

The finalized version of the form consisted of 2 components: the first focused on psychiatric concerns, and the second focused on broader psychosocial issues. This was intended to ensure that in their interviews, clinicians probed and documented both social and psychological challenges. In the first component, the form assesses the presence of psychiatric concerns (yes or no), which have been operationalized as thought disturbances, anxiety, mood-related issues, and suicidal behavior. If psychiatric concerns were endorsed, the clinician was asked to holistically rate and describe concern severity: mild, moderate, or severe. At each severity level, the data collection tool had general descriptors characterizing severity-specific distress, functional impairment, and resource needs to support consistent score interpretation across participating clinicians. The form included relevant International Classification of Diseases, 10th revision coding categories for clinicians to broadly check off areas of concern that might apply to students. These codes were intended to flag suspected concern areas and were not meant to be diagnostically applied. Clinicians also had the option to "write-in" concerns that were not adequately reflected on the form. Finally, clinicians identified any outstanding psychiatric resource needs using a prespecified list of mental health resources that range in intensity (self-directed to urgent

professional care). Again, clinicians had the option of writing in resources or services not mentioned on the form. For each recommended service, clinicians were asked to indicate the time frame within which the student should access the resource: immediately, within 72 hours, within 1 week, or after 1 week.

The structure and content of the second component were similar to those of the first component. Clinicians were asked to rate the presence, severity, and types of broader psychosocial issues that the student may be experiencing, as well as specific resource needs. Broader psychosocial concerns were operationalized as challenges involving the interaction of both social and psychological stressors in relation to students' functioning (eg, relationships), environment (eg, housing and finances), behaviors (high-risk sexual activities and substance use), and development (eg, learning disabilities and cognition).

Pediatric Quality of Life-Young Adult Version

The Pediatric Quality of Life-Young Adult (PedsQL-YA) is a 13-item HRQOL instrument designed to measure physical, emotional, social, and work or school functioning among individuals aged 18 to 25 years. Among college students living with chronic health conditions and those living without chronic health conditions (n=1264), the PedsQL-YA has demonstrated strong internal consistency, known-group discriminant validity, and convergent validity with the Short Form-8 Health Survey [42]. For the current data, the Cronbach α coefficient for the total questionnaire was .92. For the subscales, α estimates were .81 (physical), .83 (emotional), .80 (social), and .77 (school).

Mental Health Continuum-Short Form

The Mental Health Continuum-Short Form (MHC-SF) is a 14-item measure of positive mental health based on a 3-factor model of well-being: emotional, psychological, and social. In cross-national samples, including Canadian people, the MHC-SF has demonstrated strong evidence of internal consistency; structural validity; and criterion validity against concepts of psychological distress and negative social interactions and the World Health Organization Disability Assessment Schedule [43-45]. In this study's sample, the estimated Cronbach α coefficients for the subscales were .88 (emotional), .83 (psychological), and .82 (social).

9-Item Patient Health Questionnaire

The 9-item Patient Health Questionnaire (PHQ-9) is a self-administered unidimensional instrument for screening Diagnostic and Statistical Manual of Mental Disorders, 4th edition symptom criteria for major depressive disorder. The PHQ-9 has been extensively validated in cross-national samples, demonstrating strong evidence of internal consistency, structural validity, and criterion validity against psychological functioning and health care use. Measurement invariance has been demonstrated across racial groups and genders [46-50]. In diverse college student populations, a 1-factor model has been supported [49]. In the current sample, a Cronbach α value of .84 was observed.

7-item Generalized Anxiety Disorder

The 7-item Generalized Anxiety Disorder (GAD-7) is a self-administered, unidimensional instrument for screening Diagnostic and Statistical Manual of Mental Disorders, 4th edition criteria for generalized anxiety disorder. The GAD-7 has demonstrated strong reliability and structural validity evidence among college-attending young adults and, more broadly, evidence of criterion validity against a mental health professional's diagnosis in the adult population [51-53]. Exploratory and confirmatory factor analyses supported a 1-factor model for the GAD-7 in college student populations [51]. Within this study's sample, a Cronbach α coefficient of .91 was observed.

Suicide Behaviors Questionnaire–Revised

The Suicide Behaviors Questionnaire–Revised (SBQ-R) is a 4-item measure for identifying individuals at risk of engaging in suicidal behaviors. In a diverse mixed clinical and community-based sample, the SBQ-R has demonstrated criterion validity against inpatient psychiatric admission (yes or no), specifically showing high sensitivity (93%) and specificity (95%) among college students [54]. Two previous studies by Osman et al [54] and Aloba et al [55] supported a 1-factor model for the SBQ-R. For the current data, a Cronbach α coefficient of .84 was observed.

Study Procedure

Each study session involved 2 separate 30-minute components conducted consecutively (up to 48 hours apart). In the first component, a research assistant briefly introduced the HEARTSMAP-U tool and sent participants a secure link to self-administer the web-based tool version, along with 2 additional self-administered psychological instruments: the PedsQL-YA and MHC-SF. The second component, conducted by a research clinician, involved a psychosocial evaluation based on the clinician's own standards of practice and professional experience. Clinicians were blinded to students' HEARTSMAP-U assessment results. For each participant, clinicians reported on the presence or absence of psychiatric and broader psychosocial concerns and their respective severity levels and types and appropriate campus and community-based resources. After completing their session, students received a copy of their HEARTSMAP-U report, which included service recommendations triggered based on their HEARTSMAP-U scoring pattern. The same clinician completed a student's baseline and follow-up assessments to control for interclinician variability. A total of 5 clinicians carried out assessments, 2 mental health nurses and 3 registered counselors, all employed with the UBC Health Service or Counselling Services.

All procedures were repeated at the 3-month follow-up with several modifications. After their HEARTSMAP-U assessment, participants completed a qualitative survey gauging their experiences accessing care following their baseline session. Students were asked to describe whether they had (1) begun accessing care or (2) attempted to access care (unsuccessful) and (3) intended to access care in the future. In addition, students were asked to report any barriers or challenges they experienced in accessing the tool-recommended resources. At follow-up, participants self-reported on a new set of secondary psychological instruments, which included the PHQ-9, GAD-7, and SBQ-R.

Analytic Approach

Predictive Validity

The study was powered to measure HEARTSMAP-U's sensitivity in detecting psychiatric concerns. On the basis of previous studies, we hypothesized a 2-week psychiatric concern prevalence of 35% and estimated that HEARTSMAP-U would demonstrate 90% sensitivity in identifying psychiatric concerns [41,56]. We considered enrolling 100 students. This sample size would provide ±10% precision with 95% confidence around the 90% estimated sensitivity [57]. We evaluated HEARTSMAP-U's ability to predict both concern severity and resource needs. First, we calculated the sensitivity and specificity of HEARTSMAP-U in detecting psychiatric concerns (*Psychiatry* domain score ≥1) and assessed its ability to distinguish between mild and moderate or severe issues. Second, we calculated the tool's ability to identify different psychiatric resource needs: urgent care (eg, crisis line), same-day primary care, nonurgent primary care, counseling services, peer support, and self-directed care. Owing to the small sample size, our resource-related analysis was restricted to the baseline data. All estimates are reported with 95% CIs. We calculated HEARTSMAP-U's sensitivity and specificity in identifying broader psychosocial issues (eg, housing, relationships, and substance use). The total numbers of true positives, false positives, true negatives, and false negatives are descriptively reported. A summary of our predictive validity analysis has been presented in Multimedia Appendix 3.

Convergent Validity

We evaluated the convergent validity between HEARTSMAP-U and comparator instruments using nonparametric Spearman correlation coefficients and 95% CIs. On the basis of Cohen (1988) conventions, absolute correlation values near r=0.10 were considered weak, near r=0.30 were considered moderate, and near r=0.50 were considered strong [58]. All comparisons were decided on a priori. We hypothesized a strong negative correlation (>0.50) between HEARTSMAP-U's *Psychiatry* domain and both the PedsQL-YA *Emotional Functioning* and MHC-SF *Emotional Well-being* and *Psychological Well-being* subscales. Similarly, we hypothesized a strong negative correlation between HEARTSMAP-U's *Function* domain and the PedsQL-YA *School Functioning* subscale.

Scoring convergence was also evaluated between HEARTSMAP-U's individual psychiatric sections, *Mood*, *Thoughts & Anxiety*, and *Safety*, and composite scores on the PHQ-9, GAD-7, and SBQ-R instruments, respectively. We hypothesized strong correlations between the total score on the PHQ-9 and the *Mood* section, total score on the GAD-7 and the *Thoughts & Anxiety* section, and total score on the SBQ-R and the *Safety* section.

We used the mutual information method to evaluate the level of agreement among the severity classifications in HEARTSMAP-U's *Psychiatry* domain, the PedsQL-YA's *Emotional* subscale, and the total MHC-SF score [59]. Each HEARTSMAP-U domain has 4 severity classifications (no issues, mild, moderate, and severe); the PedsQL-YA has 2 classifications ("not at-risk" and "at-risk" of impaired HRQOL); and the total MHC-SF score produces 3 classifications ("flourishing," "moderately mentally healthy," and "languishing"). Agreement on severity classification is crucial because the 2 measures can have high convergent validity but assess different severities, which can impact health decision-making. Multimedia Appendix 3 provides a summary of the hypothesized correlations and classification-related analyses.

Clinical Utility

At follow-up, participants completed a survey to measure their experiences accessing tool-recommended resources. HEARTSMAP-U makes two types of support recommendations: (1) self-directed resources for maintaining mental well-being and (2) service-based resources for the identified psychiatric and psychosocial needs. Participants were asked whether they had accessed their respective tool-recommended resources (yes or no). Those who had not were asked whether they intended to access the resources in the future (yes or no) and whether they had tried to access the resources (yes or no). If they had tried, they were asked to describe any barriers they experienced. Those who had not attempted to access the resources were asked to briefly explain why and check off any prespecified barriers, including time, cost, cultural sensitivity, transportation, COVID-19, and service availability. Participants could also add any challenges or issues that they felt were not reflected. We report the proportions of students who had begun accessing the resources, those who tried to access the resources but were unsuccessful, and those who made no attempt to access the resources. We also summarize the barriers, challenges, or explanations for each subsample.

Results

Demographic Characteristics

Of the 102 enrolled students who completed informed consent procedures, 100 (98%) students completed all study participation and 2 (2%) dropped out after enrollment but before participation. There was no loss to follow-up between baseline and 3-month follow-up. A total of 530 eligible students expressed interest in the study, the demographic details of whom are summarized in Multimedia Appendix 4. We illustrate participant flow from recruitment to participation in Multimedia Appendix 5.

Among the study participants, balanced distributions were observed for gender, sex, and the year of study. Three-quarters (74/100, 74%) of the participants were enrolled in an undergraduate degree program. A majority were full-time students (97/100, 97%), living off campus (63/100, 63%), single (52/100, 52%), and not currently employed (55/100, 55%). Two-thirds (66/100, 66%) of the participants self-identified with a non-European ethnic background, 70% (70/100) identified as straight, and 20% (20/100) were international students. Similar proportions of students had never (42/100, 42%) or only previously accessed (43/100, 43%) mental health support. A sizable number of students reported living with a learning (12/100, 12%) or physical (6/100, 6%) disability. A complete demographic profile of the study participants is reported in Table 1.

Table 1. A descriptive summary of the demographic, lifestyle, and health-related characteristics reported for all the study participants (n=100).

Demographic characteristics	Values
Age (years), mean (SD)	22 (3.4)
Year of study, n (%)	
1	27 (27)
2	22 (22)
3	31 (31)
4	14 (14)
≥5	6 (6)
Sex, n (%)	
Male	48 (48)
Female	51 (51)
Prefer not to answer	1 (1)
Gender identity, n (%)	
Man	46 (46)
Woman	49 (49)
A different gender identity	5 (5)
Transgender, n (%)	
Yes	0 (0)
No	95 (95)
Prefer not to answer	5 (5)
Sexual identity, n (%)	
Straight	70 (70)
Gay	8 (8)
Bisexual	15 (15)
A different sexual identity	4 (4)
Prefer not to answer	3 (3)
Ethnicity, n (%)	
Aboriginal person	5 (5)
African	3 (3)
East and South Asian	24 (24)
European	34 (34)
South American	3 (3)
South Asian	14 (14)
West Asian and Middle Eastern	8 (8)
Multiethnic	9 (9)
Living on campus (yes), n (%)	37 (37)
Employment status, n (%)	
Full time (>30 h/wk)	6 (6)
Part time (<30 h/wk)	40 (40)
Not employed	55 (55)
Prefer not to answer	1 (1)
Relationship status, n (%)	
Single	52 (52)

Demographic characteristics	Values
Dating	37 (37)
Common law or married	10 (10)
Prefer not to answer	1 (1)
Full-time student (yes), n (%)	97 (97)
Program type, n (%)	
Undergraduate	74 (74)
Graduate	19 (19)
Professional	7 (7)
International student (yes), n (%)	20 (20)
Accessed any mental health support, n (%)	
Currently	29 (29)
Previously	43 (43)
Never	42 (42)
Physical disability, n (%)	
Yes	6 (6)
No	93 (93)
Prefer not to answer	1 (1)
Learning disability (yes), n (%)	12 (12)

Scoring Distribution

Most students scored 0 or 1 on the tool sections at both baseline (80/100, 80% to 96/100, 96%) and follow-up (78/100, 78% to 98/100, 98%), with few severe issues reported (1/100, 1% to 5/100, 5%). No significant difference was found in sectional scoring distributions between baseline and follow-up (chi-square test; P=.06). Clinicians assessed that most students were not currently experiencing psychiatric issues (49/100, 49% to 57/100, 57%), and the tool scored most cases as "mild" (68/100, 68% to 71/100, 71%). The tool and clinical assessments significantly differed in their classification of psychiatric concern severity (P<.001), but within each assessment format, severity classifications remained consistent between baseline and follow-up (P=.06). See Tables 2 and 3 for participants' score distributions on HEARTSMAP-U and psychiatric concerns by severity classification, respectively.

Table 2. Students' score distribution across HEARTSMAP-U's 10 sections at baseline and follow-up (n=100).

Tool section	No issue (0), n (%)		Mild (1), n (%)		Moderate (2), n (%)		Severe (3), n (%)	
	B[a]	F[b]	B	F	B	F	B	F
Housing & Material Security	79 (79)	84 (84)	17 (17)	14 (14)	3 (3)	0 (0)	1 (1)	2 (2)
Education & Activities	40 (40)	50 (50)	40 (40)	42 (42)	18 (18)	7 (7)	2 (2)	1 (1)
Relationships	43 (43)	55 (55)	45 (45)	35 (35)	10 (10)	9 (9)	2 (2)	1 (1)
Thoughts & Anxiety	30 (30)	22 (22)	56 (56)	64 (64)	17 (17)	13 (13)	2 (2)	1 (1)
Substances & Behavioural Dependencies	40 (40)	42 (42)	55 (55)	56 (56)	5 (5)	2 (2)	0 (0)	0 (0)
Safety	78 (78)	81 (81)	18 (18)	17 (17)	4 (4)	1 (1)	0 (0)	0 (0)
Sexual Wellness	78 (78)	82 (82)	9 (9)	6 (6)	10 (10)	8 (8)	3 (3)	4 (4)
Mood	31 (31)	40 (40)	51 (51)	48 (48)	13 (13)	10 (10)	5 (5)	2 (2)
Abuse	55 (55)	61 (61)	29 (29)	27 (27)	16 (16)	12 (12)	0 (0)	0 (0)
Professionals & Resources	59 (59)	73 (73)	25 (25)	19 (19)	11 (11)	7 (7)	5 (5)	1 (1)

[a]B: baseline.

[b]F: follow-up.

Table 3. Number of students classified by HEARTSMAP-U and clinicians at each psychiatric concern severity level at baseline and follow-up (n=100).

Assessment type	No issue, n		Mild, n		Moderate, n		Severe, n	
	B[a]	F[b]	B	F	B	F	B	F
HEARTSMAP-U	10	15	68	71	21	14	1	0
Clinician	49	57	28	31	23	12	1	0

[a]B: baseline.

[b]F: follow-up.

Predictive Validity

At both baseline (100%, 95% CI 93%-100%) and follow-up (95%, 95% CI 84%-99%), HEARTSMAP-U displayed high sensitivity in detecting the presence of any psychiatric concern, as shown in Table 4. Its specificity in distinguishing the presence and absence of psychiatric issues was 25% (95% CI 13%-41%) and 21% (95% CI 11%-34%) at baseline and follow-up, respectively. When the "no issues" and "mild issues" categories were collapsed and treated as a negative screen, the adjusted specificity was high (100%, 95% CI 92%-100%). The initial specificity and high false positive rate all reflected instances where HEARTSMAP-U identified "mild" psychiatric issues, but the clinician identified none. HEARTSMAP-U's sensitivity in triggering resource recommendations ranged from 33% (95% CI 10%-65%) for counseling services to 89% (95% CI 76%-96%) for self-directed resources (eg, workshops, web-based cognitive behavioral therapy, and self-care apps), as reported in Table 5.

Table 4. HEARTSMAP-U's sensitivity and specificity in predicting any psychiatric concerns (mild to severe) identified through a clinician-administered assessment (gold standard).

	Sensitivity (%; 95% CI)	Specificity (%; 95% CI)	Adjusted specificity (%; 95% CI)	Clinician-identified concerns, n (%)			
				Yes		No	
				TP[a]	FN[b]	FP[c]	TN[d]
Any concerns							
Baseline (n=100)	100 (93-100)	25 (13-41)	100 (0.92-1.00)	51 (51)	0 (0)	37 (37)	12 (12)
Follow-up (n=100)	95 (84-99)	21 (11-34)	100 (0.92-1.00)	41 (41)	2 (2)	45 (45)	12 (12)
Moderate or severe							
Baseline (n=51)	67 (43-85)	N/A[e]	77 (58-90)	16 (16)	8 (8)	6 (6)	21 (21)
Follow-up (n=43)	70 (43-95)	N/A	77 (41-83)	7 (7)	3 (3)	8 (8)	25 (25)

[a]TP: true positive.

[b]FN: false negative.

[c]FP: false positive.

[d]TN: true negative.

[e]N/A: not applicable.

Table 5. HEARTSMAP-U's sensitivity and specificity in predicting the psychiatric support needs identified through a clinician-administered assessment (gold standard) at baseline (n=100).

Support recommendation	Sensitivity (%; 95% CI)	Specificity (%; 95% CI)	Clinician-identified concerns, n (%)			
			Issue		No issue	
			TP[a]	FN[b]	FP[c]	TN[d]
Urgent or severe	100 (3-100)	93 (86-97)	1 (1)	0 (0)	7 (7)	92 (92)
GP[e] or counseling services	83 (61-95)	40 (29-52)	19 (19)	4 (4)	46 (46)	31 (31)
GP	81 (54-96)	41 (29-52)	13 (13)	3 (3)	50 (50)	34 (34)
Counseling services	33 (10-65)	96 (89-99)	4 (4)	8 (8)	4 (4)	88 (88)
Peer support	88 (62-98)	24 (15-34)	14 (14)	2 (2)	64 (64)	20 (20)
Self-management	89 (76-96)	4 (0-13)	40 (40)	5 (5)	53 (53)	2 (2)

[a]TP: true positive.

[b]FN: false negative.

[c]FP: false positive.

[d]TN: true negative.

[e]GP: general practitioner.

At baseline, HEARTSMAP-U's sensitivity in detecting the broader psychosocial challenges identified by a clinician-administered assessment ranged from 72% (95% CI 51%-88%) for relationship issues to 100% for substances and behavioral dependencies (95% CI 40%-100%), high-risk sexual behaviors (95% CI 3%-100%), and abuse (95% CI 59%-100%). Specificity ranged from 45% (95% CI 34%-57%) for detecting educational or work-related issues to 86% (95% CI 69%-96%) for detecting housing- and basic need-related issues. As shown in Tables 6 and 7, baseline and follow-up findings were consistent, with the exception of *Housing & Material Security*, where sensitivity decreased to 33% (95% CI 4%-78%).

Table 6. HEARTSMAP-U's sensitivity and specificity in predicting the broader psychosocial concerns identified through a clinician-administered assessment (gold standard) at baseline (n=100).

Section	Sensitivity (%; 95% CI)	Specificity (%; 95% CI)	Clinician-identified concerns, n (%)			
			Yes		No	
			TP[a]	FN[b]	FP[c]	TN[d]
Housing	80 (0.44-0.98)	86 (0.69-0.96)	8 (8)	2 (2)	13 (13)	77 (77)
Education	88 (0.62-0.98)	45 (0.34-0.57)	14 (14)	2 (2)	46 (46)	38 (38)
Relationships	72 (0.51-0.88)	48 (0.36-0.60)	18 (18)	7 (7)	39 (39)	36 (36)
Substances	100 (0.40-1.00)	41 (0.31-0.51)	4 (4)	0 (0)	57 (57)	39 (39)
Sexual	100 (0.03-1.00)	77 (0.67-0.85)	1 (1)	0 (0)	23 (23)	76 (76)
Abuse	100 (0.59-1.00)	65 (0.54-0.74)	7 (7)	0 (0)	33 (33)	60 (60)

[a]TP: true positive.

[b]FN: false negative.

[c]FP: false positive.

[d]TN: true negative.

Table 7. HEARTSMAP-U's sensitivity and specificity in predicting the broader psychosocial concerns identified through a clinician-administered assessment (gold standard) at follow-up (n=100).

Section	Sensitivity (%; 95% CI)	Specificity (%; 95% CI)	Clinician-identified concerns, n (%)			
			Yes		No	
			TP[a]	FN[b]	FP[c]	TN[d]
Housing	33 (0.04-0.78)	85 (0.76-0.92)	2 (2)	4 (4)	14 (14)	80 (80)
Education	75 (0.43-0.95)	53 (0.43-0.64)	9 (9)	3 (3)	41 (41)	47 (47)
Relationships	75 (0.51-0.91)	63 (0.51-0.73)	15 (15)	5 (5)	30 (30)	50 (50)
Substances	100 (0.40-1.00)	45 (0.35-0.55)	4 (4)	0 (0)	55 (55)	43 (43)
Sexual	100 (0.03-1.00)	85 (0.76-0.91)	1 (1)	0 (0)	15 (15)	84 (84)
Abuse	100 (0.40-1.00)	58 (0.48-0.68)	4 (4)	0 (0)	40 (40)	56 (56)

[a]TP: true positive.

[b]FN: false negative.

[c]FP: false positive.

[d]TN: true negative.

Convergent Validity

As shown in Table 8, strong correlations were found between HEARTSMAP-U's *Psychiatry* domain and the PedsQL-YA *Emotional* subscale (0.68, 95% CI 0.56-0.78). Similarly, HEARTSMAP-U's *Psychiatry* domain demonstrated moderate-to-strong correlations with the MHC-SF *Emotional* (0.63, 95% CI 0.49-0.73) and *Psychological* (0.54, 95% CI 0.38-0.67) subscales, as shown in Table 9.

Table 8. Spearman rank correlation coefficients[a] of HEARTSMAP-U domains and PedsQL-YA[b] subscales.

HEARTSMAP-U	PedsQL-YA, correlation coefficient (95% CI)			
	Emotional	School	Social	Psychosocial summary
Psychiatry	0.68 (0.56-0.78)	0.53 (0.38-0.66)	0.50 (0.34-0.63)	0.69 (0.57-0.78)
Function	0.54 (0.39-0.67)	0.64 (0.50-0.74)	0.46 (0.29-0.60)	0.63 (0.50-0.74)
Social	0.36 (0.18-0.52)	0.33 (0.14-0.49)	0.31 (0.12-0.47)	0.39 (0.21-0.55)
Student Health	0.35 (0.17-0.51)	0.33 (0.14-0.49)	0.34 (0.15-0.50)	0.39 (0.21-0.54)

[a]All correlations are significant at an α of .01 (2 tailed).

[b]PedsQL-YA: Pediatric Quality of Life-Young Adult.

Table 9. Spearman rank correlation coefficients[a] of HEARTSMAP-U domains and MHC-SF[b] subscales.

HEARTSMAP-U	MHC-SF, correlation coefficient (95% CI)			
	Emotional	Psychological	Social	Total
Psychiatry	0.63 (0.49-0.73)	0.54 (0.38-0.67)	0.53 (0.37-0.66)	0.60 (0.45-0.71)
Function	0.49 (0.33-0.63)	0.55 (0.40-0.67)	0.39 (0.21-0.54)	0.53 (0.38-0.66)
Social	0.31 (0.12-0.48)	0.32 (0.13-0.49)	0.275 (0.08-0.45)	0.34 (0.15-0.50)
Student Health	0.28 (0.09-0.45)	0.35 (0.17-0.51)	0.21 (0.02-0.39)	0.31 (0.12-0.47)

[a]All correlations are significant at an α of .01 (2-tailed).

[b]MHC-SF: Mental Health Continuum-Short Form.

Concern severity classification (none to severe) in HEARTSMAP-U's *Psychiatry* domain and the PedsQL-YA ("not at-risk" vs "at-risk") showed significant agreement (Tables 6 and 7). Significant agreement was observed between HEARTSMAP-U's *Psychiatry* severity classifications and the MHC-SF's "languishing," "moderately mentally healthy," and "flourishing" classifications (Table 10).

For broader psychosocial issues, strong correlation and classification agreement were observed between HEARTSMAP-U's *Function* domain and the PedsQL-YA *School* subscale (0.64, 95% CI 0.50-0.74). HEARTSMAP-U's *Mood*, *Anxiety*, and *Safety* sectional scores demonstrated moderate-to-strong correlations with condition-specific instruments: PHQ-9 (0.61, 95% CI 0.48-0.74), GAD-7 (0.71,

95% CI 0.60-0.82), and SBQ-R (0.65, 95% CI 0.60-0.82), respectively.

Table 10. Association between severity classifications in HEARTSMAP-U and the PedsQL-YA[a] and MHC-SF[b] instruments using measures of mutual information.

First and second instruments	Mutual information agreement, I agreement	Mutual information disagreement, I disagreement	Mutual information, I agreement + I disagreement	Chi-square (*df*)	*P* value	Association outcome[c]
HEARTSMAP-U Psychiatry						
PedsQL-YA emotional	0.25	−0.01	0.24	33.51 (3)	<.001	Agreement
MHC-SF total	0.19	0.07	0.26	35.68 (6)	<.001	Agreement
HEARTSMAP-U Function						
PedsQL-YA work or school	0.18	0.00	0.18	24.99 (3)	<.001	Agreement
MHC-SF total	0.10	0.11	0.20	28.12 (6)	<.001	Disagreement

[a]PedsQL-YA: Pediatric Quality of Life-Young Adult.

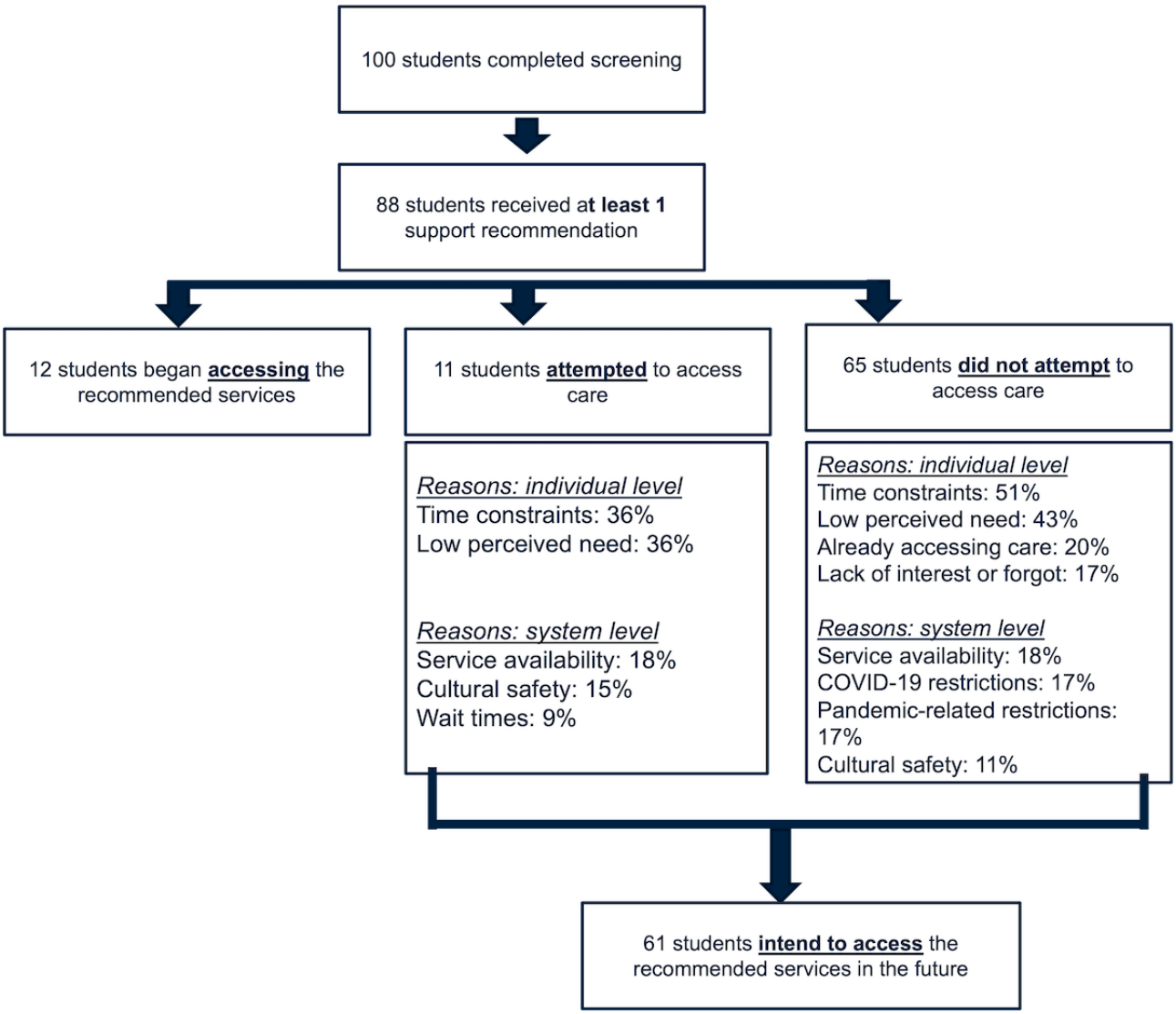
[b]MHC-SF: Mental Health Continuum-Short Form.

[c]If I agreement > I disagreement and *P*<.10, then there is a significant agreement [59].

Clinical Utility

As shown in Figure 2, most participants (88/100, 88%) received at least 1 tool-triggered psychiatric or broader psychosocial support recommendation for the identified needs of any severity level. A smaller fraction (12/100, 12%) of the students solely received recommendations for self-care and maintaining mental well-being. Of the 88 (88%) out of 100 students who received a needs-based support recommendation, 12 (14%) had started accessing the recommended care before their follow-up visit. Accessed services included counseling or primary care services (50/100, 50%), academic advising support (25/100, 25%), and web-based resources (25/100, 25%).

Figure 2. A schematic outlining participants' connection with psychosocial resources at 3-month follow-up. For a more nuanced understanding of students' help-seeking, we measured access, attempt, and intention to seek care. As per the British Columbia Centre for Disease Control, culturally safe care is "an outcome based on respectful engagement that recognizes and strives to address power imbalances inherent in the healthcare system. It results in an environment free of racism and discrimination, where people feel safe" [60].



After baseline, 13% (11/88) of the students had attempted to access needs-based services but were unsuccessful in establishing a connection with service providers. Students described the individual- and system-level factors that impeded their service access. At the individual level, time constraints (32/88, 36%) and low perceived need (32/88, 36%) were the most common explanations. At the system level, issues with service availability (16/88, 18%) were the most frequently reported barrier. Additional reasons are outlined in Figure 2.

Of all the students who received needs-based support recommendations (n=88), a total of 65 (74%) students did not attempt to access the specified support. Time constraints (45/88, 51%) and low perceived need (38/88, 43%) were the most common reasons for not accessing care. Additional explanations are summarized in Figure 2. Despite a large proportion of students not accessing care, a majority of these students (54/88, 61%) expressed their intention to do so in the future, either when they perceived the need for them or had more time.

Discussion

Principal Findings

The aim of this study was to determine whether HEARTSMAP-U was fit for purpose as a psychosocial self-screening tool for postsecondary students. HEARTSMAP-U displayed high sensitivity and lower specificity in identifying psychiatric concerns and resource needs. Moderate to strong convergent validity evidence was demonstrated between conceptually similar domains in HEARTSMAP-U and other general and condition-specific patient-reported outcome measures. At follow-up, 25% (22/100) of the students who had received service recommendations from HEARTSMAP-U had attempted to access the services. Of those who could not connect with a service provider or those who did not attempt to access care, a majority (61/76, 80%) intended to act on them when they had time or perceived a sufficient need.

HEARTSMAP-U shows high sensitivity in detecting psychiatric and psychosocial concerns and resource needs but consistent

low specificity, leading to false positives or overscreening [61]. False positives can be concerning when they result in psychological distress or overwhelm health systems [62,63]. Most of HEARTSMAP-U's false negatives were for mild support needs (eg, self-direct resources and peer support) or soft recommendations for primary care ("consider accessing..."). To address this, we incorporated an additional recommendation into HEARTSMAP-U's algorithm, encouraging students experiencing mild issues to secure access to a primary care provider in case of future need escalation (eg, build resource literacy and have a help-seeking plan). Students described soft recommendations as being helpful, "if and when I need it." Thus, HEARTSMAP-U's false positives were not perceived by participants as being distressful and are not likely to have significant system-level repercussions. Nevertheless, within the context of standard 2-stage screening procedures [64], HEARTSMAP-U is intended as an initial screening tool and requires follow-up assessment for firm diagnosis and treatment planning.

We observed a moderate-to-high correlation between HEARTSMAP-U's *Mood* and *Anxiety* sections and total 8-item Patient Health Questionnaire and GAD-7 scores. Consistent with our findings, Alschuler et al [65] observed moderate to high correlation between the College Health Questionnaire and Patient Health Questionnaire (r=0.37-0.47), a general screener for common mental health disorders. Similarly, Downs et al [66] showed that the Symptoms and Assets Screening Scale's *Anxiety Symptoms* and *Depressive Symptoms* subscales were significantly correlated with the Beck Anxiety Inventory (r=0.68) and the 8-item Patient Health Questionnaire (r=0.73), respectively. HEARTSMAP-U's convergent validity lends evidence in support of our previously developed conceptual framework. The consistency of these findings with both the College Health Questionnaire and Symptoms and Assets Screening Scale is also promising, as the current work builds on these seminal instruments and extends screening utility beyond assessment but includes resource recommendations and navigational support.

Approximately three-quarters (65/88, 74%) of the participants who received HEARTSMAP-U resource recommendations did not attempt to access care by the 3-month follow-up, with many citing low perceived need and time constraints as common barriers. Similarly, in a random sample of 2785 American college students, Eisenberg et al [67] reported that between 37% and 84% of positive screens had unmet service needs, with a lack of perceived need and a lack of time as the most commonly reported barriers. The extant literature has demonstrated consistently low help seeking and, at best, a preference for informal help seeking (eg, friends and family) among postsecondary students [5,68,69]. Indeed, this may be a concerning observation, as a number of our participants may have benefited from mild-to-moderate mental health resources but did not perceive a need for them, which may result in their concerns escalating and impacting their daily functioning and academic success.

Interestingly, over two-thirds (61/88, 69%) of the students who did not access care expressed an intention to use tool-recommended resources in the future. Although consistent with the existing understanding that students exposed to behavioral persuasion messaging report greater help-seeking intention [70], this study extends this knowledge to the context in which screening interventions are coupled with resource messaging. Students experienced system- and individual-level challenges as barriers to care. Several students (4/100, 4%) called for more self-referral or integrated services, voicing frustrations that even with resource information, accessing care involved too many steps. These comments echo longstanding challenges with screening programs, as their utility is often contingent on system capacity and readiness [17]. In primary care settings, the colocation of behavioral screening with specialists has shown increased referral completion and service use [71]. Similarly, linking HEARTSMAP-U with integrated digital mental health services may reduce logistical barriers (eg, opening new webpages and getting a general practitioner referral) and support more seamless connection with resources after screening. To this effect, we have worked with institutional partners and researchers to embed HEARTSMAP-U within an e-mental health app called "Minder," which includes built-in and immediately accessible life coaching, e-counseling, peer support, and self-directed resources [72].

This paper describes a comprehensive evaluation of HEARTSMAP-U's measurement properties and clinical utility. Together, the evidence of HEARTSMAP-U's validity and students' intention to seek tool-recommended resources demonstrate that HEARTSMAP-U may be fit for purpose as a psychosocial self-screening tool for the postsecondary educational setting. However, care must be taken to ensure appropriate access to secondary assessments to "rule out" false positives. The National College Health Assessment and Canadian Campus Well-being Survey offer population-level data on student health and academic outcomes [73,74]. However, to our knowledge, these instruments do not offer individual-level information that could facilitate students' mental health literacy, self-awareness, and understanding of personal support needs. Validated at the individual level, HEARTSMAP-U has the potential to be scaled and offered alongside institution-facing measures to promote measurement that is guided by the principles of student centeredness, as outlined in Canada's *National Standard for Post-Secondary Student Mental Health and Well-being* [7].

A strength of this study lies in working with campus-based mental health clinicians nurses, as their assessment served as a pragmatic and realistic "gold standard" for the evaluation of predicative validity. This is especially important, as it allowed us to identify nonspecific or subthreshold concerns that may not meet the rigid diagnostic criteria. In addition, the gold standard should be driven by student needs and put their voices first, which is best done by campus mental health clinicians who work with students on a daily basis. Our study was not without limitations. HEARTSMAP-U's scoring distribution was skewed toward mild to moderate concerns, limiting our ability to validate the instrument's ability to identify severe psychiatric and psychosocial needs. However, the study's concern prevalence was consistent with the population-level prevalence and appropriate for the tool's use as a universal screening measure. Our small sample size also limited the

precise evaluation of HEARTSMAP-U's measurement properties for specific concern severity levels and resource recommendations. Further studies are planned to address these limitations and evaluate HEARTSMAP-U's predictive performance in different student subpopulations (eg, varying in gender and race).

Conclusions

Our study offers a preliminary evaluation of HEARTSMAP-U's measurement properties and clinical utility for multidomain, universal psychosocial screening in the postsecondary educational setting. HEARTSMAP-U displays high sensitivity but lower specificity in identifying psychiatric concerns and resource needs as well as moderate-to-high convergent validity with other conceptually similar instruments. Our results suggest that HEARTSMAP-U has the potential to be scaled and implemented alongside institution-facing measures (eg, early alerts) to promote the prevention and early detection of mental health issues in the learning environment. Studies are currently underway to evaluate the tool's measurement properties on a larger scale and predictive validity performance across diverse student subpopulations (eg, varying in age, gender, and race).

Acknowledgments

The study reported in this paper is part of the primary investigator's (PV) doctoral dissertation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Variables used to conduct quota-based purposive sampling and target sample sizes for each variable and stratum.
[DOCX File , 16 KB - mental_v10i1e48709_app1.docx]

Multimedia Appendix 2
A paper version of the HEARTSMAP-U instrument.
[DOCX File , 1576 KB - mental_v10i1e48709_app2.docx]

Multimedia Appendix 3
Summary of predictive validity, convergent validity, and severity classification analyses.
[DOCX File , 14 KB - mental_v10i1e48709_app3.docx]

Multimedia Appendix 4
Demographic characteristics of the students who expressed interest in participating in the pilot study.
[DOCX File , 14 KB - mental_v10i1e48709_app4.docx]

Multimedia Appendix 5
Participant flow and retention from recruitment through study participation.
[DOCX File , 50 KB - mental_v10i1e48709_app5.docx]

References

1. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007-2017). Psychiatr Serv 2019 Jan 01;70(1):60-63 [FREE Full text] [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]
2. Linden B, Boyes R, Stuart H. Cross-sectional trend analysis of the NCHA II survey data on Canadian post-secondary student mental health and wellbeing from 2013 to 2019. BMC Public Health 2021 Mar 25;21(1):590 [FREE Full text] [doi: 10.1186/s12889-021-10622-1] [Medline: 33765965]
3. Kitzrow MA. The mental health needs of today's college students: challenges and recommendations. NASPA J 2003 Dec 02;41(1):167-181 [FREE Full text] [doi: 10.2202/1949-6605.1310]
4. Cooke M, Huntly J. White Paper on Postsecondary Student Mental Health. The Coordinating Committee of Vice Presidents Students. 2015. URL: https://occccco.files.wordpress.com/2015/05/ccvps-white-paper-on-postsecondary-student-mental-health-april-2015.pdf [accessed 2023-04-15]
5. Eisenberg D, Golberstein E, Gollust SE. Help-seeking and access to mental health care in a university student population. Med Care 2007 Jul;45(7):594-601. [doi: 10.1097/MLR.0b013e31803bb4c1] [Medline: 17571007]
6. Broglia E, Millings A, Barkham M. Challenges to addressing student mental health in embedded counselling services: a survey of UK higher and further education institutions. Br J Guid Counc 2017 Sep 29;46(4):441-455 [FREE Full text] [doi: 10.1080/03069885.2017.1370695]

7. National standard for mental health and well-being for post-secondary students. Mental Health Commission of Canada. 2010. URL: https://mentalhealthcommission.ca/studentstandard/ [accessed 2023-04-15]
8. Kim EH, Coumar A, Lober WB, Kim Y. Addressing mental health epidemic among university students via web-based, self-screening, and referral system: a preliminary study. IEEE Trans Inf Technol Biomed 2011 Mar;15(2):301-307. [doi: 10.1109/TITB.2011.2107561] [Medline: 21257386]
9. Canady VA. University mental health screening kiosk promotes awareness, prevention. Ment Health Wkly 2015 Sep 21;25(36):1-5 [FREE Full text] [doi: 10.1002/mhw.30334]
10. Wiljer D, Shi J, Lo B, Sanches M, Hollenberg E, Johnson A, et al. Effects of a mobile and web app (Thought Spot) on mental health help-seeking among college and university students: randomized controlled trial. J Med Internet Res 2020 Oct 30;22(10):e20790 [FREE Full text] [doi: 10.2196/20790] [Medline: 33124984]
11. VanHeerwaarden N, Ferguson G, Abi-Jaoude A, Johnson A, Hollenberg E, Chaim G, et al. The optimization of an eHealth solution (thought spot) with transition-aged youth in postsecondary settings: participatory design research. J Med Internet Res 2018 Mar 06;20(3):e79 [FREE Full text] [doi: 10.2196/jmir.8102] [Medline: 29510970]
12. Martikainen P, Bartley M, Lahelma E. Psychosocial determinants of health in social epidemiology. Int J Epidemiol 2002 Dec;31(6):1091-1093. [doi: 10.1093/ije/31.6.1091] [Medline: 12540696]
13. Chen SP, Chang WP, Stuart H. Self-reflection and screening mental health on Canadian campuses: validation of the mental health continuum model. BMC Psychol 2020 Jul 29;8(1):76 [FREE Full text] [doi: 10.1186/s40359-020-00446-w] [Medline: 32727614]
14. Lee S, Lim J, Lee S, Heo Y, Jung D. Group-tailored feedback on online mental health screening for university students: using cluster analysis. BMC Prim Care 2022 Jan 25;23(1):19 [FREE Full text] [doi: 10.1186/s12875-021-01622-6] [Medline: 35172741]
15. Dowdy E, Furlong M, Raines TC, Bovery B, Kauffman B, Kamphaus RW, et al. Enhancing school-based mental health services with a preventive and promotive approach to universal screening for complete mental health. J Educ Psychol Consult 2015;25(2-3):178-197 [FREE Full text] [doi: 10.1080/10474412.2014.929951]
16. Kodish T, Lau AS, Gong-Guy E, Congdon E, Arnaudova I, Schmidt M, et al. Enhancing racial/ethnic equity in college student mental health through innovative screening and treatment. Adm Policy Ment Health 2022 Mar;49(2):267-282 [FREE Full text] [doi: 10.1007/s10488-021-01163-1] [Medline: 34505211]
17. Shepardson RL, Funderburk JS. Implementation of universal behavioral health screening in a university health setting. J Clin Psychol Med Settings 2014 Sep;21(3):253-266. [doi: 10.1007/s10880-014-9401-8] [Medline: 25037065]
18. Forbes FM, Whisenhunt BL, Citterio C, Jordan AK, Robinson D, Deal WP. Making mental health a priority on college campuses: implementing large scale screening and follow-up in a high enrollment gateway course. J Am Coll Health 2021 Apr;69(3):275-282. [doi: 10.1080/07448481.2019.1665051] [Medline: 31566504]
19. Jaworska N, De Somma E, Fonseka B, Heck E, MacQueen GM. Mental health services for students at postsecondary institutions: a national survey. Can J Psychiatry 2016 Dec;61(12):766-775 [FREE Full text] [doi: 10.1177/0706743716640752] [Medline: 27310230]
20. Williams A, Larocca R, Chang T, Trinh NH, Fava M, Kvedar J, et al. Web-based depression screening and psychiatric consultation for college students: a feasibility and acceptability study. Int J Telemed Appl 2014;2014:580786 [FREE Full text] [doi: 10.1155/2014/580786] [Medline: 24799895]
21. Bíró É, Ádány R, Kósa K. A simple method for assessing the mental health status of students in higher education. Int J Environ Res Public Health 2019 Nov 27;16(23):4733 [FREE Full text] [doi: 10.3390/ijerph16234733] [Medline: 31783500]
22. Locke BD, McAleavey AA, Zhao Y, Lei PW, Hayes JA, Castonguay LG, et al. Development and initial validation of the Counseling Center Assessment of Psychological Symptoms–34. Meas Eval Couns Dev 2012 Jul;45(3):151-169 [FREE Full text] [doi: 10.1177/0748175611432642]
23. Virk P, Arora R, Burt H, Gadermann A, Barbic S, Nelson M, et al. HEARTSMAP-U: adapting a psychosocial self-screening and resource navigation support tool for use by post-secondary students. Front Psychiatry 2022 Feb 22;13:812965 [FREE Full text] [doi: 10.3389/fpsyt.2022.812965] [Medline: 35280181]
24. Keshavarz H, Fitzpatrick-Lewis D, Streiner DL, Maureen R, Ali U, Shannon HS, et al. Screening for depression: a systematic review and meta-analysis. CMAJ Open 2013 Oct;1(4):E159-E167 [FREE Full text] [doi: 10.9778/cmajo.20130030] [Medline: 25077118]
25. Gilbody S, House AO, Sheldon TA. Screening and case finding instruments for depression. Cochrane Database Syst Rev 2005 Oct 19;2005(4):CD002792 [FREE Full text] [doi: 10.1002/14651858.CD002792.pub2] [Medline: 16235301]
26. Bradford S, Rickwood D. Psychosocial assessments for young people: a systematic review examining acceptability, disclosure and engagement, and predictive utility. Adolesc Health Med Ther 2012 Dec 21;3:111-125 [FREE Full text] [doi: 10.2147/AHMT.S38442] [Medline: 24600292]
27. Siu AL, US Preventive Services Task Force. Screening for depression in children and adolescents: US preventive services task force recommendation statement. Pediatrics 2016 Mar;137(3):e20154467. [doi: 10.1542/peds.2015-4467] [Medline: 26908686]

28. Barbic S, Celone D, Dewsnap K, Duque SR, Osiecki S, Valevicius D, et al. Mobilizing Measurement Into Care: A Guidebook for Service Providers Working With Young Adults Aged 12–24 Years. Vancouver, BC: University of British Columbia and McGill University; 2021.
29. Edwards MC, Slagle A, Rubright JD, Wirth RJ. Fit for purpose and modern validity theory in clinical outcomes assessment. Qual Life Res 2018 Jul;27(7):1711-1720. [doi: 10.1007/s11136-017-1644-z] [Medline: 28688085]
30. Prinsen CA, Mokkink LB, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN methodology for systematic reviews of Patient‐Reported Outcome Measures (PROMs):user manual. Cosmin. 2018. URL: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf [accessed 2023-04-15]
31. Carrozzino D, Patierno C, Guidi J, Berrocal Montiel C, Cao J, Charlson ME, et al. Clinimetric criteria for patient-reported outcome measures. Psychother Psychosom 2021;90(4):222-232 [FREE Full text] [doi: 10.1159/000516599] [Medline: 34038901]
32. Demographics overview. Planning and Institutional Research Office. URL: http://pair.ubc.ca/student-demographics/demographics/ [accessed 2021-03-24]
33. Wiens K, Bhattarai A, Dores A, Pedram P, Williams JV, Bulloch AG, et al. Mental health among Canadian postsecondary students: a mental health crisis? Can J Psychiatry 2020 Jan;65(1):30-35 [FREE Full text] [doi: 10.1177/0706743719874178] [Medline: 31939333]
34. Canadian reference group: executive summary spring 2019. American College Health Association. 2019. URL: https://www.acha.org/documents/ncha/NCHA-II_SPRING_2019_US_REFERENCE_GROUP_EXECUTIVE_SUMMARY.pdf [accessed 2023-04-15]
35. Lee A, Deevska M, Stillwell K, Black T, Meckler G, Park D, et al. A psychosocial assessment and management tool for children and youth in crisis. CJEM 2019 Jan;21(1):87-96. [doi: 10.1017/cem.2018.1] [Medline: 29587889]
36. Virk P, Stenstrom R, Doan Q. Reliability testing of the HEARTSMAP psychosocial assessment tool for multidisciplinary use and in diverse emergency settings. Paediatr Child Health 2018 Dec;23(8):503-508 [FREE Full text] [doi: 10.1093/pch/pxy017] [Medline: 30842695]
37. Gill C, Arnold B, Nugent S, Rajwani A, Xu M, Black T, et al. Reliability of HEARTSMAP as a tool for evaluating psychosocial assessment documentation practices in emergency departments for pediatric mental health complaints. Acad Emerg Med 2018 Dec;25(12):1375-1384 [FREE Full text] [doi: 10.1111/acem.13506] [Medline: 29924893]
38. Virk P, Laskin S, Gokiert R, Richardson C, Newton M, Stenstrom R, et al. MyHEARTSMAP: development and evaluation of a psychosocial self-assessment tool, for and by youth. BMJ Paediatr Open 2019 Jul 24;3(1):e000493 [FREE Full text] [doi: 10.1136/bmjpo-2019-000493] [Medline: 31414065]
39. Lamoureux E, Ishikawa T, Yeates KO, Brooks BL, Beauchamp MH, Craig W, Pediatric Emergency Research Canada (PERC) A-CAP concussion team. Convergent validity of myheartsmap: a pediatric psychosocial health screening tool. Child Psychiatry Hum Dev 2023 Feb;54(1):66-75. [doi: 10.1007/s10578-021-01221-7] [Medline: 34350505]
40. Li BC, Wright B, Black T, Newton AS, Doan Q. Utility of MyHEARTSMAP in youth presenting to the emergency department with mental health concerns. J Pediatr 2021 Aug;235:124-129. [doi: 10.1016/j.jpeds.2021.03.062] [Medline: 33819465]
41. Doan Q, Wright B, Atwal A, Hankinson E, Virk P, Azizi H, et al. Utility of MyHEARTSMAP for universal psychosocial screening in the emergency department. J Pediatr 2020 Apr;219:54-61.e1. [doi: 10.1016/j.jpeds.2019.12.046] [Medline: 32106963]
42. Varni JW, Limbers CA. The PedsQL 4.0 generic core scales young adult version: feasibility, reliability and validity in a university student population. J Health Psychol 2009 May;14(4):611-622. [doi: 10.1177/1359105309103580] [Medline: 19383661]
43. Franken K, Lamers SM, Ten Klooster PM, Bohlmeijer ET, Westerhof GJ. Validation of the mental health continuum-short form and the dual continua model of well-being and psychopathology in an adult mental health setting. J Clin Psychol 2018 Dec;74(12):2187-2202 [FREE Full text] [doi: 10.1002/jclp.22659] [Medline: 29978482]
44. Robitschek C, Keyes CL. Keyes's model of mental health with personal growth initiative as a parsimonious predictor. J Couns Psychol 2009 Apr;56(2):321-329 [FREE Full text] [doi: 10.1037/a0013954]
45. Keyes CL. Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. Am Psychol 2007 Feb;62(2):95-108. [doi: 10.1037/0003-066X.62.2.95] [Medline: 17324035]
46. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
47. Garlow SJ, Rosenberg J, Moore JD, Haas AP, Koestner B, Hendin H, et al. Depression, desperation, and suicidal ideation in college students: results from the American Foundation for Suicide Prevention College Screening Project at Emory University. Depress Anxiety 2008;25(6):482-488. [doi: 10.1002/da.20321] [Medline: 17559087]
48. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. J Gen Intern Med 2006 Jun;21(6):547-552 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00409.x] [Medline: 16808734]
49. Keum BT, Miller MJ, Inkelas KK. Testing the factor structure and measurement invariance of the PHQ-9 across racially diverse U.S. college students. Psychol Assess 2018 Aug;30(8):1096-1106. [doi: 10.1037/pas0000550] [Medline: 29565614]

50. Knight A, Weiss P, Morales K, Keren R. A50: mental health and healthcare utilization in children with SLE and MCTD. Arthritis Rheumatol 2014 Apr 04;66:S75 [FREE Full text] [doi: 10.1002/art.38466]
51. Byrd-Bredbenner C, Eck K, Quick V. Psychometric properties of the Generalized Anxiety Disorder-7 and Generalized Anxiety Disorder-mini in United States University Students. Front Psychol 2020 Sep 24;11:550533 [FREE Full text] [doi: 10.3389/fpsyg.2020.550533] [Medline: 33071867]
52. Ballester L, Alayo I, Vilagut G, Almenara J, Cebrià AI, Echeburúa E, UNIVERSAL study group. Accuracy of online survey assessment of mental disorders and suicidal thoughts and behaviors in Spanish university students. Results of the WHO World Mental Health- International College Student initiative. PLoS One 2019 Sep 05;14(9):e0221529 [FREE Full text] [doi: 10.1371/journal.pone.0221529] [Medline: 31487306]
53. Bártolo A, Monteiro S, Pereira A. Factor structure and construct validity of the Generalized Anxiety Disorder 7-item (GAD-7) among Portuguese college students. Cad Saude Publica 2017 Sep 28;33(9):e00212716 [FREE Full text] [doi: 10.1590/0102-311X00212716] [Medline: 28977285]
54. Osman A, Bagge CL, Gutierrez PM, Konick LC, Kopper BA, Barrios FX. The Suicidal Behaviors Questionnaire-Revised (SBQ-R): validation with clinical and nonclinical samples. Assessment 2001 Dec;8(4):443-454. [doi: 10.1177/107319110100800409] [Medline: 11785588]
55. Aloba O, Ojeleye O, Aloba T. The psychometric characteristics of the 4-item Suicidal Behaviors Questionnaire-Revised (SBQ-R) as a screening tool in a non-clinical sample of Nigerian university students. Asian J Psychiatr 2017 Apr;26:46-51. [doi: 10.1016/j.ajp.2017.01.017] [Medline: 28483090]
56. The impact of COVID-19 on college student well-being. The Healthy Minds Network and American College Health Association. 2020. URL: https://healthymindsnetwork.org/wp-content/uploads/2020/07/Healthy_Minds_NCHA_COVID_Survey_Report_FINAL.pdf [accessed 2023-04-15]
57. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad Emerg Med 1996 Sep;3(9):895-900 [FREE Full text] [doi: 10.1111/j.1553-2712.1996.tb03538.x] [Medline: 8870764]
58. Eccleston C. Research methods in psychology. J Psychosom Res 1996 May;40(5):551-552. [doi: 10.1016/0022-3999(95)00555-2]
59. Kang Y, Steis MR, Kolanowski AM, Fick D, Prabhu VV. Measuring agreement between healthcare survey instruments using mutual information. BMC Med Inform Decis Mak 2016 Jul 26;16:99 [FREE Full text] [doi: 10.1186/s12911-016-0335-y] [Medline: 27456095]
60. Culturally safe care. BC Centre for Disease Control. 2023. URL: http://www.bccdc.ca/health-professionals/clinical-resources/covid-19-care/education-and-training/culturally-safe-care#:~:text=What [accessed 2023-06-08]
61. Jenniskens K, de Groot JA, Reitsma JB, Moons KG, Hooft L, Naaktgeboren CA. Overdiagnosis across medical disciplines: a scoping review. BMJ Open 2017 Dec 27;7(12):e018448 [FREE Full text] [doi: 10.1136/bmjopen-2017-018448] [Medline: 29284720]
62. van der Velde JL, Blanker MH, Stegmann ME, de Bock GH, Berger MY, Berendsen AJ. A systematic review of the psychological impact of false-positive colorectal cancer screening: what is the role of the general practitioner? Eur J Cancer Care (Engl) 2017 May;26(3). [doi: 10.1111/ecc.12709] [Medline: 28493375]
63. Nielssen O, Wallace D, Large M. Pokorny's complaint: the insoluble problem of the overwhelming number of false positives generated by suicide risk assessment. BJPsych Bull 2017 Feb;41(1):18-20 [FREE Full text] [doi: 10.1192/pb.bp.115.053017] [Medline: 28184312]
64. Longford NT. Classification in two-stage screening. Stat Med 2015 Nov 10;34(25):3281-3297. [doi: 10.1002/sim.6554] [Medline: 26082302]
65. Alschuler KN, Hoodin F, Byrd MR. Rapid assessment for psychopathology in a college health clinic: utility of college student specific questions. J Am Coll Health 2009 Sep;58(2):177-179. [doi: 10.1080/07448480903221210] [Medline: 19892655]
66. Downs A, Boucher LA, Campbell DG, Dasse M. Development and initial validation of the Symptoms and Assets Screening Scale. J Am Coll Health 2013;61(3):164-174. [doi: 10.1080/07448481.2013.773902] [Medline: 25158014]
67. Eisenberg D, Gollust SE, Golberstein E, Hefner JL. Prevalence and correlates of depression, anxiety, and suicidality among university students. Am J Orthopsychiatry 2007 Oct;77(4):534-542. [doi: 10.1037/0002-9432.77.4.534] [Medline: 18194033]
68. Czyz EK, Horwitz AG, Eisenberg D, Kramer A, King CA. Self-reported barriers to professional help seeking among college students at elevated risk for suicide. J Am Coll Health 2013;61(7):398-406 [FREE Full text] [doi: 10.1080/07448481.2013.820731] [Medline: 24010494]
69. Zivin K, Eisenberg D, Gollust SE, Golberstein E. Persistence of mental health problems and needs in a college student population. J Affect Disord 2009 Oct;117(3):180-185. [doi: 10.1016/j.jad.2009.01.001] [Medline: 19178949]
70. Christopher MS, Skillman GD, Kirkhart MW, D'Souza JB. The effect of normative and behavioral persuasion on help seeking in Thai and American college students. J Multicult Couns Devel 2011 Dec 23;34(2):80-93 [FREE Full text] [doi: 10.1002/j.2161-1912.2006.tb00029.x]

71. Hacker K, Arsenault L, Franco I, Shaligram D, Sidor M, Olfson M, et al. Referral and follow-up after mental health screening in commercially insured adolescents. J Adolesc Health 2014 Jul;55(1):17-23. [doi: 10.1016/j.jadohealth.2013.12.012] [Medline: 24525226]
72. App Development. Minder: Student e-Mental Health Project. 2022. URL: https://www.myminder.ca/e-intervention [accessed 2023-08-03]
73. Canadian reference group executive summary Spring 2016. American College Health Association and ACHA-NCHA II. 2016. URL: https://www.acha.org/documents/ncha/NCHA-II%20SPRING%202016%20CANADIAN%20REFERENCE%20GROUP%20EXECUTIVE%20SUMMARY.pdf [accessed 2023-04-15]
74. Faulkner G, Ramanathan S, Kwan M, CCWS Expert Panel Group. Developing a coordinated Canadian post-secondary surveillance system: a Delphi survey to identify measurement priorities for the Canadian Campus Wellbeing Survey (CCWS). BMC Public Health 2019 Jul 11;19(1):935 [FREE Full text] [doi: 10.1186/s12889-019-7255-6] [Medline: 31296190]

Abbreviations

GAD-7: 7-item Generalized Anxiety Disorder
HRQOL: health-related quality of life
MHC-SF: Mental Health Continuum-Short Form
PedsQL-YA: Pediatric Quality of Life-Young Adult
PHQ-9: 9-item Patient Health Questionnaire
SBQ-R: Suicide Behaviors Questionnaire–Revised
UBC: University of British Columbia

Edited by J Torous; submitted 04.05.23; peer-reviewed by E Morton, A Dodd; comments to author 19.05.23; revised version received 08.06.23; accepted 11.06.23; published 09.08.23.

Please cite as:
Virk P, Arora R, Burt H, Finnamore C, Gadermann A, Barbic S, Doan Q
Evaluating the Psychometric Properties and Clinical Utility of a Digital Psychosocial Self-Screening Tool (HEARTSMAP-U) for Postsecondary Students: Prospective Cohort Study
JMIR Ment Health 2023;10:e48709
URL: https://mental.jmir.org/2023/1/e48709
doi:10.2196/48709
PMID:37556180

Original Paper

Efficacy of an Internet-Delivered Intervention for Improving Insomnia Severity and Functioning in Veterans: Randomized Controlled Trial

Sarra Nazem[1,2], PhD; Sean M Barnes[2,3], PhD; Jeri E Forster[2,4], PhD; Trisha A Hostetter[2], MPH; Lindsey L Monteith[2,3,4], PhD; Emily B Kramer[2], BA; Laurel A Gaeddert[2], MS; Lisa A Brenner[2,3,4,5], PhD

[1]Dissemination & Training Division, National Center for Posttraumatic Stress Disorder, Menlo Park, CA, United States

[2]Veterans Affairs Rocky Mountain Mental Illness Research, Education and Clinical Center, Aurora, CO, United States

[3]Department of Psychiatry, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

[4]Department of Physical Medicine and Rehabilitation, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

[5]Department of Neurology, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

Corresponding Author:
Sarra Nazem, PhD
Dissemination & Training Division
National Center for Posttraumatic Stress Disorder
795 Willow Road
Building 334
Menlo Park, CA, 94025
United States
Phone: 1 650 796 8208
Fax: 1 650 617 2701
Email: Sarra.Nazem@va.gov

Abstract

Background: Despite a growing evidence base that internet-delivered cognitive behavioral therapy for insomnia (iCBT-I) is associated with decreased insomnia severity, its efficacy has been minimally examined in veterans.

Objective: The objective of this study was to evaluate the efficacy of an unguided iCBT-I (Sleep Healthy Using the Internet [SHUTi]) among veterans.

Methods: We conducted a single-blind, randomized controlled trial in Operation Enduring Freedom, Operation Iraqi Freedom, and Operation New Dawn veterans eligible for Veterans Health Administration care. Participants were randomly assigned (1:1) to receive SHUTi (a self-guided and interactive program) or an Insomnia Education Website (IEW) that provided nontailored and fixed insomnia information. Web-based assessments were administered at baseline, postintervention, 6 months postintervention, and 1 year postintervention. The primary outcome was self-reported insomnia severity (Insomnia Severity Index [ISI]). Secondary outcomes were self-reported mental and physical health functioning (Veterans RAND 36-item Health Survey). Exploratory outcomes comprised sleep diary parameters.

Results: Of the 231 randomized participants (mean age 39.3, SD 7.8 years; 170/231, 73.5% male sex; 26/231, 11.3% Black; 172/231, 74.5% White; 10/231, 4.3% multiracial; and 17/231, 7.4% other; 36/231, 15.6% Hispanic) randomized between April 2018 and January 2019, a total of 116 (50.2%) were randomly assigned to SHUTi and 115 (49.8%) to the IEW. In intent-to-treat analyses, SHUTi participants experienced significantly larger ISI decreases compared with IEW participants at all time points (generalized η^2 values of 0.13, 0.12, and 0.10, respectively; all $P<.0001$). These corresponded to estimated larger differences in changes of –3.47 (95% CI –4.78 to –2.16), –3.80 (95% CI –5.34 to –2.27), and –3.42 (95% CI –4.97 to 1.88) points on the ISI for the SHUTi group. SHUTi participants experienced significant improvements in physical (6-month generalized $\eta^2=0.04$; $P=.004$) and mental health functioning (6-month and 1-year generalized $\eta^2=0.04$; $P=.009$ and $P=.005$, respectively). Significant sleep parameter improvements were noted for SHUTi (all $P<.05$), though the pattern and magnitude of these reductions varied by parameter. No adverse events were reported.

Conclusions: Self-administered iCBT-I was associated with immediate and long-term improvements in insomnia severity. Findings suggest that leveraging technology to meet insomnia treatment demands among veterans may be a promising approach.

Trial Registration: ClinicalTrials.gov NCT03366870; https://clinicaltrials.gov/ct2/show/NCT03366870

(JMIR Ment Health 2023;10:e50516) doi:10.2196/50516

KEYWORDS

cognitive behavioral therapy; insomnia; internet intervention; online intervention; randomized controlled trial; RCT; RCTs; sleep; treatment; veteran; veterans; veterans' health

Introduction

Insomnia is a public health and clinical problem affecting US veterans. In a nationally representative sample of US veterans, 35% endorsed symptoms consistent with clinical or subthreshold insomnia [1]. This is especially concerning given that insomnia precedes and exacerbates functionally impairing mental health concerns and is associated with suicide risk [2-7]. Chronic insomnia is also linked to an increased risk of chronic diseases (eg, diabetes, cardiovascular diseases, and cancer) [8,9]. To improve veterans' mental and physical health, it is imperative that veterans receive evidence-based psychotherapies (EBPs) that reduce insomnia.

Cognitive behavioral therapy for insomnia (CBT-I) is one of the most efficacious treatments for insomnia [10-12]. Long-term effects surpass those of sleep medications, with the added benefit of being associated with few adverse side effects [13,14]. Despite successful Veterans Health Administration (VHA) efforts that have improved access to CBT-I, the resources necessary to implement provider-delivered CBT-I outstrip treatment demands [15-19]. Self-guided, internet-delivered cognitive behavioral therapy (iCBT) offers a potentially efficacious means of delivering insomnia treatments by providing individuals with the opportunity to engage in an EBP on the web. Veterans have an interest in digital health technologies, including internet-delivered CBT-I (iCBT-I), and leveraging technology is critical to meeting treatment demand and patient preferences [20-24].

From both policy and clinical perspectives, iCBTs have garnered attention due to their potential to overcome barriers to treatment (eg, cost, scheduling and travel demands, stigma, and lack of trained EBP clinicians) [25,26]. Recent meta-analyses found that iCBT-I improved insomnia severity, sleep parameters, and subjective sleep quality, with effect sizes comparable to those found in face-to-face (F2F) CBT-I (ie, Hedges *g* ranging from 0.21 to 1.09) [27,28]. iCBT-I programs that included more CBT-I elements, had longer intervention periods, and used highly interactive program designs optimized outcomes and facilitated better treatment engagement and adherence [29].

Despite veterans' interest in digital health technologies and a growing evidence base for iCBT-I interventions, relatively few studies have examined iCBT-I interventions in veteran populations. Initial, formative studies examined the use of the patient mobile app, CBT-I Coach, which was developed and designed to help facilitate patients' completion of CBT-I tasks while receiving F2F CBT-I. A pilot randomized controlled trial (RCT; n=18) demonstrated the feasibility and acceptability of integrating CBT-I Coach with individual F2F CBT-I [23]. A total of 2 additional small studies (n<35) found that self-administered use of CBT-I Coach paired with supplemental app-delivered worksheets and a self-management guide was associated with a reduction of insomnia symptoms in veterans reporting chronic insomnia [30,31]. Researchers have also demonstrated the feasibility and potential efficacy of stand-alone, self-guided iCBT-I interventions in 3 studies. First, in an uncontrolled pre-post intervention study of a 6-session self-administered iCBT-I, veterans seeking care in a Department of Veterans Affairs (VA) substance use disorder outpatient clinic achieved clinically significant improvements in insomnia severity following the iCBT-I intervention [32]. In a pilot RCT, veterans with probable insomnia disorder (n=50) were randomized to Insomnia Coach, a self-management CBT-I app, or a waitlist control. Veterans found Insomnia Coach to be feasible and acceptable, with preliminary efficacy data suggesting that app use was associated with clinically significant improvement in insomnia severity [33]. Finally, using a cohort trial design, Hermes and colleagues [34] examined the feasibility of implementing Sleep Healthy Using the Internet (SHUTi) in VA primary care clinics. Although the primary aim of this study was to examine differences in implementation outcomes based on low-intensity implementation strategy modifications, improved clinical outcomes were observed, with 20% of veterans achieving clinically meaningful change in their insomnia symptoms.

Previous studies examining iCBT-I interventions, including apps, in veteran populations have established the acceptability and feasibility of digital health interventions, with preliminary evidence suggesting that iCBT-I interventions are associated with improved insomnia outcomes as observed in nonveteran populations. Most of these studies, however, were not powered for efficacy outcomes, and no RCT has compared an evidence-based iCBT-I with a nonwaitlist control in a veteran population. Further, no study has conducted long-term follow-up beyond 12 weeks, and most studies have enrolled veterans without a confirmed insomnia diagnosis. To strengthen the evidence base, this study was designed to evaluate the efficacy of an evidence-based iCBT-I intervention, SHUTi, to decrease insomnia severity among Operation Enduring Freedom (OEF), Operation Iraqi Freedom (OIF), and Operation New Dawn (OND) veterans. Although SHUTi is one of the most empirically supported iCBT-I programs in nonveteran populations [35-44], to our knowledge, no RCT has been conducted to examine the efficacy of SHUTi in a veteran sample. We hypothesized that SHUTi participants would show greater improvements in insomnia severity and mental and physical health functioning across all time points, compared with participants randomized to an Insomnia Education Website (IEW) control. We also evaluated whether SHUTi was associated with improvements across sleep diary parameters (exploratory objective).

Methods

Ethical Considerations

The study design and protocol (Multimedia Appendix 1 [2,15,25,29,40-75]) were approved by the following supporting agencies: Colorado Multiple Institutional Review Board (COMIRB 17-0920), Rocky Mountain Regional Veterans Affairs Medical Center Research and Development Committee, and the US Army Medical Research and Development Command Human Research Protection Office (HRPO Log Number A-19051.2; ORHO Log Number E03947.2). Additional information on the informed consent process, data privacy and security measures, and compensation are provided below and in Multimedia Appendix 1.

Trial Design

This is a single-blind (participants blinded to treatment allocation) 2-group (SHUTi vs IEW) longitudinal (4 time points) RCT. Assessments were administered at baseline (T1), 9 weeks postintervention (T2), 6 months postintervention (T3), and 1 year postintervention (T4). Data were collected on the web, and participants provided informed consent before study enrollment. The RCT was registered with ClinicalTrials.gov (NCT03366870).

Randomization

Due to the empirical association between insomnia and suicide risk, randomization was stratified by self-reported suicide attempt history (“yes” or “no”). A statistically generated allocation sequence, using random blocks of 4 and 6, stratified by suicide attempt history, was conducted using SAS PROC PLAN (SAS Institute). Randomization tables were generated by the trial’s statistician and programmed to automatically assign a participant to the next allocation immediately after T1 completion.

Participants

OEF, OIF, and OND veterans aged between 18 and 55 years were eligible to participate if they met the following criteria: current *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) [45] insomnia diagnosis; reliable access to the internet; eligibility to receive VHA care; English-speaking; and ability to provide informed consent. Exclusion criteria included the following: being currently enrolled in another intervention research study, the presence of another untreated sleep disorder, currently receiving psychological treatment for insomnia (excluding sleep medications), a change in schedule or dosage of sleep medications in the past 3 months, and the presence of a condition contraindicated with CBT-I. Examples of conditions contraindicated with CBT-I included untreated seizures or seizure disorder, pregnancy or a plan to become pregnant, irregular shift work, and significant cognitive impairment. Veterans with comorbid psychiatric presentations were eligible to participate in the RCT unless they had used any nonalcohol substances (excluding cannabis) more than once within the past 3 months; endorsed a problematic pattern of alcohol use associated with significant impairment or distress (past 3 months); or had ever been diagnosed with bipolar I disorder, schizophrenia, schizoaffective, or a psychotic disorder. Those with comorbid medical conditions were included unless conditions were deemed active, unstable, or degenerative in a manner expected to influence sleep. Research staff received extensive training on conducting semistructured interviews focused on clinical sleep history to inform differential diagnoses (eg, untreated sleep apnea, periodic limb movement disorder, parasomnia, and circadian rhythm disorder) and structured clinical interviews for DSM-5 Research Version [76] modules (eg, insomnia disorder and alcohol use disorder) to determine RCT eligibility; cases were regularly reviewed by the team and principal investigator to ensure accuracy.

Total compensation for completion of all study assessments was US $208, and the participant payment for each study assessment was as follows: US $1 for each preintervention sleep diary (up to US $14 maximum), US $50 for T1 and T2 assessments, US $1 for each postintervention diary (up to US $14 maximum), and US $40 for T3 and T4 assessments.

Outcomes

The primary outcome measure was the Insomnia Severity Index (ISI) [77], one of the most empirically examined and clinically used self-report scales assessing insomnia symptoms. The ISI is comprised of 7 items that assess current (ie, past 2 weeks) insomnia severity. Each item is scored on a 0-4 scale, with total scores ranging from 0-28; each 7-point increment is associated with increasing levels of insomnia severity. The ISI has been shown to have adequate test-retest reliability over 3 months and concurrent validity with sleep diaries and polysomnography [46] and has been validated for web-based use [78]. In the current sample, the ISI demonstrated adequate internal consistency (Cronbach α ranged from .78 to .90 across the 4 time points).

Our secondary outcome measure was the Veterans RAND 36-item Health Survey (VR-36) [79]. The VR-36 assesses perceived general health-related quality of life; respondents indicate the impact of health on physical functioning, role limitations due to physical problems, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional problems, and mental health. The VR-36 was formerly called the Short Form Health Survey for Veterans (Veterans SF-36) [80]. The VR-36 has been thoroughly validated and has strong psychometric properties [79,80]. Scores are used to derive 2 summary scales (ie, Physical Component Summary [PCS] and Mental Component Summary [MCS]). Scores range from 0-100, with a mean score of 50 (SD 10). Lower scores indicate poorer health-related quality of life.

Exploratory Outcomes

Daily sleep diaries based on consensus sleep diary [81] questions were collected for 2 weeks before T1 (preintervention sleep diaries) and immediately following T2 (postintervention sleep diaries). Sleep parameters included sleep onset latency (SOL; time to fall asleep), wake after sleep onset (WASO) + early morning awakenings (EMA; combined time awake after falling asleep and before final awakening), time in bed (TIB), total sleep time (TST), and sleep efficiency (percentage of time asleep while in bed).

Clinical Characteristics

A total of 4 psychometrically sound assessments (Multimedia Appendix 2 [47,48,82-85]) were used to characterize baseline clinical characteristics: (1) Adult Suicidal Ideation Questionnaire [47], a 25-item self-report instrument that assesses past-month suicidal ideation frequency; (2) Beck Anxiety Inventory [48], a 21-item self-report rating inventory that measures past-week subjective, somatic, and panic-related symptoms of anxiety; (3) Beck Depression Inventory II [82], a 21-item self-report rating inventory that measures past 2-week characteristic attitudes and symptoms of depression; and (4) Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) [83], a 20-item self-report measure that assesses past-month severity of posttraumatic stress disorder symptoms.

Interventions

SHUTi Program

SHUTi is a self-guided, interactive, and tailored internet-delivered program modeled on the primary tenants of F2F CBT-I [44]. Intervention content is delivered through 6 "cores" where users obtain access to a new core based on a time- and event-based schedule. The SHUTi program relies on user-entered web-based sleep diaries to track progress and tailor treatment recommendations. Each SHUTi core acts as a web-based analog for the weekly sessions typically used when delivering CBT-I in a F2F format, following the same general structure: (1) core objectives (what will be learned and why this information is important), (2) review of previous week's homework and sleep diary data, (3) new intervention material, (4) assignment of homework (treatment strategies for the coming week), and (5) a summary of the core's main points. Intervention content is enhanced through a variety of interactive features, including personalized goal setting, graphical feedback based on inputted symptoms, animations and illustrations to enhance comprehension, quizzes to test users' knowledge, patient vignettes, and video-based expert explanation. The SHUTi program relies on user-entered web-based sleep diaries to track progress and tailor treatment recommendations. Automated emails are sent throughout the intervention to encourage program adherence.

IEW Control

The IEW used in this RCT has been previously used in SHUTi trials [37,44]. Informed by what typically constitutes "treatment as usual" for patient education websites targeting insomnia, the IEW provided static information about insomnia symptoms, causes, and strategies to improve sleep (ie, sleep hygiene, stimulus control, and behavioral recommendations). In contrast to SHUTi, IEW strategies were presented as text with no tailoring, interactivity, or feedback.

Procedure

Veterans were screened over the phone for eligibility. Of the 844 veterans who expressed interest in the study, a total of 61 were unable to be reached for screening, and 494 were ineligible for participation. The top reasons for ineligibility were identification of another untreated sleep disorder (162/494, 32.8%); not serving in the OEF, OIF, or OND conflict (112/494, 22.7%); multiple exclusion criteria (54/494, 10.9%); and not meeting criteria for insomnia disorder (52/494, 10.5%). After screening, a total of 289 veterans were eligible for the study, and 250 were enrolled.

Most participants (234/250, 93.6%) enrolled were recruited from targeted mailings sent to veterans with diagnosed sleep problems, as identified by medical codes in the VA's Corporate Data Warehouse. An additional 3.6% (9/250) of enrolled participants were recruited based on their expressed interest in learning more about the study. Participants were enrolled in the RCT between April 2018 and January 2019. The final T4 assessment was completed in June 2020.

After providing informed consent, participants were given instructions on how to complete a web-based, 2-week sleep diary assessment and informed that they had to complete 10 out of 14 sleep diaries to move forward in the study; a total of 6% (15/250) of enrolled veterans did not complete the sleep diary requirement. Baseline assessments (T1) were unlocked once the requirement was fulfilled. A total of 3 participants did not complete T1 assessments, resulting in a total of 231 participants (92.4%) who were randomized.

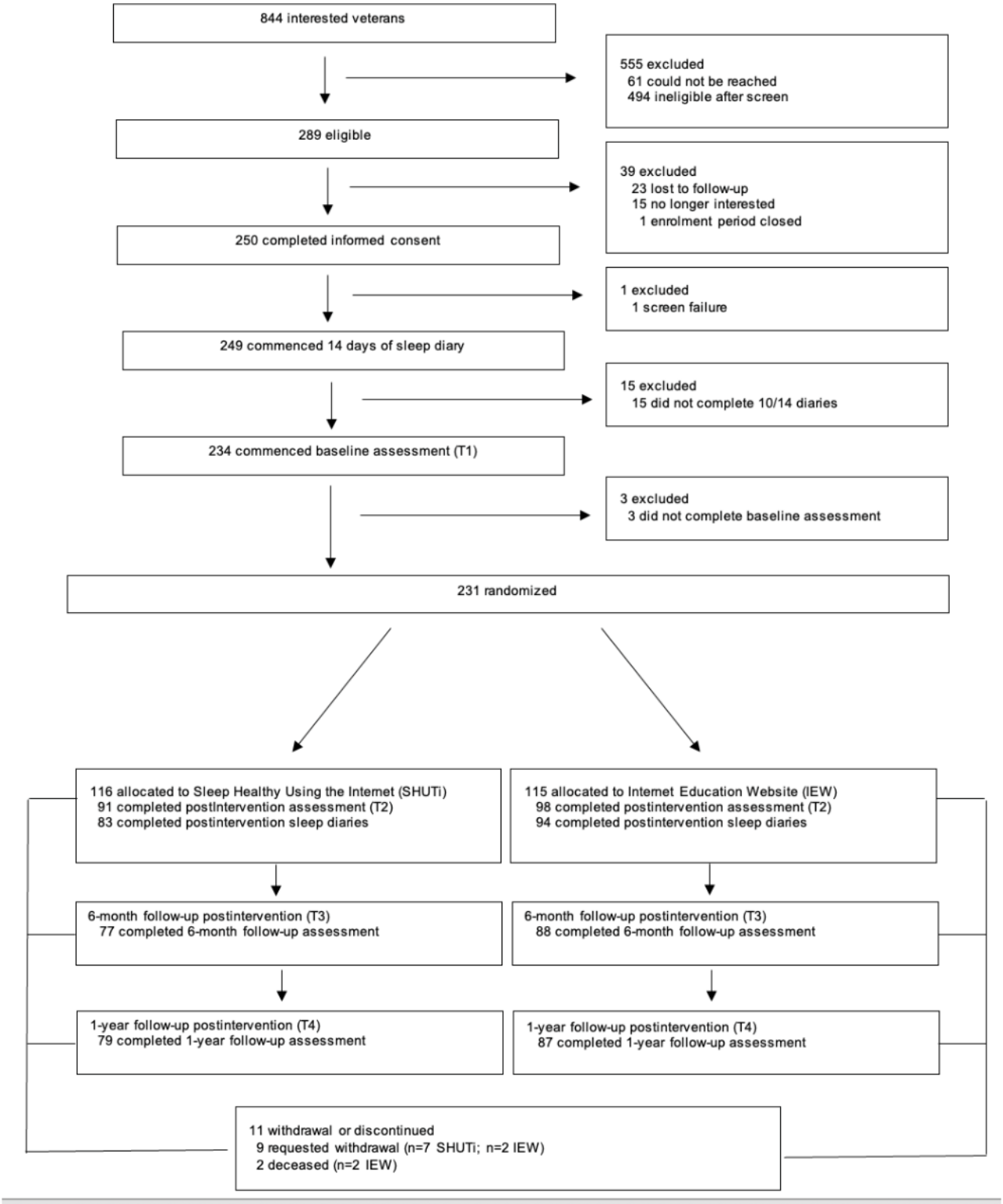
A total of 84.8% (196/231) of all randomized participants completed at least 1 postintervention assessment, of which 151 participants completed all 3 assessments (77%), another 21 completed 2 assessments (10.7%), and 24 completed 1 assessment (12.2%). A total of 9 participants asked to be withdrawn from the study: 1 during the intervention, 2 at T2, 5 at T3, and 1 at T4. A total of 2 participants randomized to the IEW died during the study (deaths unrelated to research procedures); these deaths occurred between T2 and T3, and T3 and T4. Figure 1 depicts the study enrollment flow.

Participants received reminder calls from study staff to complete initial sleep diaries and the T1 assessment. During the 9-week intervention window, study staff did not initiate any contact with participants, so participants only received automated emails built into the interventions during the intervention window. Participants received reminder calls, automated emails, and reminder postcards to support postintervention (T2-T4) assessment completion.

Participants were compared across demographic variables, ISI, and VR-36 summary scale scores to examine whether there were differences between participants who did not complete a postintervention assessment (n=35) and participants who completed at least 1 postintervention assessment (n=196). The only statistically significant difference observed was that participants who did not complete a postintervention assessment were younger (mean 35.4, SD 6.4 years) than participants who completed at least 1 postintervention assessment (mean 40.0, SD 7.8 years; 2-tailed $t_{229}=3.33$; $P=.001$).

Of the 35 participants who did not complete a postintervention assessment, a total of 20 randomized to SHUTi demonstrated significantly lower VR-36 MCS scores (mean 31.9, SD 8.2) than the 15 randomized to the IEW (mean 40.1, SD 12.6; 2-tailed $t_{33}=2.35$; $P=.02$). No other statistically significant differences were observed between groups.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram of study enrollment flow. IEW: Insomnia Education Website; SHUTi: Sleep Healthy Using the Internet.



Statistical Methods

Because data were collected on the web, steps were taken to verify data quality before conducting analyses, including examining abnormally low survey completion times and surveys with potentially biased or inconsistent responses. No inconsistent, biased, or inaccurate response patterns were identified in the data set. Missing data was minimal across measures and was handled in accordance with manual instructions. Sleep diaries were excluded from analyses if participants provided too few (less than 7) usable sleep diaries during the sleep diary assessment window (T1, n=2; and T2, n=4).

We used any pair power and Holm sequential procedure [86] to control the familywise error rate, considering all 3 primary (ISI: T2-T4) and 6 secondary outcomes (PCS: T2-T4, and MCS:

T2-T4). A final sample size of at least 170 completing T2 assessments (85 per group) would provide 80% power to detect an effect size of 0.6 for each outcome. Assuming an attrition rate of 25%, we sought to randomize at least 226 participants to achieve at least 85 per group completing T2 assessments.

The analysis of covariance was used to model the change from T1 to T2, from T1 to T3, and from T1 to T4 for each outcome as a function of group, the baseline value of the respective outcome, and suicide attempt (stratification variable). Generalized η^2 [87] is reported with all analysis of covariance results, where values of 0.01 are considered small, 0.06 as medium, and 0.14 as large. Fisher exact tests were used for post hoc analyses that examined the clinical significance of ISI changes at each postintervention assessment. All analyses were conducted using SAS (version 9.4; SAS Institute) statistical software [88].

Results

Overview

There were no significant differences between groups on demographic or baseline clinical characteristics (Table 1 and Multimedia Appendix 2). At T1, participants on average endorsed moderate insomnia and reported deficits in functioning, with PCS and MCS scores about 1 SD below the normative US population sample but consistent with the VHA normative sample. Table 2 contains the ISI, PCS, and MCS descriptive statistics over time.

Table 1. Demographic and baseline clinical characteristics.

	Sleep Healthy Using the Internet (n=116)	Insomnia Education Website (n=115)	*P* value
Age (years), mean (SD)	39.1 (8.0)	39.5 (7.6)	.72
Male sex, n (%)	86 (74)	86 (7)	.91
Gender[a], n (%)			.88[b]
Man	84 (73)	86 (75)	
Woman	30 (26)	29 (25)	
Transwoman	1 (1)	0 (0)	
Race[c], n (%)			.92
Black or African American	13 (12)	13 (12)	
White or Caucasian	85 (76)	87 (77)	
Multiracial	6 (5)	4 (4)	
Other	8 (7)	9 (8)	
Hispanic[d], n (%)	16 (14)	20 (18)	.45
Relationship status, n (%)			.74
Married	67 (58)	67 (58)	
Single	16 (14)	21 (18)	
Cohabitating	11 (9)	9 (8)	
Widowed, divorced, or separated	22 (19)	18 (16)	
Employment, n (%)			.96
Full time	68 (59)	65 (57)	
Part time	9 (8)	8 (7)	
Unemployed and not seeking	10 (9)	10 (9)	
Unemployed and seeking	13 (11)	12 (10)	
Retired	16 (14)	20 (17)	
Deployed	99 (85)	106 (92)	.10
Combat[a]	87 (76)	87 (76)	>.99
Era (not mutually exclusive), n (%)			
Post-Vietnam	10 (9)	9 (8)	.83
Desert Storm	34 (29)	29 (25)	.48
OEF[e] or OIF[f]	113 (97)	113 (98)	>.99[b]
Other	1 (1)	1 (1)	>.99[b]
History of a suicide attempt, mean (SD)	23 (20)	23 (20)	.97
ISI[g] total score, mean (SD)	16.9 (4.6)	16.3 (4.3)	.35
Veterans SF-36[h] PCS[i], mean (SD)	41.8 (10.1)	42.9 (9.7)	.40
Veterans SF-36 MCS[j], mean (SD)	37.0 (12.4)	37.3 (11.6)	.86
ASIQ[k] total score[d], mean (SD)	14.4 (20.4)	13.2 (18.4)	.64
BAI[l] total score, mean (SD)	15.6 (13.4)	15.3 (12.5)	.83
BDI[m] total score[n], mean (SD)	20.8 (12.6)	19.7 (12.1)	.49
PCL-5[o] total sore, mean (SD)	34.6 (18.6)	33.0 (19.3)	.51
Provisional PTSD[p], n (%)	66 (57)	53 (46)	.10

	Sleep Healthy Using the Internet (n=116)	Insomnia Education Website (n=115)	*P* value
Probable PTSD, n (%)	61 (53)	53 (46)	.32

[a]n=115, Sleep Healthy Using the Internet.
[b]Fisher exact test.
[c]n=112, Sleep Healthy Using the Internet; n=113, Insomnia Education Website.
[d]n=115, Sleep Healthy Using the Internet; n=114, Insomnia Education Website.
[e]OEF: Operation Enduring Freedom.
[f]OIF: Operation Iraqi Freedom.
[g]ISI: Insomnia Severity Index.
[h]Veteran SF-36: Short Form Health Survey for Veterans.
[i]PCS: Physical Component Summary.
[j]MCS: Mental Component Summary.
[k]ASIQ: Adult Suicidal Ideation Questionnaire.
[l]BAI: Beck Anxiety Inventory.
[m]BDI: Beck Depression Inventory.
[n]n=114, Insomnia Education Website.
[o]PCL-5: Posttraumatic Stress Disorder Checklist for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
[p]PTSD: posttraumatic stress disorder.

Table 2. Primary and secondary outcome descriptive statistics over time.

Outcome	Time 1		Time 2		Time 3		Time 4	
	n	Frequency	n	Frequency	n	Frequency	n	Frequency
Sleep Healthy Using the Internet								
ISI[a], mean (SD)	116	16.9 (4.6)	91	11.6 (6.7)	77	10.5 (6.8)	79	10.7 (6.7)
ISI categorical, n (%)	116		91		77		79	
0-7		1 (1)		28 (31)		29 (38)		32 (41)
8-14		40 (34)		33 (36)		27 (35)		26 (33)
15-21		54 (47)		23 (25)		15 (19)		15 (19)
22-28		21 (18)		7 (8)		6 (8)		6 (8)
PCS[b], mean (SD)	116	41.8 (10.1)	91	42.5 (10.5)	77	43.7 (10.6)	78	42.6 (10.2)
MCS[c], mean (SD)	116	37.0 (12.4)	91	42.5 (13.4)	77	43.9 (13.8)	78	43.7 (13.2)
Insomnia Education Website								
ISI, mean (SD)	115	16.3 (4.3)	98	14.9 (4.9)	88	13.9 (5.1)	87	13.8 (4.8)
ISI categorical, n (%)	115		98		88		87	
0-7		2 (2)		4 (4)		6 (7)		11 (13)
8-14		40 (35)		45 (46)		48 (55)		40 (46)
15-21		57 (50)		41 (42)		25 (28)		29 (33)
22-28		16 (14)		8 (8)		9 (10)		7 (8)
PCS, mean (SD)	115	42.9 (9.7)	97	41.8 (9.7)	88	42.3 (8.9)	87	42.0 (10.4)
MCS, mean (SD)	115	37.3 (11.6)	97	38.7 (12.5)	88	38.6 (11.7)	87	38.2 (12.2)

[a]ISI: Insomnia Severity Index.
[b]PCS: Physical Component Summary of the Veterans RAND 36-item Health Survey.
[c]MCS: Mental Component Summary of the Veterans RAND 36-item Health Survey.

Primary and Secondary Outcomes

SHUTi participants experienced a significantly larger T1 to T2 (–3.47, 95% CI –4.78 to –2.16 points; *P*<.0001), T1 to T3 (–3.80, 95% CI –5.34 to –2.27 points; *P*<.0001) and T1 to T4 (–3.42, 95% CI –4.97 to –1.88 points; *P*<.0001) decreases on the ISI compared with IEW participants. The magnitude of the

results was maintained across all time points (generalized η^2 values of 0.13, 0.12, and 0.10; Table 3).

SHUTi participants experienced a significantly greater T1 to T3 improvement on the VR-36 PCS (3.22, 95% CI 1.07-5.36 points; generalized η^2=0.04; *P*=.004) compared with IEW participants (Table 3). Significance was not observed for the change from T1 to T2 (*P*=.03) or from T1 to T4 (*P*=.14). For the VR-36 MCS, SHUTi participants had a significantly greater T1 to T3 (4.33, 95% CI 1.11-7.55 points; generalized η^2=0.03; *P*=.009) and T1 to T4 (4.53, 95% CI 1.39-7.67 points; generalized η^2=0.04; *P*=.005) improvement, but the difference in change from T1 to T2 did not achieve statistical significance (2.58, 95% CI –0.07 to 5.24 points; *P*=.06; Table 3).

Table 3. Estimated difference in change between groups. All models control for the baseline value of the outcome and history of suicide attempt ("yes" or "no"). Dependent variables were calculated as the later time point minus baseline, and the difference in change was estimated as Sleep Healthy Using the Internet minus Insomnia Education Website. The italicized text indicates statistical significance based on the Holm sequential procedure, considering all 9 outcomes.

Models	Change time 1 to time 2[a,b]			Change time 1 to time 3[c]			Change time 1 to time 4[d,e]		
	Estimated difference between groups, SE	Generalized η^2	*P* value	Estimated difference between groups, SE	Generalized η^2	*P* value	Estimated difference between groups, SE	Generalized η^2	*P* value
ISI[f]	–3.47 (0.66)	0.13	*<.0001*	–3.80 (0.78)	0.12	*<.0001*	–3.42 (0.78)	0.10	*<.0001*
PCS[g]	2.07 (0.96)	0.02	.03	3.22 (1.08)	0.04	*.004*	1.81 (1.22)	0.01	.14
MCS[h]	2.58 (1.35)	0.02	.06	4.33 (1.63)	0.04	*.009*	4.53 (1.59)	0.04	*.005*

[a]n=98, Insomnia Education Website (for ISI); n=91, Sleep Healthy Using the Internet (for ISI).

[b]n=97, Insomnia Education Website (for PCS and MCS); n=91, Sleep Healthy Using the Internet (for PCS and MCS).

[c]n=88, Insomnia Education Website (for all outcomes); n=77, Sleep Healthy Using the internet (for all outcomes).

[d]n=87, Insomnia Education Website (for ISI); n=79, Sleep Healthy Using the Internet (for ISI).

[e]n=87, Insomnia Education Website (for PCS and MCS); n=78, Sleep Healthy Using the Internet (for PCS and MCS).

[f]ISI: Insomnia Severity Index.

[g]PCS=Physical Component Summary of the Veterans RAND 36-item Health Survey.

[h]MCS=Mental Component Summary of the Veterans RAND 36-item Health Survey.

Exploratory Outcomes: Sleep Diary Parameters

Results for the sleep diary parameters are displayed in Table 4. SHUTi participants experienced a significantly greater T1 to T2 decrease in SOL (*P*=.0001), WASO+EMA (*P*=.006), TIB (*P*<.0001), and TST (*P*=.04), compared with IEW participants, and a significantly greater improvement in sleep efficiency (*P*=.02). Specifically, SOL decreased by 16.9 (95% CI –25.6 to –8.3; generalized η^2=0.04) more minutes for SHUTi participants, WASO+EMA decreased by 15.3 (95% CI –26.1 to –4.4; generalized η^2=0.03) minutes more, TIB decreased by 48 (95% CI –65.4 to –30.0; generalized η^2=0.12) minutes more, and TST decreased by 18.6 (95% CI –36.6 to –0.60; generalized η^2=0.02) minutes more. Sleep efficiency improved for SHUTi participants 3.74% (95% CI 0.72%-6.75%; generalized η^2=0.03) more than for the IEW participants.

Table 4. Exploratory sleep parameter models. All models control for the baseline value of the outcome and history of suicide attempt ("yes" or "no"). Dependent variables were calculated as the later time point minus baseline, and the difference in change was estimated as Sleep Healthy Using the Internet minus Insomnia Education Website. The italicized text indicates statistical significance at the *P*<.05 level.

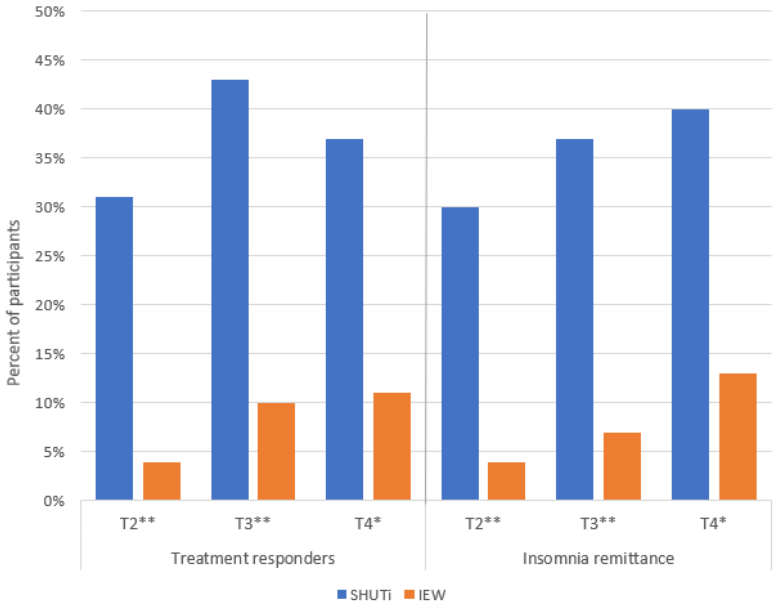
Outcome	Change Time 1 to Time 2[a]		
	Group parameter estimate, SE	Generalized η^2	*P* value
Sleep onset latency (minutes)	–16.95 (4.37)	0.04	*.0001*
Wake after sleep onset + early morning awakenings (minutes)	–15.26 (5.51)	0.03	*.006*
Time in bed (hours)	–0.80 (0.15)	0.12	*<.0001*
Total sleep time (hours)	–0.31 (0.15)	0.02	*.04*
Sleep efficiency (%)	3.74 (1.53)	0.03	*.02*

[a]n=89, Insomnia Education Website; n=83, Sleep Healthy Using the Internet.

Post Hoc Analyses

A significantly greater proportion of the SHUTi group were deemed treatment responders (reduction of greater than 7 points on the ISI) at all time points compared with IEW participants (31% vs 4% at T2; 43% vs 10% at T3, and 37% vs 11% at T4; all *P*<.001; Figure 2). Additionally, a greater proportion of the SHUTi group met criteria for insomnia remittance (change from ≥8 at T1 to <8 on the ISI at T2, T3, and T4) at all time points compared with IEW participants (30% vs 4% at T2; 37% vs 7% at T3, and 40% vs 13% at T4; all *P*<.001; Figure 2).

Figure 2. Treatment responders and insomnia remittance by group. Treatment responders are defined as a reduction of >7 points on the Insomnia Severity Index (ISI) between T2, T3, and T4, respectively, and T1. Insomnia remittance is defined as an ISI change from ≥8 at T1 to <8 at T2, T3, and T4. IEW: Insomnia Education Website; SHUTI: Sleep Healthy Using the Internet. **P*<.001; ***P*<.0001.



Discussion

Principal Findings

Despite the empirical support for CBT-I as an efficacious first-line treatment for insomnia and evidence that iCBT-I improves insomnia severity with comparable effect sizes to those found in F2F CBT-I, there is a paucity of RCTs examining iCBT-I in veterans. Data from this RCT confirmed the efficacy of SHUTi for the treatment of insomnia in OEF, OIF, and OND veterans, and this is the first study to demonstrate that improvements in insomnia severity are sustained up to 1 year postintervention. SHUTi participants reported significant changes in insomnia severity across all time points compared with IEW participants. These effects were medium-large and consistent with civilian SHUTi trials, an iCBT-I trial [89] in a military sample, and self-guided iCBT-I interventions in veteran samples [32-34]. SHUTi participants also experienced significant improvements in sleep diary parameters compared with IEW participants. Although SHUTi participants experienced a reduction in total sleep time compared with IEW participants, they spent less time in bed and improved their sleep efficiency, which are consistent with CBT-I desired outcomes.

IEW participants completed pre- and postintervention sleep diaries and had access to high-quality, patient-oriented insomnia materials, constituting a more "active" control compared with waitlist designs. For example, behavioral tracking can facilitate awareness of patterns that foster insight and promote behavior change. Given this, the fact that SHUTi participants significantly outperformed IEW participants in reducing their insomnia severity demonstrates the value of metered, interactive, and tailored iCBT-I content. These findings add to the robust empirical data that supports the superiority of CBT-I interventions over approaches such as sleep hygiene education. This study extends those by demonstrating the efficacy of self-administered iCBT-I in a veteran population.

SHUTi participants were also more likely than IEW participants to report significant improvements in health-related quality of life as assessed by the VR-36 during follow-up, though effect sizes were small. These findings suggest that mental and physical health functioning may require sustained improvements in insomnia to ultimately improve functioning. Furthermore, given the many health factors impacting veterans' health-related quality of life, it is possible that small effect sizes were due to improved sleep but the continued negative impact of comorbid issues. Our findings were consistent with those observed in a military sample [90] but differed slightly from findings reported in a systematic review on F2F CBT-I and functioning [91], suggesting potential differences in the insomnia-functioning association across populations. Although our findings fall below the change typically indicative of clinical significance at an individual level, they are above the cutoff for population clinical significance [79,92]. Paired with the insomnia severity findings, this suggests that health care implementation of iCBT-I could be an especially promising approach to improving veteran outcomes (eg, systems level).

SHUTi participants were more likely to achieve clinically meaningful changes in insomnia severity compared with IEW

participants across all time points, consistent with improvements observed at posttreatment by veterans following use of Insomnia Coach [33]. The proportion of treatment responders (31%) and those who achieved remittance (30%) at posttreatment was greater in our SHUTi sample compared with an implementation-focused study of SHUTi in VA primary care (20% responders, 18.6% remitted) [34]. These differences may have been driven by sample characteristics, as participants in this study were younger (mean age of 39.3, SD 7.8 years) than those in the study of Hermes et al [34] (mean age of 67, SD 16.9 years). Furthermore, participants in this study may have been more motivated to engage in the iCBT-I as enrollment originated from self-interest versus provider referral. Despite SHUTi participants achieving clinically meaningful changes in insomnia, responder and remittance percentages were lower than what had been observed in nonveteran SHUTi trials. IEW participants, however, were also less likely to naturally recover or receive benefit from the IEW when compared with nonveteran IEW participants, suggesting that differences may be attributable to methodology and sample clinical complexity. Given that this study was not designed to specifically examine comorbidity, future studies are necessary to investigate how clinical comorbidity impacts iCBT-I engagement and benefit.

For self-administered approaches to be a viable health care option, individuals must be interested in using technology to receive an intervention and effectively engage. OEF, OIF, and OND veterans' interest in this study was high, with almost 900 veterans reporting interest in approximately 9 months. Engagement was also high, with 58% of SHUTi participants engaging in the intervention to receive an adequate dose (ie, completion of more than 4 cores) [44]. This finding is notable given there was no clinician or research staff support during the intervention, demonstrating that some veterans can achieve positive outcomes from a self-administered iCBT-I without clinician or coaching support.

Limitations

Although this trial was adequately powered and included a longitudinal design with a high retention rate in an important clinical population, several limitations are important to note. To mitigate primary outcome confounding, there was no ISI inclusion requirement. As such, about one-third of randomized participants reported insomnia severity below the ISI clinically significant threshold. Although CBT-I has been shown to provide benefit in subthreshold insomnia [93], this has not been widely studied in veteran populations and may have resulted in a floor effect. Most participants in this study identified as male, White, and non-Hispanic. Thus, our results may not generalize to veterans with different identities who face additional barriers to accessing medical care. To ensure safety, participants were ineligible if they had untreated sleep apnea (a commonly cooccurring sleep disorder with insomnia) [30,94]; our findings may not generalize to veterans with insomnia and untreated sleep apnea. During the consent process, participants were told that the purpose of the study was to learn whether participation in computerized insomnia programs may reduce insomnia symptoms and improve functioning in veterans (ie, SHUTi and IEW were not described or compared). Although participants were blinded to their treatment allocation, based on their experiences (eg, access to static IEW information vs access to SHUTi, seeking information on ClinicalTrials.gov), they may have speculated that their assigned condition was better or worse than the other. To better understand how participant perceptions of assigned conditions impact outcomes, asking participants what condition they believe they received would be of benefit to future iCBT-I trials. Given the RCT design (ie, no initiated contact during the intervention), we do not have a reliable way to ascertain the duration or engagement in SHUTi or the IEW. Although this design improves the ecological validity of our findings, it limits our ability to determine whether there were moderating factors associated with trajectories of improvement. Future research examining levels of support will be integral in determining CBT-I treatment options (ie, clinician-delivered CBT-I and clinician-supported and self-administered iCBT-I) [33,95-97]. Although participants must have experienced insomnia symptoms for at least 3 months to meet the inclusion criteria, we did not systematically collect information on insomnia symptom duration, sleep medication use, or medical comorbidity; this information would have further characterized the sample and will be important to include in future iCBT-I research. Finally, although commonly used as an important indicator, sleep outcomes were based on self-report, with no objective indicators. Future research examining potential differences in subjective versus objective outcomes following SHUTi will be critical to understanding the mechanisms of change.

Conclusions

This study is the first known longitudinal RCT comparing an evidence-based iCBT-I to a nonwaitlist control in a veteran population meeting diagnostic criteria for insomnia disorder. In this RCT, we observed significant reductions in the primary outcome of insomnia severity between groups immediately after the intervention and up to 1 year postintervention. Although the secondary outcomes of improved physical and mental health functioning were not detected immediately after the 9-week intervention, SHUTi was efficacious in improving mental health functioning at both follow-up time points and physical health functioning at 6 months. In exploratory analyses, participants randomized to SHUTi experienced significant improvements across sleep parameters. The results of this RCT support the use of a self-administered iCBT-I in OEF, OIF, and OND veterans who meet criteria for insomnia disorder.

Acknowledgments

The authors wish to acknowledge Christine Jackson, BA, for her contributions to the success of this research, and Lee Ritterband, PhD, for his consultation and support in the execution of this research.

This research was funded by the Military Suicide Research Consortium (principal investigator SN; W81XWH-16-2-0004) and supported with resources and the use of facilities at the Rocky Mountain Mental Illness Research, Education, and Clinical Center (MIRECC), which is located within the Department of Veterans Affairs (VA) Rocky Mountain Regional Medical Center.

Conflicts of Interest

SN reports grants from the VA and editorial board membership and remuneration from Wiley Periodicals, LLC. LAB reports grants from the VA, the Department of Defense (DOD), the National Institutes of Health (NIH), and the State of Colorado; editorial remuneration from Wolters Kluwer and the Rand Corporation; and royalties from the American Psychological Association and Oxford University Press. In addition, she consults with sports leagues through her university affiliation. JEF reports grants from the VA, DOD, NIH, and the State of Colorado. SMB reports grants from the VA. LLM reports grants from the VA and DOD.

Multimedia Appendix 1
Trial protocol.
[DOCX File , 82 KB - mental_v10i1e50516_app1.docx]

Multimedia Appendix 2
Exploratory measures.
[DOCX File , 17 KB - mental_v10i1e50516_app2.docx]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1196 KB - mental_v10i1e50516_app3.pdf]

References

1. Byrne SP, McCarthy E, DeViva JC, Southwick SM, Pietrzak RH. Prevalence, risk correlates, and health comorbidities of insomnia in US military veterans: results from the 2019-2020 national health and resilience in veterans study. J Clin Sleep Med 2021;17(6):1267-1277 [FREE Full text] [doi: 10.5664/jcsm.9182] [Medline: 33656983]
2. Ohayon MM, Roth T. Place of chronic insomnia in the course of depressive and anxiety disorders. J Psychiatr Res 2003;37(1):9-15. [doi: 10.1016/s0022-3956(02)00052-3] [Medline: 12482465]
3. Ulmer CS, van Voorhees E, Germain AE, Voils CI, Beckham JC, VA Mid-Atlantic Mental Illness Research Education and Clinical Center Registry Workgroup. A comparison of sleep difficulties among Iraq/Afghanistan theater veterans with and without mental health diagnoses. J Clin Sleep Med 2015;11(9):995-1005 [FREE Full text] [doi: 10.5664/jcsm.5012] [Medline: 26094928]
4. Pigeon WR, Britton PC, Ilgen MA, Chapman B, Conner KR. Sleep disturbance preceding suicide among veterans. Am J Public Health 2012;102(Suppl 1):S93-S97 [FREE Full text] [doi: 10.2105/AJPH.2011.300470] [Medline: 22390611]
5. Malik S, Kanwar A, Sim LA, Prokop LJ, Wang Z, Benkhadra K, et al. The association between sleep disturbances and suicidal behaviors in patients with psychiatric diagnoses: a systematic review and meta-analysis. Syst Rev 2014;3(1):18 [FREE Full text] [doi: 10.1186/2046-4053-3-18] [Medline: 24568642]
6. Littlewood D, Kyle SD, Pratt D, Peters S, Gooding P. Examining the role of psychological factors in the relationship between sleep problems and suicide. Clin Psychol Rev 2017;54:1-16 [FREE Full text] [doi: 10.1016/j.cpr.2017.03.009] [Medline: 28371648]
7. Zuromski KL, Cero I, Witte TK. Insomnia symptoms drive changes in suicide ideation: a latent difference score model of community adults over a brief interval. J Abnorm Psychol 2017;126(6):739-749. [doi: 10.1037/abn0000282] [Medline: 28557509]
8. Zhang Y, Jiang X, Liu J, Lang Y, Liu Y. The association between insomnia and the risk of metabolic syndrome: a systematic review and meta-analysis. J Clin Neurosci 2021;89:430-436 [FREE Full text] [doi: 10.1016/j.jocn.2021.05.039] [Medline: 34074598]
9. Taylor DJ, Mallory LJ, Lichstein KL, Durrence HH, Riedel BW, Bush AJ. Comorbidity of chronic insomnia with medical problems. Sleep 2007;30(2):213-218 [FREE Full text] [doi: 10.1093/sleep/30.2.213] [Medline: 17326547]
10. Siebern AT, Manber R. New developments in cognitive behavioral therapy as the first-line treatment of insomnia. Psychol Res Behav Manag 2011;4:21-28 [FREE Full text] [doi: 10.2147/PRBM.S10041] [Medline: 22114532]
11. Mysliwiec V, Martin JL, Ulmer CS, Chowdhuri S, Brock MS, Spevak C, et al. The management of chronic insomnia disorder and obstructive sleep apnea: synopsis of the 2019 U.S. Department of Veterans Affairs and U.S. Department of Defense Clinical Practice Guidelines. Ann Intern Med 2020;172(5):325-336 [FREE Full text] [doi: 10.7326/M19-3575] [Medline: 32066145]

12. Taylor DJ, Pruiksma KE. Cognitive and behavioural therapy for insomnia (CBT-I) in psychiatric populations: a systematic review. Int Rev Psychiatry 2014;26(2):205-213 [FREE Full text] [doi: 10.3109/09540261.2014.902808] [Medline: 24892895]
13. National Institutes of Health. NIH state-of-the-science conference statement on manifestations and management of chronic insomnia in adults. NIH Consens State Sci Statements 2005;22(2):1-30 [FREE Full text] [Medline: 17308547]
14. Sweetman A, Putland S, Lack L, McEvoy RD, Adams R, Grunstein R, et al. The effect of cognitive behavioural therapy for insomnia on sedative-hypnotic use: a narrative review. Sleep Med Rev 2021;56:101404 [FREE Full text] [doi: 10.1016/j.smrv.2020.101404] [Medline: 33370637]
15. Edinger JD, Means MK. Cognitive-behavioral therapy for primary insomnia. Clin Psychol Rev 2005;25(5):539-558 [FREE Full text] [doi: 10.1016/j.cpr.2005.04.003] [Medline: 15951083]
16. Manber R, Carney C, Edinger J, Epstein D, Friedman L, Haynes PL, et al. Dissemination of CBTI to the non-sleep specialist: protocol development and training issues. J Clin Sleep Med 2012;8(2):209-218 [FREE Full text] [doi: 10.5664/jcsm.1786] [Medline: 22505869]
17. Trockel M, Karlin BE, Taylor CB, Manber R. Cognitive behavioral therapy for insomnia with veterans: evaluation of effectiveness and correlates of treatment outcomes. Behav Res Ther 2014;53:41-46 [FREE Full text] [doi: 10.1016/j.brat.2013.11.006] [Medline: 24412462]
18. Koffel E, Farrell-Carnahan L. Feasibility and preliminary real-world promise of a manualized group-based cognitive behavioral therapy for insomnia protocol for veterans. Mil Med 2014;179(5):521-528 [FREE Full text] [doi: 10.7205/milmed-d-13-00455]
19. Sarmiento K, Rossettie J, Stepnowsky C, Atwood C, Calvitti A, VA Sleep Network. The state of Veterans Affairs sleep medicine programs: 2012 inventory results. Sleep Breath 2016;20(1):379-382 [FREE Full text] [doi: 10.1007/s11325-015-1184-0] [Medline: 25924933]
20. Bramoweth AD, Germain A. Deployment-related insomnia in military personnel and veterans. Curr Psychiatry Rep 2013;15(10):401 [FREE Full text] [doi: 10.1007/s11920-013-0401-4] [Medline: 24005883]
21. Gutner CA, Pedersen ER, Drummond SPA. Going direct to the consumer: examining treatment preferences for veterans with insomnia, PTSD, and depression. Psychiatry Res 2018;263:108-114 [FREE Full text] [doi: 10.1016/j.psychres.2018.02.045] [Medline: 29524908]
22. Epstein DR, Babcock-Parziale JL, Haynes PL, Herb CA. Insomnia treatment acceptability and preferences of male Iraq and Afghanistan combat veterans and their healthcare providers. J Rehabil Res Dev 2012;49(6):867-878 [FREE Full text] [doi: 10.1682/jrrd.2011.06.0114] [Medline: 23299258]
23. Koffel E, Kuhn E, Petsoulis N, Erbes CR, Anders S, Hoffman JE, et al. A randomized controlled pilot study of CBT-I coach: feasibility, acceptability, and potential impact of a mobile phone application for patients in cognitive behavioral therapy for insomnia. Health Informatics J 2018;24(1):3-13 [FREE Full text] [doi: 10.1177/1460458216656472] [Medline: 27354394]
24. Vincent N, Lionberg C. Treatment preference and patient satisfaction in chronic insomnia. Sleep 2001;24(4):411-417 [FREE Full text] [doi: 10.1093/sleep/24.4.411] [Medline: 11403525]
25. Cartreine JA, Ahern DK, Locke SE. A roadmap to computer-based psychotherapy in the United States. Harv Rev Psychiatry 2010;18(2):80-95. [doi: 10.3109/10673221003707702] [Medline: 20235773]
26. Simon L, Steinmetz L, Feige B, Benz F, Spiegelhalder K, Baumeister H. Comparative efficacy of onsite, digital, and other settings for cognitive behavioral therapy for insomnia: a systematic review and network meta-analysis. Sci Rep 2023;13(1):1929 [FREE Full text] [doi: 10.1038/s41598-023-28853-0] [Medline: 36732610]
27. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia-a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016;30:1-10 [FREE Full text] [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
28. Soh HL, Ho RC, Ho CS, Tam WW. Efficacy of digital cognitive behavioural therapy for insomnia: a meta-analysis of randomised controlled trials. Sleep Med 2020;75:315-325 [FREE Full text] [doi: 10.1016/j.sleep.2020.08.020] [Medline: 32950013]
29. Cheng SK, Dizon J. Computerised cognitive behavioural therapy for insomnia: a systematic review and meta-analysis. Psychother Psychosom 2012;81(4):206-216 [FREE Full text] [doi: 10.1159/000335379] [Medline: 22585048]
30. Reilly ED, Robinson SA, Petrakis BA, Kuhn E, Pigeon WR, Wiener RS, et al. Mobile app use for insomnia self-management: pilot findings on sleep outcomes in veterans. Interact J Med Res 2019;8(3):e12408 [FREE Full text] [doi: 10.2196/12408] [Medline: 31342904]
31. Reilly ED, Robinson SA, Petrakis BA, Gardner MM, Wiener RS, Castaneda-Sceppa C, et al. Mobile intervention to improve sleep and functional health of veterans with insomnia: randomized controlled trial. JMIR Form Res 2021;5(12):e29573 [FREE Full text] [doi: 10.2196/29573] [Medline: 34889746]
32. Hermes EDA, Rosenheck RA. Implementing computer-based psychotherapy among veterans in outpatient treatment for substance use disorders. Psychiatr Serv 2016;67(2):176-183 [FREE Full text] [doi: 10.1176/appi.ps.201400532] [Medline: 26620286]

33. Kuhn E, Miller KE, Puran D, Wielgosz J, YorkWilliams SL, Owen JE, et al. A pilot randomized controlled trial of the insomnia coach mobile app to assess its feasibility, acceptability, and potential efficacy. Behav Ther 2022;53(3):440-457 [FREE Full text] [doi: 10.1016/j.beth.2021.11.003] [Medline: 35473648]
34. Hermes EDA, Rosenheck RA, Burrone L, Dante G, Lukens C, Martino S. The implementation and effectiveness of digital cognitive behavioral therapy for insomnia in primary care: a pilot study. Implement Res Pract 2021;2:26334895211053659 [FREE Full text] [doi: 10.1177/26334895211053659] [Medline: 37090002]
35. Cavanagh RC, Mackey R, Bridges L, Gleason A, Ciulla R, Micheel L, et al. The use of digital health technologies to manage insomnia in military populations. J Technol Behav Sci 2019;5(1):61-69 [FREE Full text] [doi: 10.1007/s41347-019-00114-w]
36. Vedaa Ø, Kallestad H, Scott J, Smith ORF, Pallesen S, Morken G, et al. Effects of digital cognitive behavioural therapy for insomnia on insomnia severity: a large-scale randomised controlled trial. Lancet Digit Health 2020;2(8):e397-e406 [FREE Full text] [doi: 10.1016/S2589-7500(20)30135-7] [Medline: 33328044]
37. Hagatun S, Vedaa Ø, Nordgreen T, Smith ORF, Pallesen S, Havik OE, et al. The short-term efficacy of an unguided internet-based cognitive-behavioral therapy for insomnia: a randomized controlled trial with a six-month nonrandomized follow-up. Behav Sleep Med 2019;17(2):137-155 [FREE Full text] [doi: 10.1080/15402002.2017.1301941] [Medline: 28345961]
38. Shi X, Buysse DJ, Ritterband LM, Sereika SM, Strollo PJ, Wenzel SE, et al. Solving insomnia electronically: sleep treatment for asthma (SIESTA): a study protocol for a randomized controlled trial. Contemp Clin Trials 2019;79:73-79 [FREE Full text] [doi: 10.1016/j.cct.2019.02.011] [Medline: 30825525]
39. Zhou ES, Ritterband LM, Bethea TN, Robles YP, Heeren TC, Rosenberg L. Effect of culturally tailored, internet-delivered cognitive behavioral therapy for insomnia in Black women: a randomized clinical trial. JAMA Psychiatry 2022;79(6):538-549 [FREE Full text] [doi: 10.1001/jamapsychiatry.2022.0653] [Medline: 35442432]
40. Ritterband LM, Thorndike FP, Gonder-Frederick LA, Magee JC, Bailey ET, Saylor DK, et al. Efficacy of an internet-based behavioral intervention for adults with insomnia. Arch Gen Psychiatry 2009;66(7):692-698 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.66] [Medline: 19581560]
41. Thorndike FP, Ritterband LM, Gonder-Frederick LA, Lord HR, Ingersoll KS, Morin CM. A randomized controlled trial of an internet intervention for adults with insomnia: effects on comorbid psychological and fatigue symptoms. J Clin Psychol 2013;69(10):1078-1093 [FREE Full text] [doi: 10.1002/jclp.22032] [Medline: 24014057]
42. Ritterband LM, Bailey ET, Thorndike FP, Lord HR, Farrell-Carnahan L, Baum LD. Initial evaluation of an internet intervention to improve the sleep of cancer survivors with insomnia. Psychooncology 2012;21(7):695-705 [FREE Full text] [doi: 10.1002/pon.1969] [Medline: 21538678]
43. Thorndike FP, Saylor DK, Bailey ET, Gonder-Frederick L, Morin CM, Ritterband LM. Development and perceived utility and impact of an internet intervention for insomnia. E J Appl Psychol 2008;4(2):32-42 [FREE Full text] [doi: 10.7790/ejap.v4i2.133] [Medline: 20953264]
44. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. JAMA Psychiatry 2017;74(1):68-75 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.3249] [Medline: 27902836]
45. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.
46. Bastien CH, Vallières A, Morin CM. Validation of the insomnia severity index as an outcome measure for insomnia research. Sleep Med 2001;2(4):297-307 [FREE Full text] [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
47. Reynolds WM. Psychometric characteristics of the adult suicidal ideation questionnaire in college students. J Pers Assess 1991;56(2):289-307 [FREE Full text] [doi: 10.1207/s15327752jpa5602_9] [Medline: 2056423]
48. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988;56(6):893-897 [FREE Full text] [doi: 10.1037//0022-006x.56.6.893] [Medline: 3204199]
49. Armed Forces Health Surveillance Center. Insomnia, active component, US Armed Forces, January 2000-December 2009. MSMR 2010;17(5):12-15. [doi: 10.21236/ada614179]
50. Lew HL, Pogoda TK, Hsu PT, Cohen S, Amick MM, Baker E, et al. Impact of the "polytrauma clinical triad" on sleep disturbance in a Department of Veterans Affairs outpatient rehabilitation setting. Am J Phys Med Rehabil 2010;89(6):437-445 [FREE Full text] [doi: 10.1097/PHM.0b013e3181ddd301] [Medline: 20489391]
51. Alexander M, Ray MA, Hébert JR, Youngstedt SD, Zhang H, Steck SE, et al. The national veteran sleep disorder study: descriptive epidemiology and secular trends, 2000-2010. Sleep 2016;39(7):1399-1410 [FREE Full text] [doi: 10.5665/sleep.5972] [Medline: 27091538]
52. Germain A, Richardson R, Stocker R, Mammen O, Hall M, Bramoweth AD, et al. Treatment for insomnia in combat-exposed OEF/OIF/OND military veterans: preliminary randomized controlled trial. Behav Res Ther 2014;61:78-88 [FREE Full text] [doi: 10.1016/j.brat.2014.07.016] [Medline: 25194223]
53. Vasterling JJ, Verfaellie M, Sullivan KD. Mild traumatic brain injury and posttraumatic stress disorder in returning veterans: perspectives from cognitive neuroscience. Clin Psychol Rev 2009;29(8):674-684 [FREE Full text] [doi: 10.1016/j.cpr.2009.08.004] [Medline: 19744760]

54. Gilbert KS, Kark SM, Gehrman P, Bogdanova Y. Sleep disturbances, TBI and PTSD: implications for treatment and recovery. Clin Psychol Rev 2015;40:195-212 [FREE Full text] [doi: 10.1016/j.cpr.2015.05.008] [Medline: 26164549]
55. Vincent N, Lewycky M. Logging on for better sleep: RCT of the effectiveness of online treatment for insomnia. Sleep 2009;32(6):807-815 [FREE Full text] [doi: 10.1093/sleep/32.6.807] [Medline: 19544758]
56. Okajima I, Komada Y, Inoue Y. A meta-analysis on the treatment effectiveness of cognitive behavioral therapy for primary insomnia. Sleep Biol Rhythms 2011;9(1):24-34. [doi: 10.1111/j.1479-8425.2010.00481.x]
57. Morin CM, Culbert JP, Schwartz SM. Nonpharmacological interventions for insomnia: a meta-analysis of treatment efficacy. Am J Psychiatry 1994;151(8):1172-1180 [FREE Full text] [doi: 10.1176/ajp.151.8.1172] [Medline: 8037252]
58. Murtagh DR, Greenwood KM. Identifying effective psychological treatments for insomnia: a meta-analysis. J Consult Clin Psychol 1995;63(1):79-89. [doi: 10.1037//0022-006x.63.1.79] [Medline: 7896994]
59. Smith MT, Perlis ML, Park A, Smith MS, Pennington J, Giles DE, et al. Comparative meta-analysis of pharmacotherapy and behavior therapy for persistent insomnia. Am J Psychiatry 2002;159(1):5-11 [FREE Full text] [doi: 10.1176/appi.ajp.159.1.5] [Medline: 11772681]
60. Espie CA. "Stepped care": a health technology solution for delivering cognitive behavioral therapy as a first line insomnia treatment. Sleep 2009;32(12):1549-1558 [FREE Full text] [doi: 10.1093/sleep/32.12.1549] [Medline: 20041590]
61. Christensen H, Batterham PJ, Gosling JA, Ritterband LM, Griffiths KM, Thorndike FP, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight study): a randomised controlled trial. Lancet Psychiatry 2016;3(4):333-341 [FREE Full text] [doi: 10.1016/S2215-0366(15)00536-2] [Medline: 26827250]
62. Woosley JA, Lichstein KL, Taylor DJ, Riedel BW, Bush AJ. Insomnia complaint versus sleep diary parameters: predictions of suicidal ideation. Suicide Life Threat Behav 2016;46(1):88-95. [doi: 10.1111/sltb.12173] [Medline: 26052862]
63. Kazis LW. The Veterans SF-36 Health Status Questionnaire: development and application in the Veterans Health Administration. Med Outcomes 2000;5(1):1-18.
64. Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The medical outcomes study. an application of methods for monitoring the results of medical care. JAMA 1989;262(7):925-930 [FREE Full text] [doi: 10.1001/jama.262.7.925] [Medline: 2754793]
65. Brown GK. A Review of Suicide Assessment Measures for Intervention Research with Adults and Older Adults. Pennsylvania: University of Pennsylvania; 2001.
66. Beck AT, Ward CH, Mendelson J, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4(6):561-571. [doi: 10.1001/archpsyc.1961.01710120031004] [Medline: 13688369]
67. Holmes TH, Rahe RH. The social readjustment rating scale. J Psychosom Res 1967;11(2):213-218 [FREE Full text] [doi: 10.1016/0022-3999(67)90010-4] [Medline: 6059863]
68. Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 1976;4(2):97-110. [Medline: 1027738]
69. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. Posttraumatic stress disorder checklist for DSM-5 (PCL-5). U.S. Department of Veterans Affairs. 2013. URL: https://www.ohsu.edu/sites/default/files/2022-08/%28PCL-5%29%20PTSD%20Checklist%20for%20DSM5.pdf [accessed 2015-12-06]
70. Morin CM. Insomnia: Psychological Assessment and Management. New York, NY: Guilford Publications; 1993.
71. Morin CM, Hauri PJ, Espie CA, Spielman AJ, Buysse DJ, Bootzin RR. Nonpharmacologic treatment of chronic insomnia. an American Academy of Sleep Medicine Review. Sleep 1999;22(8):1134-1156 [FREE Full text] [doi: 10.1093/sleep/22.8.1134] [Medline: 10617176]
72. Hasson-Ohayon I, Roe D, Kravetz S. A qualitative approach to the evaluation of psychosocial interventions for persons with severe mental illness. Psychol Serv 2006;3(4):262-273. [doi: 10.1037/1541-1559.3.4.262]
73. Nguyen TD, Attkisson CC, Stegner BL. Assessment of patient satisfaction: development and refinement of a service evaluation questionnaire. Eval Program Plann 1983;6(3-4):299-313. [doi: 10.1016/0149-7189(83)90010-1] [Medline: 10267258]
74. Bagalman E. The number of veterans that use VA health care services: a fact sheet. Congressional Research Service. 2014. URL: https://sgp.fas.org/crs/misc/R43579.pdf [accessed 2015-12-06]
75. Morin CM, Bootzin RR, Buysse DJ, Edinger JD, Espie CA, Lichstein KL. Psychological and behavioral treatment of insomnia:update of the recent evidence (1998-2004). Sleep 2006;29(11):1398-1414 [FREE Full text] [doi: 10.1093/sleep/29.11.1398] [Medline: 17162986]
76. First MB, Williams JB, Karg RS, Spitzer RL. Structured clinical interview for DSM-5-Research version (SCID-5 for DSM-5, research version; SCID-5-RV). Arlington, VA: American Psychiatric Association; 2015.
77. Morin CM, Belleville G, Bélanger L, Ivers H. The insomnia severity index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
78. Thorndike FP, Ritterband LM, Saylor DK, Magee JC, Gonder-Frederick LA, Morin CM. Validation of the insomnia severity index as a web-based measure. Behav Sleep Med 2011;9(4):216-223 [FREE Full text] [doi: 10.1080/15402002.2011.606766] [Medline: 22003975]

79. Kazis LE, Miller DR, Clark JA, Skinner KM, Lee A, Ren XS, et al. Improving the response choices on the veterans SF-36 health survey role functioning scales: results from the veterans health study. J Ambul Care Manage 2004;27(3):263-280. [doi: 10.1097/00004479-200407000-00010] [Medline: 15287216]
80. Kazis LE, Ren XS, Lee A, Skinner K, Rogers W, Clark J, et al. Health status in VA patients: results from the veterans health study. Am J Med Qual 1999;14(1):28-38 [FREE Full text] [doi: 10.1177/106286069901400105] [Medline: 10446661]
81. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
82. Beck AT, Steer RA, Brown GK. BDI-II, Beck Depression Inventory. San Antonio, Tex, Boston: Psychological Corporation; 2011.
83. Weathers FW, Litz B, Herman D, Juska J, Keane T. PTSD checklist-civilian version. American Psychological Association. 2012. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft02622-000 [accessed 2023-11-08]
84. Batterham PJ, Ftanou M, Pirkis J, Brewer JL, Mackinnon AJ, Beautrais A, et al. A systematic review and evaluation of measures for suicidal ideation and behaviors in population-based research. Psychol Assess 2015;27(2):501-512. [doi: 10.1037/pas0000053] [Medline: 25496086]
85. Marx BP, Lee DJ, Norman SB, Bovin MJ, Sloan DM, Weathers FW, et al. Reliable and clinically significant change in the clinician-administered PTSD scale for DSM-5 and PTSD checklist for DSM-5 among male veterans. Psychol Assess 2022;34(2):197-203 [FREE Full text] [doi: 10.1037/pas0001098] [Medline: 34941354]
86. Eichstaedt KE, Kovatch K, Maroof DA. A less conservative method to adjust for familywise error rate in neuropsychological research: the Holm's sequential bonferroni procedure. NeuroRehabilitation 2013;32(3):693-696. [doi: 10.3233/NRE-130893] [Medline: 23648625]
87. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
88. SAS Institute Inc. 2016. URL: https://www.sas.com/en_in/home.html [accessed 2023-11-08]
89. Taylor DJ, Peterson AL, Pruiksma KE, Young-McCaughan S, Nicholson K, Mintz J, et al. Internet and in-person cognitive behavioral therapy for insomnia in military personnel: a randomized clinical trial. Sleep 2017;40(6):1-12 [FREE Full text] [doi: 10.1093/sleep/zsx075] [Medline: 28472528]
90. Taylor DJ, Peterson AL, Pruiksma KE, Hale WJ, Young-McCaughan S, Wilkerson A, et al. Impact of cognitive behavioral therapy for insomnia disorder on sleep and comorbid symptoms in military personnel: a randomized clinical trial. Sleep 2018;41(6):1-11 [FREE Full text] [doi: 10.1093/sleep/zsy069] [Medline: 29618098]
91. Xu D, Cardell E, Broadley SA, Sun J. Efficacy of face-to-face delivered cognitive behavioral therapy in improving health status of patients with insomnia: a meta-analysis. Front Psychiatry 2021;12:798453 [FREE Full text] [doi: 10.3389/fpsyt.2021.798453] [Medline: 35002813]
92. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3):220-233 [FREE Full text] [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
93. Denis D, Eley TC, Rijsdijk F, Zavos HMS, Keers R, Espie CA, et al. Is digital cognitive behavioural therapy for insomnia effective in treating sub-threshold insomnia: a pilot RCT. Sleep Med 2020;66:174-183 [FREE Full text] [doi: 10.1016/j.sleep.2019.10.007] [Medline: 31901759]
94. e Cruz MM, Kryger MH, Morin CM, Palombini L, Salles C, Gozal D. Comorbid insomnia and sleep apnea: mechanisms and implications of an underrecognized and misinterpreted sleep disorder. Sleep Med 2021;84:283-288 [FREE Full text] [doi: 10.1016/j.sleep.2021.05.043] [Medline: 34214960]
95. Luik AI, Kyle SD, Espie CA. Digital Cognitive Behavioral Therapy (dCBT) for insomnia: a state-of-the-science review. Curr Sleep Med Rep 2017;3(2):48-56 [FREE Full text] [doi: 10.1007/s40675-017-0065-4] [Medline: 28553574]
96. Luik AI, van der Zweerde T, van Straten A, Lancee J. Digital delivery of cognitive behavioral therapy for insomnia. Curr Psychiatry Rep 2019;21(7):50 [FREE Full text] [doi: 10.1007/s11920-019-1041-0] [Medline: 31161406]
97. Hermes EDA, Lyon AR, Schueller SM, Glass JE. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J Med Internet Res 2019;21(1):e11752 [FREE Full text] [doi: 10.2196/11752] [Medline: 30681966]

Abbreviations

CBT-I: Cognitive behavioral therapy for insomnia
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EBP: evidence-based psychotherapy
EMA: early morning awakenings
F2F: face-to-face
iCBT: internet-delivered cognitive behavioral therapy
iCBT-I: internet-delivered cognitive behavioral therapy for insomnia
IEW: Insomnia Education Website

ISI: Insomnia Severity Index
MCS: Mental Component Summary
OEF: Operation Enduring Freedom
OIF: Operation Iraqi Freedom
OND: Operation New Dawn
PCL-5: Posttraumatic Stress Disorder Checklist for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
PCS: Physical Component Summary
RCT: randomized controlled trial
SHUTi: Sleep Healthy Using the Internet
SOL: sleep onset latency
TIB: time in bed
TST: total sleep time
VA: Department of Veterans Affairs
Veterans SF-36: Short Form Health Survey for Veterans
VHA: Veterans Health Administration
VR-36: Veterans RAND 36-item Health Survey
WASO: wake after sleep onset

Edited by J Torous; submitted 03.07.23; peer-reviewed by H Sun, L Yang; comments to author 05.08.23; revised version received 26.09.23; accepted 06.10.23; published 24.11.23.

Please cite as:
Nazem S, Barnes SM, Forster JE, Hostetter TA, Monteith LL, Kramer EB, Gaeddert LA, Brenner LA
Efficacy of an Internet-Delivered Intervention for Improving Insomnia Severity and Functioning in Veterans: Randomized Controlled Trial
JMIR Ment Health 2023;10:e50516
URL: https://mental.jmir.org/2023/1/e50516
doi:10.2196/50516
PMID:37999953

Original Paper

A Digital Single-Session Intervention (Project Engage) to Address Fear of Negative Evaluation Among College Students: Pilot Randomized Controlled Trial

Arka Ghosh[1,2], PhD; Katherine A Cohen[1], MA; Laura Jans[1], MA; Carly A Busch[3], BA; Riley McDanal[4], MA; Yuanyuan Yang[5], MA; Katelyn M Cooper[3], PhD; Jessica L Schleider[1,2], PhD

[1]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Center for Behavioral Intervention Technologies, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Research for Inclusive STEM Education Center, School of Life Sciences, Arizona State University, Tempe, AZ, United States

[4]Department of Psychology, Stony Brook University, Stony Brook, NY, United States

[5]Department of Psychology, University of Kansas, Lawrence, KS, United States

Corresponding Author:
Arka Ghosh, PhD
Department of Medical Social Sciences
Feinberg School of Medicine
Northwestern University
625 N Michigan Ave, 21st Fl
Chicago, IL, 60611
United States
Phone: 1 3125033403
Email: arka.ghosh@northwestern.edu

Abstract

Background: Increasingly, college science courses are transitioning from a traditional lecture format to active learning because students learn more and fail less frequently when they engage in their learning through activities and discussions in class. Fear of negative evaluation (FNE), defined as a student's sense of dread associated with being unfavorably evaluated while participating in a social situation, discourages undergraduates from participating in small group discussions, whole class discussions, and conversing one-on-one with instructors.

Objective: This study aims to evaluate the acceptability of a novel digital single-session intervention and to assess the feasibility of implementing it in a large enrollment college science course taught in an active learning way.

Methods: To equip undergraduates with skills to cope with FNE and bolster their confidence, clinical psychologists and biology education researchers developed Project Engage, a digital, self-guided single-session intervention for college students. It teaches students strategies for coping with FNE to bolster their confidence. Project Engage provides biologically informed psychoeducation, uses interactive elements for engagement, and helps generate a personalized action plan. We conducted a 2-armed randomized controlled trial to evaluate the acceptability and the preliminary effectiveness of Project Engage compared with an active control condition that provides information on available resources on the college campus.

Results: In a study of 282 upper-level physiology students, participants randomized to complete Project Engage reported a greater increase in overall confidence in engaging in small group discussions (P=.01) and whole class discussions (P<.001), but not in one-on-one interactions with instructors (P=.05), from baseline to immediately after intervention outcomes, compared with participants in an active control condition. Project Engage received a good acceptability rating (1.22 on a scale of –2 to +2) and had a high completion rate (>97%).

Conclusions: This study provides a foundation for a freely available, easily accessible intervention to bolster student confidence for contributing in class.

Trial Registration: OSF Registries osf.io/4ca68 http://osf.io/4ca68

(JMIR Ment Health 2023;10:e48926) doi:10.2196/48926

KEYWORDS

single-session intervention; fear of negative evaluation; active learning course; pilot randomized controlled trial; intervention; college student; science course; active learning; negative evaluation

Introduction

Overview

For over a decade, national calls have championed the transition of college science courses from the traditional lecture format, where instructors lecture at students who passively listen, to active learning courses, where students engage in their learning during class [1,2]. In active learning courses, students engage in learning by participating in activities, such as clicker questions or worksheets, as well as having discussions with their peers and instructors during class [3]. The national push to adopt active learning resulted from robust evidence suggesting that, on average, students learn more and fail less frequently in active learning science courses than in traditional lecture science courses [4] and demonstrating that active learning narrows the achievement gap in science for students from underrepresented groups [5]. Despite the undeniable benefits of active learning, recent research has found that active learning can cause significant anxiety in undergraduates [6-10]. These feelings of worry and apprehension are primarily due to students' fears of being negatively evaluated by their classmates and instructors, often instigated by low levels of confidence [6-8]. As a first step toward ameliorating this common student-level challenge, we conducted a randomized pilot evaluation of a novel single-session digital intervention, Project Engage, designed to help students better manage their fear of negative evaluation (FNE) and bolster their confidence within active learning environments.

Background

FNE is defined as a sense of dread associated with being unfavorably evaluated in a social situation [11,12]. It was first applied in the context of higher education in language-learning courses, where students are regularly expected to engage in discussion during class [13]. However, with the transition of science courses to active learning, there are far more social evaluative situations, or opportunities for students to be negatively judged by their peers, owing to the increased number of conversations during class. Specifically, 2 qualitative interview studies, one focused on students enrolled in large-enrollment active learning college science courses and the other on students enrolled in small-enrollment active learning college science courses, found that students describe that if they contribute their thoughts to a discussion about science and their thoughts are wrong, others will perceive them as *dumb* or *stupid* [6,8]. Students were also worried that a single interaction with their peers or instructor could lead to a lasting negative reputation.

As a result of these feelings of worry and apprehension, undergraduates described that they struggle to think through science problems, have difficulty articulating their thoughts about science in discussions, and avoid participating in conversations [6,8]. A recent study of >500 undergraduates enrolled in large-enrollment active learning science courses echoed these findings; students most commonly reported that FNE caused them to overthink their responses, participate less, struggle to speak, and struggle to think [14]. Undergraduates who are worried about being negatively evaluated may monitor their environment for the threat of potential judgment [15]. This likely increases cognitive load, consequently limiting their ability to think and perform specific tasks [16]. As such, it is unsurprising that this fear of judgment primarily hinders students' performance in class. In addition, FNE has been linked to low self-confidence [8]. If students' concerns about being judged in science courses cause them to feel less confident, this may lower their self-efficacy, ultimately negatively affecting their performance [17].

Notably, the solution to these challenges is not to eliminate active learning from college science courses. Overwhelming evidence suggests that reverting to the traditional lecture format would not only be detrimental to student learning but would also be less equitable [4,5]. Further, traditional lectures can also exacerbate feelings of anxiety in students, but for different reasons than active learning courses do [8]. In traditional lectures, students express worry that they are unable to gauge how much they have learned before summative assessments and that there is a lack of opportunities to clarify their understanding with others. Therefore, researchers have suggested a 2-pronged solution to address student FNE in active learning [18]. First, to adjust how active learning practices are implemented, and second, to bolster students' confidence and ability to cope with the FNE.

In the context of higher education, there have been few documented efforts to reduce social anxiety or bolster confidence in speaking in the classroom. One study of 20 students at the National University of Singapore found that allowing chemistry students to contribute to discussions with partial anonymity (voice only) reduced feelings of anxiety [19]. Another study of >500 business students at a large university found that in classes where instructors frequently cold-called students, defined as calling on students whose hands were not raised, students became more comfortable participating in class, compared with students in classes that did not implement cold-call as frequently [20]. However, no studies have directly targeted FNE in the context of active learning. Reducing FNE is typically considered a goal in the treatment of social anxiety disorder, suggesting that researchers interested in FNE look toward literature in clinical psychology. Intervention strategies from cognitive behavioral therapy (CBT) for social anxiety include providing psychoeducation, encouraging exposure to feared stimuli, discouraging avoidance of feared stimuli, and improving cognitive flexibility [21]. Although there is substantial evidence from meta-analytic studies that CBT-informed treatments are effective, most interventions are intensive, typically requiring between 12 and 16 sessions [22,23]. Such treatment cannot be feasibly implemented for each student engaging in active learning courses. However, there is evidence that single-session interventions (SSIs), defined

as "specific, structured programs that intentionally involve just one visit or encounter with a clinic, provider, or program" [24], can result in sustained change.

SSIs are intentionally designed to target clinically relevant mechanisms in a brief, self-contained period. They have been used for a broad array of problems, including anxiety, depression, conduct problems, alcohol abuse, and more [25]. Evidence suggests that these effects are longstanding. For example, individuals randomized to complete a 30-minute, digital, self-administered SSI designed to encourage behavioral activation showed significantly greater decreases in depression at a 3-month follow-up compared with individuals randomized to an active control condition [26]. Across multiple trials, both youths and young adults, who report limited access to traditional mental health services owing to a variety of structural- and stigma-related barriers [27], rated SSIs as highly acceptable [26,28,29].

The reduced burden of SSIs, particularly those that are delivered digitally and are self-administered, allows for their dissemination in a wide variety of settings, including classrooms. Several mental health–focused trials have found positive psychopathological outcomes of SSIs in school settings [30-33], including higher education settings [34,35]. However, SSIs have not yet been applied to clinically relevant difficulties, such as classroom anxiety or FNE, out of the context of mental health–focused trials. No SSIs designed to reduce FNE have been tested among college students. Other trials have found positive outcomes for brief (≤3 sessions) interventions on academic achievement [36-38]. Given the demonstrated utility of brief, scalable interventions in educational environments, there is reason to expect that SSIs may have utility beyond clinical contexts.

Study Purpose

To address this gap in the literature, we developed and assessed a digital, self-administered SSI designed to reduce FNE and increase confidence in different situations in a classroom setting. In this pilot study, our primary aim was to assess whether the intervention was acceptable and whether it could be feasibly implemented in a large-enrollment college science course taught in an active learning way. The SSI aimed to teach strategies for dealing with FNE in the context of anxiety. We hypothesized that the intervention would impact constructs related to student anxiety such as distress tolerance [39] and intolerance of uncertainty [40] relative to an active control program. We also hypothesized that relative to an active control program, the intervention would increase the intention to persist in science [9,41], which may be threatened by FNE. To our knowledge, this is the first study to develop and test the impact of a digital SSI on college science students with the intent to improve their experiences in the classroom.

Methods

Ethical Considerations

Before participant enrollment, all study procedures were approved by Arizona State University's institutional review board (#STUDY00015263) and preregistered on the Open Science Framework [42]. The trial results have been reported using the CONSORT (Consolidated Standards of Reporting Trials) guidelines [43].

Recruitment and Procedures

Participants were recruited from a large-enrollment upper-level physiology course taught in an active learning format at Arizona State University, a large public research university located in the southwestern region of the United States. Participant recruitment began on November 11, 2022. The intervention was created as a Qualtrics survey. Qualtrics is a software primarily used for designing and disseminating web-based surveys. The intervention was shared as a Qualtrics link as an assignment with 320 undergraduate students. They were given until November 15, 2022, to complete the survey in exchange for course points (<1%). Alternate assignments were available for students who chose not to participate. Eligibility criteria were (1) comfort speaking and writing in English, (2) consistent access to an internet-equipped device, and (3) aged ≥18 years. Exclusion criteria were (1) responding *yes* to the question "Are you taller than 7 feet?," which was included as an attention check or (2) responding that they could not commit to completing the entire 45-minute activity.

All participant data were collected and stored using Qualtrics, a secure survey platform. After completing the web-based screener, eligible participants were brought to a page allowing them to document their informed consent, and then directed to a series of preintervention questionnaires. Participants were then randomized (1:1 allocation ratio) to either the intervention condition (*Project Engage*) or an active control condition via a digitally embedded randomizer available in Qualtrics. All participants were blinded to condition assignment, as they were unaware of whether they were in the control group, which was designed to be helpful and contained a positive message, but which did not target the intended target (FNE) of the novel SSI. Participants completed a second set of questionnaires directly after intervention completion to gauge immediate shifts in the outcomes of interest as well as user acceptability.

Condition Descriptions

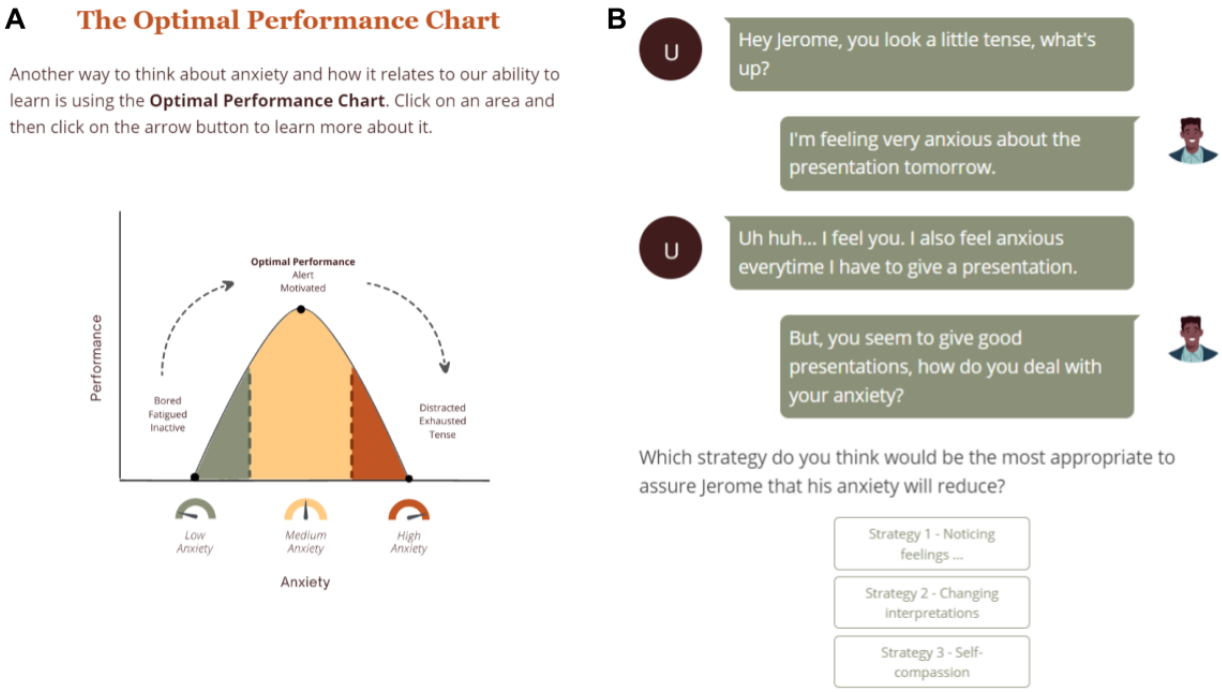
Project Engage [44] is a web-based and self-guided SSI designed to target FNE within the context of an active learning Science, Technology, Engineering, and Medicine (STEM) course. The intervention introduced anxiety in general and discussed coping strategies by providing examples related to FNE in the context of an active learning class. The coping strategies incorporate principles of CBT and mindfulness, encouraging students to reflect on and recognize patterns between their thoughts, feelings, and behaviors in a nonjudgmental manner. Apart from reducing FNE, we expected the intervention to improve student confidence in different social situations in an active learning environment. Project Engage contains 7 main components:

1. Student anxiety within active learning courses, a common outcome of the FNE, is normalized and validated.
2. The potentially adaptive nature of anxiety is described. An optimal performance chart

(which plots anxiety vs performance; Figure 1A) is used to help students differentiate between functional and nonfunctional levels of anxiety.

3. Students are presented with a list of socially evaluative situations that may arise in an active learning classroom and asked which would be most stressful for them (eg, "being involuntarily asked to speak in front of the class"). Then, students identify which feelings, thoughts, and behaviors they may experience in the selected situation.
4. Students are presented with three strategies for managing anxiety in the classroom: (1) noticing physiological responses as anxiety, (2) changing interpretation of anxiogenic situations (ie, cognitive flexibility), and (3) self-compassion.
5. Students read quotes from peers who have successfully used each of the strategies.
6. Students put the strategies into practice by helping a hypothetical peer through a socially evaluative situation via an interactive texting activity (Figure 1B).
7. Students develop an action plan to remind them of how socially stressful situations might make them feel, think, and behave as well as how to handle them.

Figure 1. (A) Snapshot of an interactive page in Project Engage. The graph shows how performance varies with anxiety. The participant can click on the colored zones to learn more about how being in that zone of anxiety would affect their performance. (B) Snapshot of an interactive conversation between the participant (left) and a simulated friend (right). Through this conversation, the participant gets a chance to think about which of the techniques they learned in Project Engage would be appropriate in a given situation.



The main theoretical framework driving the design of Project Engage was the self-determination theory [45]. The self-determination theory posits that it is possible to affect positive behavior change if an intervention supports an individual's need for competence, relatedness, and autonomy. In line with the principles of self-determination theory and other effective SSIs [24], Project Engage empowers participants to assume an *expert role* (eg, students are experts of their own undergraduate experience) to instill a sense of autonomy. Participants were presented with a set of 3 strategies and a personalized *action plan* to elicit motivation and a sense of competence [26,46,47]. Participants were also presented with testimonials from peers to give them a sense of relatedness to others. To the best of our knowledge, Project Engage is the first intervention based on self-determination theory aiming to increase student confidence in the context of higher education.

The control condition, which contained a generally positive message linked to well-being and mental health, was similarly web-based and self-guided to account for any nonspecific outcomes of completing an web-based activity. Participants also began this activity by identifying a socially evaluative situation that would elicit anxiety and then choosing the thoughts, feelings, and actions that they would associate with that situation. Instead of being provided with evidence-based strategies, the participants in the control condition were presented with a list of resources specific to their university. Similar to Project Engage, the control condition incorporated open-ended prompts to encourage a similar level of participant effort and engagement.

Measures

Demographics

The following demographic information was collected: gender identity, racial and ethnic identity, sexual orientation, LGBTQ+ (lesbian, gay, bisexual, transgender, queer, and others) identity, parental education level, financial stability, household income, primary language spoken, grade point average, international student status, and disability status.

Assessing FNE

The straightforward Brief Fear of Negative Evaluation (BFNE-S) subscale [12] was used to assess FNE. Only positively worded items (ie, assessing the presence, rather than absence, of fear) were presented. Students were asked to rate 8 statements (eg, "I am usually worried about what kind of impression I make") on a 5-point Likert scale (1="Not at all characteristic of me" and 5="Extremely characteristic of me") specifically within the context of a STEM active learning course. The total scores range from 8 to 40, with higher scores indicating greater fear of negative social evaluation. As this measure asks students to reflect on their past experiences in active learning courses, BFNE-S scores were not expected to change before intervention to immediately after the intervention. Therefore, the BFNE-S was only presented before the intervention. The BFNE-S has been found to have sound psychometric properties, including internal consistency, convergent validity, and divergent validity, in adults with social anxiety disorder [12].

Program Acceptability

The acceptability and feasibility of the program were assessed using the Program Feedback Scale (PFS) [48]. The PFS contains 7 items (eg, "I enjoyed the activity" and "I agree with the activity's message"), which students rated on a 5-point Likert scale with response anchors at –2 ("Really disagree") and 2 ("Really agree"). As per the study preregistration, an average score of ≥0.5 indicates adequate acceptability. This scale was specifically developed for use in brief digital interventions [48].

Proximal Outcomes

Overview

Changes in distress tolerance, intolerance of uncertainty, confidence, and intention to persist in science were evaluated as preintervention to immediate postintervention outcomes. Importantly, we are assessing immediate shifts in the SSI targets because immediate shifts may predict more positive long-term changes in outcomes related to well-being and mental health [49]. Therefore, the following measures were used to better understand the potential effectiveness and utility of the intervention.

Distress Tolerance

The Distress Tolerance Scale (DTS) [50] was used to assess distress tolerance. Students rated 16 statements (eg, "I can't handle feeling distressed or upset") on a 5-point Likert scale (1="Strongly disagree at this moment" and 5="Strongly agree at this moment"; we reverse-coded the scores during analysis to be congruent with the original scoring scheme) specifically within the context of a STEM active learning course. This measure was modified to assess levels of *in-the-moment* distress tolerance by asking students to consider how much each item described them *at this moment*. Total scores range from 16 to 80, with higher scores indicating higher levels of distress tolerance. This measure was presented both before and immediately after the intervention. The DTS has shown acceptable internal consistency and construct validity in previous student samples [50].

Intolerance of Uncertainty

The Intolerance of Uncertainty Scale-12 (IUS-12) [51] was used to assess intolerance of uncertainty. Students rated 12 statements (eg, "I can't stand being taken by surprise") on a 5 point Likert scale (1="Not at all like me at this moment" and 5="Entirely like me at this moment") specifically within the context of a STEM active learning course. This measure was modified to assess levels of *in-the-moment* intolerance of uncertainty by asking students to consider how much each item described them *at this moment*. Total scores range from 12 to 60, with higher scores indicating higher levels of intolerance of uncertainty. This measure was presented both before and immediately after the intervention. The IUS-12 has shown high levels of internal consistency in undergraduate samples [51].

Confidence

We aimed to assess students' confidence in contributing to discussions in different contexts: one-on-one conversations with the instructor, in small groups, and with the whole class. Despite a thorough literature search, we were unable to find suitable, previously developed, and validated measures to assess students' confidence in contributing to these unique contexts. As a result, confidence was assessed by asking students how confident they would be to complete a socially evaluative activity (eg, "Answer a question") within 3 different environments (small group discussion, whole class discussion, and one-on-one discussion with the instructor). For each environment, students completed 4 items rating their *in-the-moment* comfort levels for completing an activity in a large-enrollment college science course (1="Not at all confident at this moment" and 5="Very confident at this moment"; Multimedia Appendix 1 provides full questionnaire). Scores for each environment ranged from 4 to 20, with higher scores indicating higher confidence levels within that environment. These scales were developed in collaboration with a psychometrician who was also on the advisory board for this study. To establish cognitive validity, we conducted 6 think-aloud interviews with undergraduate science students, iteratively revising after each think-aloud to ensure that the questions were being interpreted as intended [52]. These scales were piloted with a sample of undergraduate science students (N=566). Factor analyses indicated excellent model fit (small group discussion Comparative Fit Index (CFI)=1.00, whole class discussion CFI=0.99, and one-on-one discussion with instructor CFI=0.98), and each of the 3 scales was negatively correlated with FNE, as would be expected. The McDonald's omega values indicated adequate internal consistency (all>0.9). Confidence scales for each environment were presented both before and immediately after the intervention.

Intention to Persist in Science

Students were asked to answer the question "To what extent do you intend to pursue a science-related research career?" at this moment on an 11-point Likert scale (0="Definitely will not" and 10="Definitely will"). This item was presented both before and immediately after the intervention. This item is hypothesized to be a proximal predictor of long-term retention in the sciences for underrepresented science students [53].

Analytic Plan

Overview

All demographic variables were reported as percentages of students who endorsed each response option. Completion rates were calculated for each condition as the percentage of students within that condition who completed the entire survey. Mean baseline levels of all primary outcome variables, as well as BFNE-S scores, were reported. Independent 2-tailed *t* tests were performed to check for differences in the mean baseline levels of each affective variable. For all analyses, the effects were considered significant if $P<.05$.

Primary Outcomes

Affective Outcomes

Multiple linear regression models were used to evaluate the effect of condition assignment, separately, on each primary outcome: distress tolerance, intolerance of uncertainty, confidence, and intention to persist in science. In each model, the baseline level of the affective outcome and the intervention condition (experimental=1, control=0) were included as a predictor. As per the preregistered cutoff, the effect of the conditions was considered significant if *P* was <.05.

Program Acceptability

For both conditions, mean scores were calculated. For experimental group participants, all scores have been reported for each item on the PFS. For the control group participants, we have reported the mean acceptability score across the 7 items.

Secondary Outcomes

Student Performance

Student performance was indexed by students' scores on the first 3 examinations. As the first 2 examinations took place before randomization, the first 2 examination averages and condition assignments were added as predictors to the regression model, which predicted the third examination score. A *P* value of <.05 for condition assignment was preregistered as a significant differential effect in this model.

Qualitative Responses

Qualitative feedback was collected about students' experiences with the digital intervention, as well as the strategies they currently use to handle anxiety in interactive classrooms. Responses were thematically analyzed using practices delineated by Braun and Clarke [54]. An inductive approach was used to identify and report commonly recurring themes in the data.

Missing Data and Correcting for Multiple Tests

List-wise deletion was used to exclude noncomplete responses (ie, participants who dropped out of either condition). No missing data were imputed. During the thematic analysis of qualitative data, participants were excluded if their responses made it apparent that they were non-English speakers. In cases where participants submitted multiple responses, their more complete response was used for the analyses. If both responses were equally complete, their initial response was used. Preregistered multiple regressions were corrected using the Benjamini-Hochberg method [55].

Software Packages

We used R (version 4.2.2; R Foundation for Statistical Computing [56]) for running statistical analyses and the MOTE package to calculate effect sizes [57].

Sample Size

As this was a pilot study, we did not perform a sample size calculation. However, we anticipated that 320 participants would be a large enough sample to assess the feasibility and acceptability of the intervention and to provide preliminary evidence of efficacy.

Results

Participant Characteristics

Of the 320 students who were invited to participate in the study, 299 (93.4%) students started the survey. Of these 299 participants, 282 (94.3%) met our inclusion criteria and were randomly assigned to the experimental (n=141, 50%) or control group (n=141, 50%; Figure 2). A total of 138 participants in the experimental group and 141 participants in the control group completed the postintervention questionnaires. The participants in both groups were found to be balanced in terms of demographic factors (Table 1) and baseline scores on the BFNE-S ($P=.59$), IUS-12 ($P=.21$), persistence scale ($P=.27$), and confidence scale ($P=.77$). Only for DTS, the baseline score of experimental group participants was significantly higher than that of the control group participants ($P=.02$).

Figure 2. The flow of participants in the pilot trial.

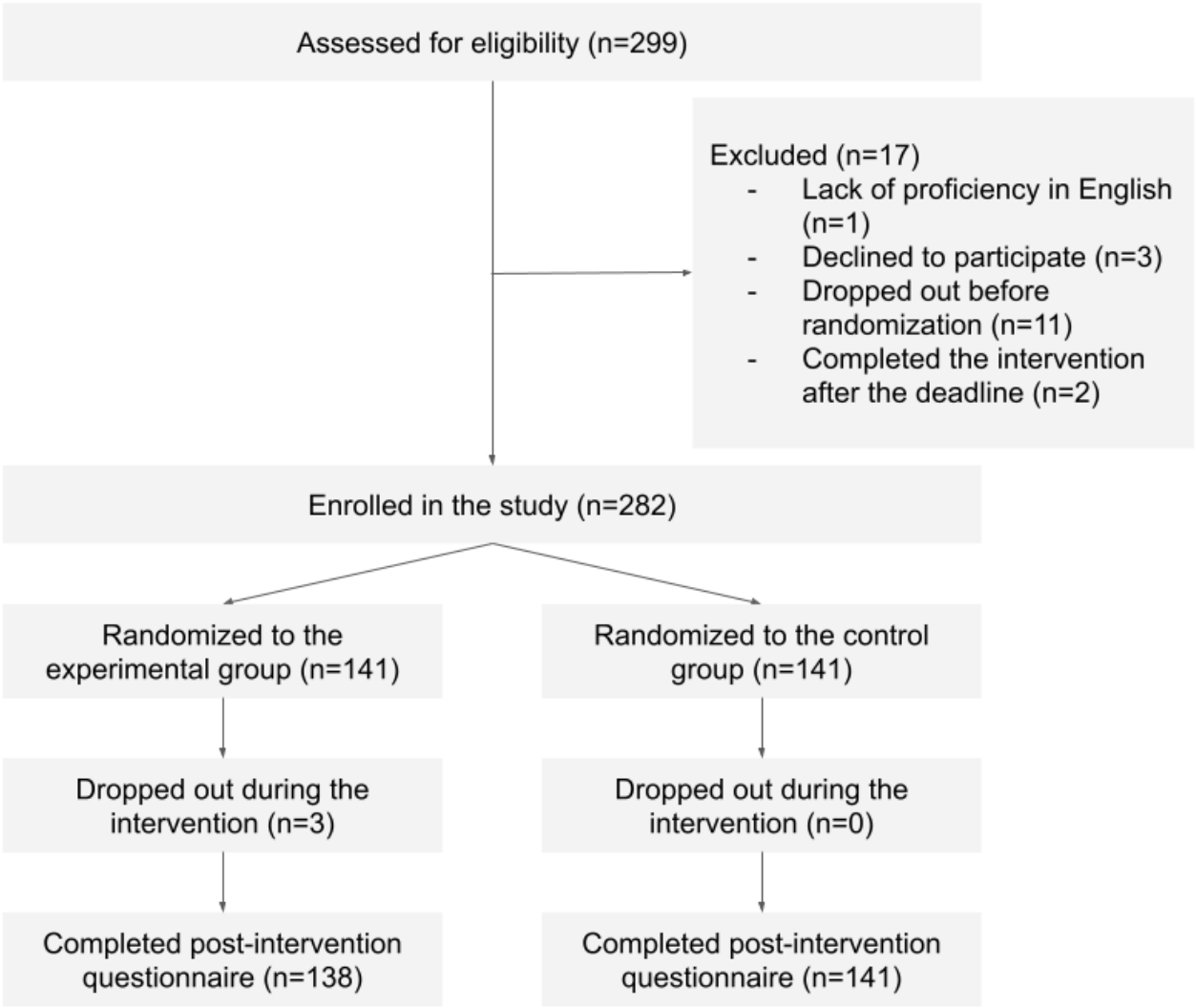


Table 1. Demographic characteristics of participants recruited in the study.

Group	Experimental (n=141)	Control (n=141)	All (n=282)
Age (y), mean (SD)	20.78 (1.89)	21.52 (3.23)	21.15 (2.67)
Gender identity, n (%)			
Men	50 (35.5)	49 (34.7)	99 (35.1)
Women	89 (63.1)	89 (63.1)	178 (63.1)
Nonbinary	2 (1.4)	2 (1.4)	4 (1.4)
Another gender	—[a]	1 (0.7)	1 (0.3)
Race or ethnicity[b], n (%)			
American Indian or Alaska Native	4 (2.8)	3 (2.1)	7 (2.5)
Asian (including South Asian)	43 (30.5)	34 (24.1)	77 (27.3)
Black or African American	10 (7.1)	8 (5.7)	18 (6.4)
Hispanic or Latinx	25 (17.7)	31 (22)	26 (19.9)
Native Hawaiian or another Pacific Islander	2 (1.4)	2 (1.4)	4 (1.4)
White (non-Hispanic; includes Middle Eastern)	69 (48.9)	79 (56)	148 (52.5)
Prefer not to answer	2 (1.4)	2 (1.4)	4 (1.4)
Other (specify)	1 (0.7)	2 (1.4)	3 (1.1)
LGBTQ+[c] status, n (%)			
Yes	27 (19.1)	26 (18.4)	53 (18.8)
No	112 (79.4)	115 (81.6)	227 (80.5)
Did not answer	2 (1.4)	—	2 (0.7)
Gender identity among participants who identify as LGBTQ+[b], n (%)			
Transgender	1 (0.7)	1 (0.7)	2 (0.7)
Gender nonbinary	5 (3.5)	0 (0)	5 (1.8)
Gender queer	0 (0)	3 (2.1)	3 (1.1)
Third gender	0 (0)	0 (0)	0 (0)
Two-spirited	0 (0)	0 (0)	0 (0)
Agender	1 (0.7)	1 (0.7)	2 (0.7)
A gender not listed	17 (12.1)	14 (9.9)	31 (10.9)
Sexual orientation of participants who identify as LGBTQ+[b], n (%)			
Asexual	0 (0)	2 (1.4)	2 (0.7)
Lesbian or gay	12 (8.5)	6 (4.2)	18 (6.4)
Bisexual	13 (9.2)	18 (12.8)	31 (10.99)
Queer	7 (4.9)	2 (1.4)	9 (3.2)
Questioning	3 (2.1)	2 (1.4)	5 (1.8)
An identity not listed	3 (2.1)	0 (0)	3 (1.1)
Parental educated, n (%)			
Did not complete high school	7 (4.9)	10 (7.1)	17 (6)
High-school diploma or GED[d]	17 (12.1)	14 (9.9)	31 (10.9)
Some college but no degree	17 (12.1)	12 (8.5)	29 (10.28)
Associate degree (eg, AA[e] and AS[f])	5 (3.5)	9 (6.4)	14 (4.9)
Bachelor's degree (eg, BA[g] and BS[h])	38 (26.9)	38 (26.9)	76 (26.9)

Group	Experimental (n=141)	Control (n=141)	All (n=282)
Master's degree (eg, MA[i], MS[j], MEd[k], MSW[l], MBA[m])	32 (22.7)	32 (22.7)	64 (22.7)
Higher than a master's degree (eg, PhD[n], MD[o], JD[p])	24 (17)	24 (17)	48 (17)
Prefer not to answer	1 (0.7)	1 (0.7)	2 (0.7)
Other	—	1 (0.7)	1 (0.3)
Financial stability, n (%)			
Yes	82 (58.2)	83 (58.9)	165 (58.5)
Yes, but only sometimes	40 (28.4)	37 (26.2)	77 (27.3)
No	13 (11.3)	19 (13.5)	35 (12.4)
Prefer to answer	3 (2.1)	2 (1.4)	5 (1.8)
Household income (US $), n (%)			
Low income (<25,000)	10 (7.1)	6 (4.3)	16 (5.7)
Middle-low income (25,000-49,999)	16 (11.3)	29 (20.6)	45 (15.9)
Middle income (50,000-99,999)	32 (22.7)	40 (28.4)	72 (25.5)
Middle-high income (100,000-199,999)	50 (35.5)	37 (26.2)	87 (30.9)
High income (≥200,000)	25 (17.7)	18 (12.8)	43 (15.2)
Prefer not to answer	8 (5.7)	11 (7.8)	19 (6.7)
Language, n (%)			
English	135 (95.7)	123 (87.2)	258 (91.5)
Spanish	4 (2.8)	7 (4.9)	11 (3.9)
Other	2 (1.4)	11 (7.8)	13 (4.6)
GPA[q], mean (SD)	3.62 (0.4)	3.62 (0.38)	3.62 (0.39)
International student status, n (%)			
Yes	4 (2.8)	4 (2.8)	8 (2.8)
No	137 (97.2)	137 (97.2)	274 (97.2)
Disability, n (%)			
Yes	19 (13.5)	15 (10.6)	34 (12.1)
No	122 (86.5)	126 (89.4)	248 (87.9)

[a]Not available.
[b]Participants could select more than 1 option.
[c]LGBTQ+: lesbian, gay, bisexual, transgender, queer, and others.
[d]GED: General Educational Development.
[e]AA: Associate of Arts.
[f]AS: Associate of Science.
[g]BA: Bachelor of Arts.
[h]BS: Bachelor of Science.
[i]MA: Master of Arts.
[j]MS: Master of Science.
[k]MEd: Master of Education.
[l]MSW: Master of Social Work.
[m]MBA: Master of Business Administration.
[n]PhD: Doctorate of Philosophy.
[o]MD: Doctor of Medicine.
[p]JD: Juris Doctor.
[q]GPA: Grade Point Average.

Acceptability, Feasibility, and Length

A total of 137 participants in the experimental group submitted the PFS. Overall, Project Engage scored above our acceptability criteria of 0.50 on the PFS (M_{exp}=1.22). The mean acceptability score of Project Engage was higher than that of the control version on the same scale (M_{con}=1.07). All individual items on the PFS received a feedback score higher than the acceptability cutoff (Figure 3). Among the different aspects of the intervention, the question on whether the activity was easy to use received the highest rating, and the question on whether the activity was enjoyable received the lowest rating. According to the text-based feedback provided by the participants on the PFS, the positive aspects of Project Engage were that (1) the techniques will be useful for dealing with anxiety, (2) it was engaging, and (3) it was easy to use. Most participants reported that they would not change anything about the intervention. Apart from that, the most common negative feedback was that Project Engage was too long. Only one participant reported that the intervention affected them negatively; they mentioned, "I am currently going through a very hard time and I felt like this activity just brought more of those emotions to the surface."

The completion rate of the overall study, including the baseline questionnaires, the intervention, and the postintervention questionnaires, was 95.22%. After randomization, only 3 participants dropped out from the experimental group, and 0 participants dropped out from the control group. For the participants randomized to the experimental and control groups, the completion rates were 97.87% and 100%, respectively.

On average, participants in the experimental and control groups spent 39.98 minutes and 26.68 minutes in the overall survey, respectively (Figure 4). Figure 4 shows that a few users have spent an unusually high amount of time with the intervention. These outliers were most likely participants who did not close the survey tab on their browser once they completed the survey. Therefore, to obtain a better estimate of the time spent by the participants, we calculated the total time spent by the participants on the individual pages of the survey and took the sum of those times to calculate the total time spent. We had these data for most but not all pages, so it is a slight underestimation of the actual amount of time spent on the survey. Using this method, we found that the average time spent by the experimental and control group participants was 27.08 minutes and 19.53 minutes, respectively.

Figure 3. Violin plot showing experimental group participants' responses on the Program Feedback Scale. Each violin corresponds to a question in the Program Feedback Scale. Black dots indicate user responses. Solid horizontal red lines indicate the mean values and the dotted red line indicates the acceptability cut-off score of 0.5.

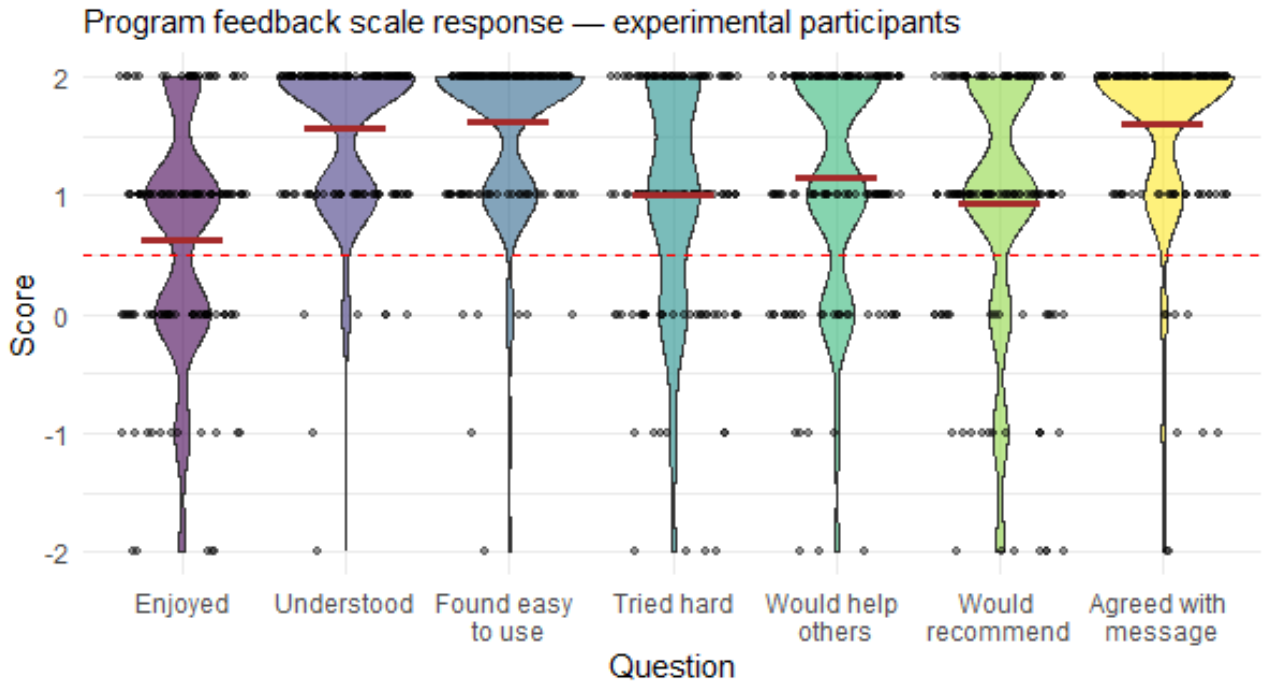
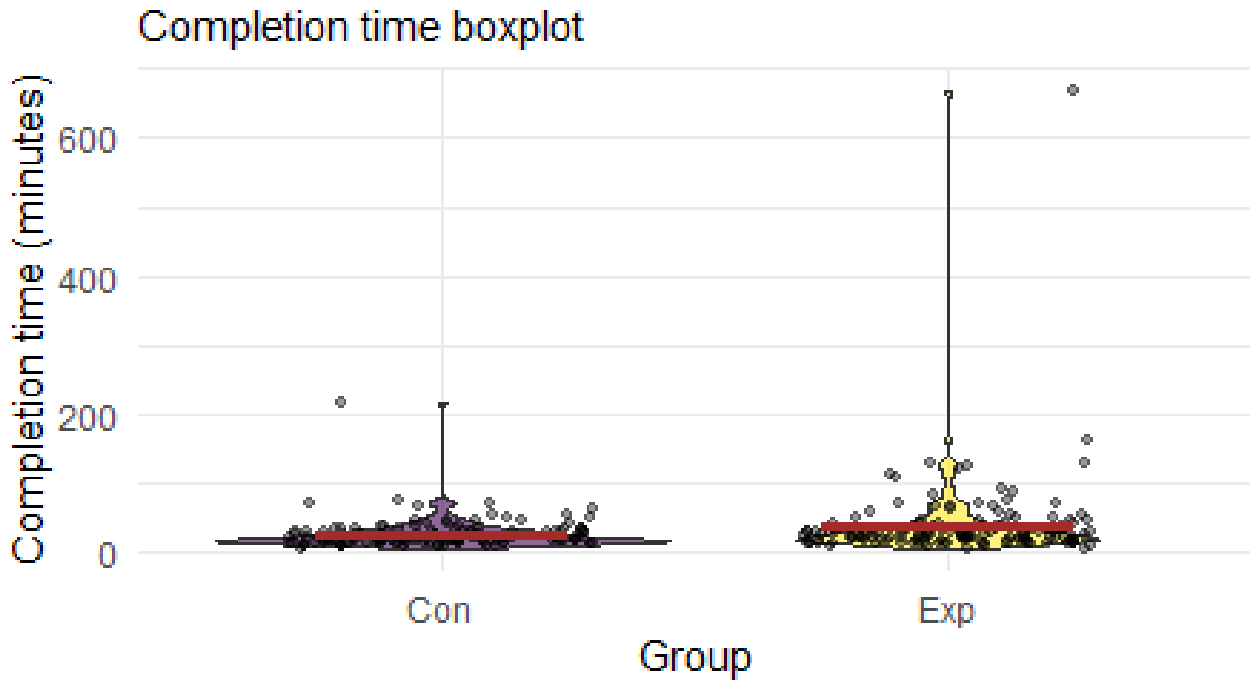


Figure 4. Violin plot showing the total time spent by the Control and Experimental group participants with Project Engage. Outliers indicate participants who completed the survey but did not close the browser tab.



Effects on Participants' Confidence

We ran multiple linear regressions with baseline confidence scores and intervention groups as predictors (Table 2 provides full regression results). The experimental group participants showed a significant increase in *in-the-moment* confidence compared with the control group participants for speaking in small groups (1.07 vs 0.08; P_{adj}=.01; Cohen *d*=0.34, 95% CI 0.10-0.58) and in the whole class (1.57 vs 0.12; P_{adj}<.001; Cohen *d*=0.51, 95% CI 0.26-0.74), but not for a one-on-one with the instructor (0.79 vs 0.13; P_{adj}=.05; Cohen *d*=0.27, 95% CI 0.03-0.51). Participants in the experimental group also reported significant increases in total confidence— calculated as the sum of the 3 subscales—compared with the control group (3.44 vs 0.34; P_{adj}<.001; Cohen *d*=0.48, 95% CI 0.24-0.72).

There was a significant difference between baseline total confidence scores between the experimental group participants whose (1) parents had attended or completed school, (2) parents had an associate degree, bachelor's degree, or had attended college but had no degree, and (3) parents had a master's or higher degree ($F_{2,134}$=4.59; *P*=.01). After the intervention, there was no significant difference between the total confidence in the 3 groups ($F_{2, 133}$=2.17, *P*=.12). We divided the experimental group into subgroups based on LGBTQ+ status (yes or no), disability status (yes or no), primary language (English or non-English), family income (low income, middle income, or high income), and parental education (first-generation college students or continuing-generation college students). We did not find significant differences in the baseline confidence levels of these subgroups.

Table 2. Results of multiple linear regression models predicting intervention effects on confidence scores.

	β (SE)	*P* value
Confidence in small group		
Intercept	3.24 (.65)	<.001
Condition	1.01 (.35)	.01[a]
Baseline score	.77 (.04)	<.001
Confidence in the whole class		
Intercept	1.69 (.42)	<.001
Condition	1.42 (.34)	<.001[a]
Baseline score	.82 (.04)	<.001
Confidence in one-on-one with the instructor		
Intercept	3.63 (.63)	<.001
Condition	.78 (.35)	.05[a]
Baseline score	.76 (.04)	<.001
Overall confidence		
Intercept	5.82 (1.57)	<.001
Condition	3.22 (.81)	<.001[a]
Baseline score	.85 (.04)	<.001

[a]*P* values have been adjusted using the Benjamini-Hochberg method.

Effects on DTS, IUS-12, and Intention to Persist in a Science-Related Career

Table 3 reports the baseline and postintervention scores (if available) in the different questionnaires administered to the participants. The changes in the scores were not statistically significant on DTS, IUS-12, or the intention to persist in science scale.

Table 3. Baseline and postintervention scores of participants recruited in the study.

Assessment scale	Time point, mean (SD)					
	Baseline		Postintervention		Postintervention—baseline change[a]	
	Experimental group	Control group	Experimental group	Control group	Experimental group	Control group
Brief Fear of Negative Evaluation-S	22.96 (8.73)	22.39 (9.34)	N/A[b]	N/A	N/A	N/A
Distress Tolerance Scale	56.69 (12.71)	60.07 (11.52)	58.01 (13.67)	60.36 (12.54)	1.32	0.29
Intolerance of Uncertainty Scale-12	31.49 (9.75)	29.99 (10.06)	29.79 (10.6)	29.33 (10.98)	–1.7	–0.66
Confidence						
Small group	14.04 (4.22)	14.21 (4.35)	15.11 (4.31)	14.29 (4.55)	1.07	0.08
Whole class	8.39 (4.34)	8.76 (4.40)	9.96 (4.63)	8.88 (4.45)	1.57	0.12
One-on-one	15.14 (4.21)	14.94 (4.87)	15.93 (4.23)	15.07 (4.71)	0.79	0.13
Total	37.55 (10.38)	37.91 (10.79)	40.99 (10.97)	38.25 (11.34)	3.44	0.34
Intention to persist in science	6.46 (2.78)	6.08 (2.99)	6.33 (2.85)	6.29 (2.94)	–0.13	0.21

[a]These values represent the change in the mean values of experimental and control group participants from baseline to postintervention. Hence, SD values are not available.

[b]N/A: not applicable (BFNE-S was not administered at postintervention).

Exploratory Analysis

Effects on Student Performance

Student performance data were available as 3 examination points (examination 1, examination 2, and final examination), participation points, and homework points. Out of these, examination 1 and examination 2 were conducted—before the intervention, and the final examination was conducted—after the intervention. We ran a multiple regression with intervention conditions (experimental=1, control=0) and an average of examination 1 percentage and examination 2 percentage as the predictor variables and final examination percentage as the predicted variable. We did not find any effects of the intervention condition on the final examination percentage (P=.96). We did not find any within-group effects on the examination performance for experimental (P=.15) or control group participants (P=.27). For participation and homework points, preintervention versus postintervention data were not available; therefore, no analysis was performed.

Analysis of Text-Based Responses for Experimental and Control Group Participants

The experimental and control group participants were asked to select the most anxiety-inducing situation in a class. The most commonly endorsed anxiety-inducing situation was "Being involuntarily asked to speak in front of the class" (Table 4). The most common feelings, thoughts, and behaviors were "Heart racing,""I shouldn't be this stressed out," and "Overthink responses," respectively. The experimental group participants were presented with 3 strategies (*noticing physiological feelings and trusting that they will reduce*, *changing interpretations*, and *practicing self-compassion*) to deal with the FNE. As a last step of Project Engage, we asked the experimental group participants to select a strategy that they would use to deal with anxiety. The most commonly selected strategy was *noticing physiological feelings and trusting that they will reduce.*

Table 4. Anxiety-inducing situations, and related feelings, thoughts, and behaviors selected by participants (N=282)[a].

	Selected by, n (%)
Situations	
Being involuntarily asked to speak in front of the class	151 (53.55)
Presenting in front of the class individually	95 (33.69)
Asking for extra help	11 (3.9)
Asking a question in front of everyone	10 (3.55)
Voluntarily answering a question in front of the class	4 (1.42)
Talking with the instructor one-on-one	4 (1.42)
Presenting in a group presentation	4 (1.42)
Participating in small group discussion	1 (0.35)
Feelings	
Heart racing	258 (91.49)
Sweaty palms	180 (63.83)
Trembling	156 (55.32)
Short, shallow breathing	115 (40.78)
Upset stomach	110 (39)
Headaches	61 (21.63)
Muscle aches	23 (8.16)
Thoughts	
I should not be this stressed out	184 (65.25)
Everyone will think I am dumb	182 (64.54)
The professor will think I am stupid	139 (49.29)
Other people can handle this situation—what's wrong with me?	134 (47.52)
Other students will laugh at me	108 (38.29)
I cannot handle this	103 (36.52)
People will make jokes about me if I get the wrong answer	89 (31.56)
The professor will be disappointed in me	73 (25.89)
This is the worst possible thing that could happen to me	68 (24.11)
Other students would not want to work with me	57 (20.21)
I am never going to succeed	48 (17.02)
Behaviors	
Overthink responses	202 (71.63)
Avoid eye contact	158 (56.03)
Prepare more	150 (53.19)
Struggle to think through things	127 (45.04)
Participate less in class	98 (34.75)
Try to get away from the stressful situation	97 (34.39)
Avoid talking	71 (25.18)
Consider dropping the class	37 (13.12)

[a]The numbers indicate how many participants selected each situation, feeling, thought, and behavior. Participants could select only one situation but multiple feelings, thoughts, and behaviors.

Discussion

Principal Findings

We have presented the results of a pilot trial evaluating a self-administered digital SSI designed to help students deal with the FNE in social-evaluative situations within the context of an active learning class. Results from the pilot trial conducted with 282 participants showed that the intervention was acceptable to the students and feasible for implementation in a large-enrollment classroom setting. In addition, the results showed that the experimental intervention improved *in-the-moment* confidence compared with the control intervention.

Overall, the SSI was acceptable to the participants, with most aspects receiving an average rating of >1 on a scale of –2 to +2. The aspect of whether the activity was enjoyable received a lower rating compared with the other aspects of the SSI. The intervention was developed and provided using Qualtrics. As Qualtrics is primarily a software designed to administer web-based surveys, it provides limited opportunities to create an interactive intervention, which might have contributed to the relatively low rating on enjoyability. In the future, bespoke software based on Project Engage might provide more opportunities to include interactive features and make the intervention more enjoyable. A bespoke software would also allow added opportunities to send reminders to the participants based on the strategies they have learned in the SSI.

The overall survey had a very high completion rate (>95%) compared with open trials of self-administered digital SSIs [29]. This indicates that the intervention is feasible for implementation and evaluation via a large-scale randomized controlled trial (RCT). Although we cannot determine the exact reason based on the available data, the high completion rate may be because of multiple factors: (1) acceptability and relevance of the intervention to the participant population, (2) being recommended by a trusted source (a course instructor in this case), (3) being offered as an assignment, and (4) being offered course points (although <1% of the total course points) for completing the survey. This finding has implications for the real-world dissemination of SSIs and other mental health interventions in university settings. Although being offered coupons or monetary benefits would be difficult to scale up in the real world, large-scale dissemination of interventions as assignments in college courses and as a way to earn course points may be relatively more feasible.

Compared with the control group, Project Engage was able to increase *in-the-moment* confidence immediately after the intervention among participants, with a medium effect size (Cohen *d*=0.48). The intervention also led to increased *in-the-moment* confidence for speaking in small groups and in the whole class. The corresponding effect sizes were medium (Cohen *d*=0.34) and large (Cohen *d*=0.51), respectively. Our results are comparable with the short-term (0-2 weeks) postintervention effects of other SSIs (Hedges *g*=0.46) [25].

The increase in confidence for speaking one-on-one with the instructor was not statistically significant. One potential explanation is that baseline confidence for speaking one-on-one with the instructor was high and therefore, there was less scope for improvement. Indeed, we found significant differences in the 3 subscales of confidence (small group discussion, whole class discussion, and one-on-one discussion with instructor) at baseline for the experimental group ($F_{2,\ 415}$=100.3, *P*<.001). Post hoc analyses using pairwise *t* tests and Benjamin-Hochberg correction revealed that baseline confidence was significantly lower for speaking in small groups (*P*=.03) and the whole class (*P*<.001) as compared with speaking one-on-one with the instructor.

In the pilot study, we did not observe an effect of Project Engage on *in-the-moment* scores on the DTS, the IUS-12, or the Intention to Persist in Science scale. Although we cannot know why the results were null, distress tolerance, intolerance of uncertainty, and the choice of a future career are stable attributes that may be unlikely to change within a short period of 30-40 minutes. In future RCTs, long-term follow-ups are required to check whether the intervention has a delayed effect on these attributes.

Comparison With Prior Work

Upon reviewing the literature, we were unable to find other digital SSIs aiming to improve student confidence in the context of higher education. However, other studies have evaluated single-session treatments targeting FNE and related constructs [58-60]. Hindo and González-Prendes [58] evaluated an SSI that gradually exposed participants from least to most anxiety-inducing situations in a group setting. The total intervention duration was 3 hours long and they found that the participants had reduced social anxiety and public speaking anxiety immediately after the intervention with a large effect size [58]. However, this study lacked a control group, and hence, the effect size might be inflated. Another study by Knutsson et al [59] evaluated a 90-minute therapist-led SSI delivering exposure or imagery rescripting-based treatments. They found that both treatments reduced FNE and that the gains were maintained at 4-week follow-up. Stefan et al [60] evaluated a 2 hours 30 minutes long web-based group session providing contextual schema therapy. This intervention reduced FNE from pre- to posttest compared with a waitlist control. Overall, these studies provide evidence that single-session treatments can help with the FNE and related constructs. Compared with these therapist-led interventions, Project Engage, being fully self-administered and digital, has a higher scalability.

Strengths and Limitations

Strengths of this study include its large sample size and low dropout rate. A limitation of this study was the lack of long-term follow-up data. Although Project Engage resulted in *in-the-moment* improvements on some measures compared with the active control group, it is unclear whether these improvements were sustained over time. This study did not collect such data as it aimed primarily to test the feasibility and acceptability of Project Engage. However, the promising results of this pilot trial suggest that collecting long-term follow-up data may be a future direction worth pursuing.

Another limitation of this study was the discrepancy between the experimental group and the control group regarding the time spent on the intervention. The experimental group participants spent significantly more time with the intervention compared with the control group participants. Although the activities were intended to be approximately time matched, this discrepancy may be because of the type of information presented in each activity. It is possible that the content presented in Project Engage provided more opportunities for reflection and deeper processing than the control condition, depending on perceived personal relevance and familiarity of the information to each student. In future studies, the active control condition may be altered to ensure that the time spent on the activity is comparable with that of the experimental intervention.

Given the measures of interest in this study—FNE, confidence in the classroom, distress tolerance, and intolerance of uncertainty—we had to rely on self-report scales, which are susceptible to common method bias.

Finally, although all participants were recruited from one upper-level physiology course at Arizona State University, they had diverse demographic backgrounds (Table 1). Consequently, our results might generalize to student populations at other US universities, but their generalizability beyond the US might be limited.

Future Directions

As college science courses continue to transition from traditional lecture format to active learning, bolstering student confidence as it relates to speaking in class is an important step in reducing FNE and improving overall student experiences [8]. This study demonstrates that the theoretical basis and scalability of clinical psychology interventions can be leveraged to improve student confidence at scale. There is substantial evidence that women have higher FNE compared with men [14,61,62]. In addition, nonbinary individuals, persons excluded because of their ethnicity or race, first-generation college students, LGBTQ+ students, and disabled students have also been found to express disproportionately high levels of FNE in the context of college science courses compared with their respective peer groups [14]. These differences in FNE likely help explain why women's voices are underrepresented in both whole-class and small-group discussions in college biology courses [63-66] and why there are suspected participation gaps among other underrepresented groups [67]. Therefore, it is imperative to test whether equipping students with the skills to cope with FNE may disproportionately benefit groups that are reluctant to participate in class for these reasons. Future studies can investigate whether Project Engage differentially affects marginalized individuals and whether the actual participation of students in these groups changes after being exposed to the intervention. Researchers hypothesize that enhanced participation may lead to additional long-term outcomes, including better performance [68,69], college retention rates, and decisions to pursue a career or leadership position in academic science [70,71]. As such, we view Project Engage as a critical step in working to create a more diverse and inclusive scientific community.

Conclusions

Active learning classes, although effective at improving students' learning, may exacerbate evaluation-related anxiety owing to the structure of coursework and learning activities, creating a need for interventions that support students in succeeding in these learning environments. Researchers from Science Education and Clinical Psychology worked together to create Project Engage, an intervention designed to provide psychoeducation and teach coping skills to students who experience FNE within active learning classes. In a pilot RCT, Project Engage had a high completion rate and received a high acceptability rating. These results demonstrate that a digital, self-administered SSI can be feasibly implemented in large classroom settings. In addition, participants randomized to complete Project Engage reported greater increases in overall confidence in engaging in small group discussions and whole class discussions, from baseline to immediately after intervention, compared with participants in an active control condition. Fully powered RCTs investigating Project Engage with a longer follow-up period are warranted to determine whether these effects are sustained over time. Such an investigation may provide stronger evidence regarding the efficacy of Project Engage in improving the classroom engagement and coping skills of students in active learning classes.

Acknowledgments

The authors would like to thank Dr Yi Zheng for helping develop the questionnaires to assess confidence. The authors would like to thank all the members of the Lab for Scalable Mental Health and the Cooper Biology Education Research Lab for going through the first version of Project Engage and providing feedback. This study was funded by an award to KMC and JLS from the National Science Foundation (NSF; 2141710). Outside the present work, KMC received funding from the NSF. JLS received funding from the National Institute of Health Office of the Director (DP5OD028123), the National Institute of Mental Health (R43MH128075), the Upswing Fund for Adolescent Mental Health, Health Research and Services Association (U3NHP45406-01-00), the Society for Clinical Child and Adolescent Psychology, HopeLab, and the Klingenstein Third Generation Foundation. Preparation of this paper was supported in part by the Implementation Research Institute (IRI) at the George Warren Brown School of Social Work, Washington University in St. Louis, through an award from the National Institute of Mental Health (R25MH080916; JLS is an IRI Fellow). CAB was supported by an NSF Graduate Research Fellowship (grant #026257-001). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Data Availability

Deidentified data may be available on reasonable request.

Authors' Contributions

KMC and JLS conceptualized this project. KMC, JLS, and AG contributed to the study design. KMC and CAB collected the data. JLS, KMC, AG, and KAC led the development of Project Engage with contributions from all authors. CAB and LJ led the development of the control version with contributions from all the authors. AG performed all data analyses. KMC and KAC wrote the *Introduction* section; LJ wrote the *Methods* section; AG wrote the *Results* section; and AG, KAC, and KMC wrote the *Discussion* section. JLS and KMC edited the manuscript. All the authors contributed to the review and editing of the final manuscript.

Conflicts of Interest

JLS serves on the Scientific Advisory Board for Walden Wise and the Clinical Advisory Board for Koko; receives consulting fees from Kooth, LLC and Woebot Health; is the cofounder and codirector of Single Session Support Solutions; and receives book royalties from New Harbinger, Oxford University Press, and Little Brown Book Group. KAC receives consulting fees from Kooth, LLC and Koko.

Multimedia Appendix 1
The full questionnaire developed for the study.
[DOCX File , 16 KB - mental_v10i1e48926_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1164 KB - mental_v10i1e48926_app2.pdf]

References

1. Vision and change in undergraduate biology education: a call to action. The American Association for the Advancement of Science (AAAS). 2009. URL: https://www.aaas.org/sites/default/files/content_files/VC_report.pdf [accessed 2023-11-04]
2. Vision and change in undergraduate biology education: chronicling change, inspiring the future. The American Association for the Advancement of Science. 2015. URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3f3a12e5f2e78e513e0a7d30cfd8b338514b9b34 [accessed 2023-11-04]
3. Driessen EP, Knight JK, Smith MK, Ballen CJ. Demystifying the meaning of active learning in postsecondary biology education. CBE Life Sci Educ 2020 Dec;19(4):ar52 [FREE Full text] [doi: 10.1187/cbe.20-04-0068] [Medline: 33001767]
4. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, et al. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A 2014 Jun 10;111(23):8410-8415 [FREE Full text] [doi: 10.1073/pnas.1319030111] [Medline: 24821756]
5. Theobald EJ, Hill MJ, Tran E, Agrawal S, Arroyo EN, Behling S, et al. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. Proc Natl Acad Sci U S A 2020 Mar 24;117(12):6476-6483 [FREE Full text] [doi: 10.1073/pnas.1916903117] [Medline: 32152114]
6. Cooper KM, Downing VR, Brownell SE. The influence of active learning practices on student anxiety in large-enrollment college science classrooms. Int J STEM Educ 2018;5(1):23 [FREE Full text] [doi: 10.1186/s40594-018-0123-6] [Medline: 30631713]
7. Cooper KM, Brownell SE. Student anxiety and fear of negative evaluation in active learning science classrooms. In: Mintzes JJ, Walter EM, editors. Active Learning in College Science. Cham, Switzerland: Springer; Feb 24, 2020.
8. Downing VR, Cooper KM, Cala JM, Gin LE, Brownell SE. Fear of negative evaluation and student anxiety in community college active-learning science courses. CBE Life Sci Educ 2020 Jun;19(2):ar20 [FREE Full text] [doi: 10.1187/cbe.19-09-0186] [Medline: 32453679]
9. England BJ, Brigati JR, Schussler EE. Student anxiety in introductory biology classrooms: perceptions about active learning and persistence in the major. PLoS One 2017 Aug 3;12(8):e0182506 [FREE Full text] [doi: 10.1371/journal.pone.0182506] [Medline: 28771564]
10. Hood S, Barrickman N, Djerdjian N, Farr M, Magner S, Roychowdhury H, et al. "I like and prefer to work alone": social anxiety, academic self-efficacy, and students' perceptions of active learning. CBE Life Sci Educ 2021 Mar;20(1):ar12 [FREE Full text] [doi: 10.1187/cbe.19-12-0271] [Medline: 33600218]
11. Watson D, Friend R. Measurement of social-evaluative anxiety. J Consult Clin Psychol 1969 Aug;33(4):448-457. [doi: 10.1037/h0027806] [Medline: 5810590]

12. Weeks JW, Heimberg RG, Fresco DM, Hart TA, Turk CL, Schneier FR, et al. Empirical validation and psychometric evaluation of the brief fear of negative evaluation scale in patients with social anxiety disorder. Psychol Assess 2005 Jun;17(2):179-190. [doi: 10.1037/1040-3590.17.2.179] [Medline: 16029105]
13. MacIntyre PD, Gardner RC. Language anxiety: its relationship to other anxieties and to processing in native and second languages*. Lang Learn 1991 Dec;41(4):513-534. [doi: 10.1111/j.1467-1770.1991.tb00691.x]
14. Busch CA, Wiesenthal NJ, Mohammed TF, Anderson S, Barstow M, Custalow C, et al. The disproportionate impact of fear of negative evaluation on first-generation college students, LGBTQ+ students, and students with disabilities in college science courses. CBE Life Sci Educ 2023 Sep;22(3):ar31 [FREE Full text] [doi: 10.1187/cbe.22-10-0195] [Medline: 37347813]
15. Heimberg RG, Brozovich FA, Rapee RM. A cognitive behavioral model of social anxiety disorder: update and extension. In: Social Anxiety (Second Edition). Cambridge, UK: Academic Press; 2010.
16. Sweller J. Cognitive load theory, learning difficulty, and instructional design. Learn Instr 1994;4(4):295-312. [doi: 10.1016/0959-4752(94)90003-5]
17. McConnell DA, Stempien JA, Perkins D, Van Der Hoeven Kraft KJ, Vislova T, Wirth KR, et al. The little engine that could – less prior knowledge but high self-efficacy is equivalent to greater prior knowledge and low self-efficacy. In: Proceedings of the 2010 Geological Society of America Annual Meeting. 2010 Presented at: 2010 Geological Society of America Annual Meeting; October 31-November 3, 2010; Denver, CO URL: https://gsa.confex.com/gsa/2010AM/webprogram/Paper181041.html
18. Yannier N, Hudson SE, Koedinger KR, Hirsh-Pasek K, Golinkoff RM, Munakata Y, et al. Active learning: "hands-on" meets "minds-on". Science 2021 Oct;374(6563):26-30. [doi: 10.1126/science.abj9957] [Medline: 34591619]
19. Yep BL, Tan TK, Fung FM. How partial anonymity may reduce students' anxiety during remote active learning—a case study using clubhouse. J Chem Educ 2023 Jan 20;100(2):459-468. [doi: 10.1021/acs.jchemed.2c00051]
20. Dallimore EJ, Hertenstein JH, Platt MB. Impact of cold-calling on student voluntary participation. J Manag Educ 2012 May 08;37(3):305-341. [doi: 10.1177/1052562912446067]
21. Heimberg RG. Cognitive-behavioral therapy for social anxiety disorder: current status and future directions. Biol Psychiatry 2002 Jan 01;51(1):101-108. [doi: 10.1016/s0006-3223(01)01183-0] [Medline: 11801235]
22. Kindred R, Bates GW, McBride NL. Long-term outcomes of cognitive behavioural therapy for social anxiety disorder: a meta-analysis of randomised controlled trials. J Anxiety Disord 2022 Dec;92:102640. [doi: 10.1016/j.janxdis.2022.102640] [Medline: 36265270]
23. Mayo-Wilson E, Dias S, Mavranezouli I, Kew K, Clark DM, Ades AE, et al. Psychological and pharmacological interventions for social anxiety disorder in adults: a systematic review and network meta-analysis. Lancet Psychiatry 2014 Oct;1(5):368-376 [FREE Full text] [doi: 10.1016/S2215-0366(14)70329-3] [Medline: 26361000]
24. Schleider JL, Dobias ML, Sung JY, Mullarkey MC. Future directions in single-session youth mental health interventions. J Clin Child Adolesc Psychol 2020;49(2):264-278 [FREE Full text] [doi: 10.1080/15374416.2019.1683852] [Medline: 31799863]
25. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
26. Schleider JL, Mullarkey MC, Fox KR, Dobias ML, Shroff A, Hart EA, et al. A randomized trial of online single-session interventions for adolescent depression during COVID-19. Nat Hum Behav 2022 Feb;6(2):258-268 [FREE Full text] [doi: 10.1038/s41562-021-01235-0] [Medline: 34887544]
27. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
28. Dobias ML, Morris RR, Schleider JL. Single-session interventions embedded within tumblr: acceptability, feasibility, and utility study. JMIR Form Res 2022 Jul 25;6(7):e39004 [FREE Full text] [doi: 10.2196/39004] [Medline: 35877163]
29. Schleider JL, Dobias M, Sung J, Mumper E, Mullarkey MC. Acceptability and utility of an open-access, online single-session intervention platform for adolescent mental health. JMIR Ment Health 2020 Jun 30;7(6):e20513 [FREE Full text] [doi: 10.2196/20513] [Medline: 32602846]
30. Bannink R, Broeren S, Joosten-van Zwanenburg E, van As E, van de Looij-Jansen P, Raat H. Effectiveness of a web-based tailored intervention (E-health4Uth) and consultation to promote adolescents' health: randomized controlled trial. J Med Internet Res 2014 May 30;16(5):e143 [FREE Full text] [doi: 10.2196/jmir.3163] [Medline: 24878521]
31. Miu AS, Yeager DS. Preventing symptoms of depression by teaching adolescents that people can change: effects of a brief incremental theory of personality intervention at 9-month follow-up. Clin Psychol Sci 2014 Sep 15;3(5):726-743. [doi: 10.1177/2167702614548317]
32. Osborn TL, Rodriguez M, Wasil AR, Venturo-Conerly KE, Gan J, Alemu RG, et al. Single-session digital intervention for adolescent depression, anxiety, and well-being: outcomes of a randomized controlled trial with Kenyan adolescents. J Consult Clin Psychol 2020 Jul;88(7):657-668. [doi: 10.1037/ccp0000505] [Medline: 32391709]

33. Schleider JL, Burnette JL, Widman L, Hoyt C, Prinstein MJ. Randomized trial of a single-session growth mind-set intervention for rural adolescents' internalizing and externalizing problems. J Clin Child Adolesc Psychol 2020;49(5):660-672 [FREE Full text] [doi: 10.1080/15374416.2019.1622123] [Medline: 31219698]
34. Samson JE, Tanner-Smith EE. Single-session alcohol interventions for heavy drinking college students: a systematic review and meta-analysis. J Stud Alcohol Drugs 2015 Jul;76(4):530-543 [FREE Full text] [doi: 10.15288/jsad.2015.76.530] [Medline: 26098028]
35. Zomback L, Barré LK. Randomized control trial of a single-session intervention targeting sociocultural pressures of body image in college sorority women. J Am Coll Health 2022 Sep 28:1-5. [doi: 10.1080/07448481.2022.2124379] [Medline: 36170556]
36. Bettinger E, Ludvigsen S, Rege M, Solli IF, Yeager D. Increasing perseverance in math: evidence from a field experiment in Norway. J Econ Behav Organ 2018 Feb;146:1-15. [doi: 10.1016/j.jebo.2017.11.032]
37. Broda M, Yun J, Schneider B, Yeager DS, Walton GM, Diemer M. Reducing inequality in academic success for incoming college students: a randomized trial of growth mindset and belonging interventions. J Res Educ Eff 2018 Apr 09;11(3):317-338. [doi: 10.1080/19345747.2018.1429037]
38. Yeager DS, Hanselman P, Walton GM, Murray JS, Crosnoe R, Muller C, et al. A national experiment reveals where a growth mindset improves achievement. Nature 2019 Sep;573(7774):364-369 [FREE Full text] [doi: 10.1038/s41586-019-1466-y] [Medline: 31391586]
39. Keough ME, Riccardi CJ, Timpano KR, Mitchell MA, Schmidt NB. Anxiety symptomatology: the association with distress tolerance and anxiety sensitivity. Behav Ther 2010 Dec;41(4):567-574 [FREE Full text] [doi: 10.1016/j.beth.2010.04.002] [Medline: 21035619]
40. Carleton RN, Collimore KC, Asmundson GJ. "It's not just the judgements--it's that I don't know": intolerance of uncertainty as a predictor of social anxiety. J Anxiety Disord 2010 Mar;24(2):189-195. [doi: 10.1016/j.janxdis.2009.10.007] [Medline: 19931391]
41. England BJ, Brigati JR, Schussler EE, Chen MM. Student anxiety and perception of difficulty impact performance and persistence in introductory biology courses. CBE Life Sci Educ 2019 Jun;18(2):ar21 [FREE Full text] [doi: 10.1187/cbe.17-12-0284] [Medline: 31120397]
42. Project engage: equipping biology undergraduates with skills to cope with fear of negative evaluation in active learning classrooms through a single session intervention. Open Science Framework. URL: https://osf.io/4ca68 [accessed 2023-11-12]
43. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
44. Project engage pilot - exp.pdf. Open Science Framework. URL: https://osf.io/m4q2z [accessed 2023-11-12]
45. Deci EL, Ryan RM. Self-determination theory. In: Handbook of Theories of Social Psychology: Volume 1. Thousand Oaks, CA: SAGE Publications; 2012.
46. Schleider JL, Sung J, Bianco A, Gonzalez A, Vivian D, Mullarkey MC. Open pilot trial of a single-session consultation service for clients on psychotherapy wait-lists. PsyArXiv Preprints. Preprint posted online May 29, 2020 2023 [FREE Full text] [doi: 10.31234/osf.io/fdwqk]
47. Schleider JL, Beidas RS. Harnessing the single-session intervention approach to promote scalable implementation of evidence-based practices in healthcare. Front Health Serv 2022 Sep 23;2:997406 [FREE Full text] [doi: 10.3389/frhs.2022.997406] [Medline: 36925822]
48. Schleider JL, Mullarkey MC, Weisz JR. Virtual reality and web-based growth mindset interventions for adolescent depression: protocol for a three-arm randomized trial. JMIR Res Protoc 2019 Jul 09;8(7):e13368 [FREE Full text] [doi: 10.2196/13368] [Medline: 31290406]
49. Schleider JL, Abel MR, Weisz JR. Do immediate gains predict long-term symptom change? Findings from a randomized trial of a single-session intervention for youth anxiety and depression. Child Psychiatry Hum Dev 2019 Oct;50(5):868-881 [FREE Full text] [doi: 10.1007/s10578-019-00889-2] [Medline: 30993499]
50. Simons JS, Gaher RM. The distress tolerance scale: development and validation of a self-report measure. Motiv Emot 2005;29:83-102. [doi: 10.1007/s11031-005-7955-3]
51. Carleton RN, Norton MP, Asmundson GJ. Fearing the unknown: a short version of the intolerance of uncertainty scale. J Anxiety Disord 2007;21(1):105-117. [doi: 10.1016/j.janxdis.2006.03.014] [Medline: 16647833]
52. Martin JP, Miller MK, Gipson K. Utilization of a think-aloud protocol to cognitively validate a survey instrument identifying social capital resources of engineering undergraduates. In: Proceedings of the 2011 American Society for Engineering Education Annual Conference and Exhibition. 2011 Presented at: 2011 American Society for Engineering Education Annual Conference and Exhibition; June 26-29, 2011; Vancouver, BC URL: https://www.researchgate.net/publication/280309582_Utilization_of_a_Think-Aloud_Protocol_to_Cognitively_Validate_a_Survey_Instrument_Identifying_Social_Capital_Resources_of_Engineering_Undergraduates [doi: 10.18260/1-2--18492]
53. Estrada M, Woodcock A, Hernandez PR, Schultz PW. Toward a model of social influence that explains minority student integration into the scientific community. J Educ Psychol 2011 Feb 01;103(1):206-222 [FREE Full text] [doi: 10.1037/a0020743] [Medline: 21552374]

54. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
55. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
56. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2022. URL: https://www.R-project.org/ [accessed 2023-11-02]
57. Buchanan E, Gillenwaters A, Scofield J, Valentine K. MOTE: measure of the effect: package to assist in effect size calculations and their confidence intervals. The Comprehensive R Archive Network. 2019. URL: https://cran.r-project.org/web/packages/MOTE/ [accessed 2023-11-02]
58. Hindo CS, González-Prendes AA. One-session exposure treatment for social anxiety with specific fear of public speaking. Res Soc Work Pract 2011 Aug 04;21(5):528-538. [doi: 10.1177/1049731510393984]
59. Knutsson J, Nilsson JE, Eriksson Å, Järild L. Imagery rescripting and exposure in social anxiety: a randomized trial comparing treatment techniques. J Contemp Psychother 2019 Dec 21;50:233-240. [doi: 10.1007/s10879-019-09448-1]
60. Stefan S, Stroian P, Fodor L, Matu S, Nechita D, Boldea I, et al. An online contextual schema therapy workshop for social anxiety symptoms – a randomized controlled trial. J Contextual Behav Sci 2023 Jul;29:67-75. [doi: 10.1016/j.jcbs.2023.06.005]
61. Al-Johani WM, AlShamlan NA, AlAmer NA, Shawkhan RA, Almayyad AH, Alghamdi LM, et al. Social anxiety disorder and its associated factors: a cross-sectional study among medical students, Saudi Arabia. BMC Psychiatry 2022 Jul 27;22(1):505 [FREE Full text] [doi: 10.1186/s12888-022-04147-z] [Medline: 35896983]
62. Yue Y, Jia Y. Fear of negative evaluation: a cross-sectional study among undergraduate nursing students. Nurse Educ Today 2023 Feb;121:105678. [doi: 10.1016/j.nedt.2022.105678] [Medline: 36527756]
63. Aguillon SM, Siegmund GF, Petipas RH, Drake AG, Cotner S, Ballen CJ. Gender differences in student participation in an active-learning classroom. CBE Life Sci Educ 2020 Jun;19(2):ar12 [FREE Full text] [doi: 10.1187/cbe.19-03-0048] [Medline: 32453677]
64. Ballen CJ, Aguillon SM, Awwad A, Bjune AE, Challou D, Drake AG, et al. Smaller classes promote equitable student participation in STEM. BioScience 2019 Aug;69(8):669-680. [doi: 10.1093/biosci/biz069]
65. Eddy SL, Brownell SE, Wenderoth MP. Gender gaps in achievement and participation in multiple introductory biology classrooms. CBE Life Sci Educ 2014;13(3):478-492 [FREE Full text] [doi: 10.1187/cbe.13-10-0204] [Medline: 25185231]
66. Nadile EM, Williams KD, Wiesenthal NJ, Stahlhut KN, Sinda KA, Sellas CF, et al. Gender differences in student comfort voluntarily asking and answering questions in large-enrollment college science courses. J Microbiol Biol Educ 2021 Jun 30;22(2):e00100-e00121 [FREE Full text] [doi: 10.1128/jmbe.00100-21] [Medline: 34594434]
67. Waugh AH, Andrews TC. Diving into the details: constructing a framework of random call components. CBE Life Sci Educ 2020 Jun;19(2):ar14 [FREE Full text] [doi: 10.1187/cbe.19-07-0130] [Medline: 32357096]
68. LaDage LD, Tornello SL, Vallejera JM, Baker EE, Yan Y, Chowdhury A. Variation in behavioral engagement during an active learning activity leads to differential knowledge gains in college students. Adv Physiol Educ 2018 Mar 01;42(1):99-103 [FREE Full text] [doi: 10.1152/advan.00150.2017] [Medline: 29357267]
69. Reinholz D, Johnson E, Andrews-Larson C, Stone-Johnstone A, Smith , J, Mullins B, et al. When active learning is inequitable: women's participation predicts gender inequities in mathematical performance. J Res Math Educ 2022 May;53(3):204-226. [doi: 10.5951/jresematheduc-2020-0143]
70. Braxton JM, Milem JF, Sullivan AS. The influence of active learning on the college student departure process: toward a revision of Tinto's theory. J Higher Educ 2000;71(5):569-590. [doi: 10.2307/2649260]
71. Braxton JM, Jones WA, Hirschy AS, Hartley HVIII. The role of active learning in college student persistence. New Dir Teach Learn 2008;115(115):71-83. [doi: 10.1002/tl.326]

Abbreviations

BFNE-S: Brief Fear of Negative Evaluation
CBT: cognitive behavioral therapy
CFI: Comparative Fit Index
CONSORT: Consolidated Standards of Reporting Trials
DTS: Distress Tolerance Scale
FNE: fear of negative evaluation
IUS-12: Intolerance of Uncertainty Scale-12
LGBTQ+: lesbian, gay, bisexual, transgender, queer, and others
PFS: Program Feedback Scale
RCT: randomized controlled trial
SSI: single-session intervention
STEM: Science, Technology, Engineering, and Medicine

Edited by J Torous; submitted 11.05.23; peer-reviewed by J Bautista, C Brogly, C Laranjeira; comments to author 17.06.23; revised version received 15.09.23; accepted 06.10.23; published 23.11.23.

Please cite as:
Ghosh A, Cohen KA, Jans L, Busch CA, McDanal R, Yang Y, Cooper KM, Schleider JL
A Digital Single-Session Intervention (Project Engage) to Address Fear of Negative Evaluation Among College Students: Pilot Randomized Controlled Trial
JMIR Ment Health 2023;10:e48926
URL: https://mental.jmir.org/2023/1/e48926
doi:10.2196/48926
PMID:37995114

Original Paper

Individualized Virtual Reality for Increasing Self-Compassion: Evaluation Study

Ilona Halim[1], MInfSc; Lehan Stemmet[2], PhD; Sylvia Hach, PhD; Richard Porter[3], MD; Hai-Ning Liang[4], PhD; Atiyeh Vaezipour[5], PhD; Julie D Henry[6], PhD; Nilufar Baghaei[7], PhD

[1]School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand

[2]Auckland Institute of Studies, Auckland, New Zealand

[3]Department of Psychological Medicine, University of Otago, Christchurch, New Zealand

[4]Department of Computing, Xi'an Jiaotong-Liverpool University, Suzhou, China

[5]Recover Injury Research Centre, The University of Queensland, St Lucia, Australia

[6]School of Psychology, The University of Queensland, St Lucia, Australia

[7]School of Electrical Engineering and Computer Science, The University of Queensland, St Lucia, Australia

Corresponding Author:
Nilufar Baghaei, PhD
School of Electrical Engineering and Computer Science
The University of Queensland
Staff House Road
St Lucia, 4072
Australia
Phone: 61 450150234
Email: n.baghaei@uq.edu.au

Abstract

Background: Depression and anxiety are common and debilitating mental disorders with severe negative repercussions at both individual and societal levels. Although virtual reality (VR) has emerged as a safe and effective tool for the treatment of anxiety disorders, studies of the therapeutic application of VR to treat depression are more limited.

Objective: The purpose of this study was to test whether a novel type of individualized VR (iVR) can be used to improve self-compassion and decrease depressive symptoms and to evaluate the usability and acceptability of this approach, as rated by participants. The iVR system was designed and developed based on the feedback obtained from a previous study, with improved appearance and feel of the avatar and enhanced graphical quality.

Methods: A total of 36 young adult participants were recruited from a university community social media site. Participants were aware that the study was investigating a treatment for depression but were not recruited based on depression diagnosis. Participants were asked to complete 2 iVR sessions, spaced 2 weeks apart. At baseline and upon completion of each iVR session, participants were asked to complete validated measures of self-compassion and depression. Upon completion of both iVR sessions, additional measures were administered to assess participants' perceptions about the perceived usability and system acceptability of the iVR approach.

Results: Self-compassion was assessed at the beginning of session 1 (preintervention baseline) and at the end of session 1 (postintervention assessment). Owing to COVID-19 constraints, 36% (13/36) of the participants were unable to complete the follow-up iVR session. Self-compassion was assessed again for the remaining 64% (23/36) of the participants at the end of session 2 (postintervention assessment). Within-group analyses revealed that self-compassion was significantly increased at the end of both session 1 (P=.01) and session 2 (P=.03) relative to baseline. There was also a nonsignificant trend for depressive symptoms to be low at the end of session 2 relative to baseline. Both quantitative and qualitative participant data supported the iVR approach as being acceptable and usable.

Conclusions: Although these data must be treated as preliminary owing to the small sample size and potential selection bias, the data provide encouraging initial evidence that iVR might be a useful tool to enhance self-compassion and reduce depressive symptoms, highlighting the need for randomized controlled trials in the future.

(JMIR Ment Health 2023;10:e47617) doi:10.2196/47617

KEYWORDS

individualized virtual reality; mental health; self-compassion; depression; depressive symptoms; mobile phone

Introduction

Background

Depression is a common mental disorder and a leading cause of disability that has been identified by the World Health Organization as a major contributor to the overall global burden of disease [1]. Psychological and pharmacological treatments, such as cognitive behavioral therapy, interpersonal psychotherapy, and antidepressant medication, are commonly used to treat depression [1]. However, despite the availability of treatment, it has been estimated that between 76% and 85% of people in low- and middle-income nations do not obtain adequate care owing to lack of resources and trained health care professionals [1,2]. Although steps have been taken to try and overcome these obstacles, more effective and accessible care is still urgently required [1].

Studies suggest that there is a link between self-compassion or self-criticism and depressive symptoms, with the former considered as a protective factor and the latter as a risk factor for depression [3]. Such links are unsurprising, given that self-compassion involves treating oneself with kindness and understanding [4], whereas self-criticism is defined by negative self-evaluation and judgment [5]. Low self-compassion and high self-criticism are associated with more severe depressive symptoms and increased risk of chronic or recurring depression [6].

Importantly, studies suggest that self-compassion can be increased with training [7]. Compassion-focused therapy (CFT) is therefore a possible method to reduce symptoms of depression. With CFT, the aim of the treatment is to increase a person's self-compassion levels through strategic exposure to situations or scenarios that encourage demonstration of compassion toward themselves or others and allow the client to focus on self-compassion when they experience negative thought processes [8]. However, although this method has been shown to be effective, it can be challenging and resource intensive to implement; thus, there is a need for innovative interventions that are more accessible and sustainable [9].

Virtual Reality in Mental Health

Virtual Reality (VR) is an emerging technology that can simulate realistic and immersive experiences within a virtual world. VR head-mounted displays (HMDs) are becoming more affordable and accessible, and there is a growing interest in their potential use in psychotherapeutic interventions [10,11]. VR exposure therapy (VRET) is one of the most common VR applications in mental health [12-14]. This exposure therapy strategy takes advantage of the high degree of presence provided by VR environments. Recent studies indicate that VR can be effective in the treatment of a variety of mental disorders, pain management, and addiction, where gains from simulated virtual exposure therapy transfer to real-life situations [15].

VR has been used to provide exposure therapy in the treatment of phobia, where people are presented with a simulation of their source of fear in a safe and controlled situation [16]. Positive results have been reported in the treatment of acrophobia, trypanophobia, claustrophobia, and vehophobia and for more complex anxiety disorders including panic disorder, social anxiety disorder, and arachnophobia [17].

In particular, intensive studies have focused on the use of VR for the treatment of acrophobia (fear of heights). Krijn et al [18] investigated whether cognitive coping self-statements would have additional benefits over VRET in people with acrophobia. Participants (N=26) were randomly assigned to receive either 2 sessions of VRET and coping self-statements followed by 2 sessions of VRET or 2 sessions of VRET followed by 2 sessions of VRET and coping self-statements. Findings indicated that VRET reduced height-related anxiety and avoidance behavior, regardless of whether coping self-statements were added. In a study conducted by Meyerbroker et al [12], VRET was used to treat acrophobia and fear of flying. Findings showed that VRET with a placebo is an effective treatment for acrophobia. More recently, Donker et al [19] conducted a study on the use of a smartphone VR app for VRET. In total, 193 participants with acrophobia were randomly assigned to a VR cognitive behavioral treatment group using cardboard VR goggles (n=96, 49.7%) or to a waiting list control group (n=97, 50.3%). Results showed reduction in acrophobia symptoms after therapy and at 3-month follow-up.

In the treatment of people with posttraumatic stress disorder (PTSD), VR has also been used as a tool for exposure therapy by presenting simulated scenarios that trigger the trauma in combination with relaxation training [20]. Although prolonged exposure involving emotional processing of traumatic situations and habituation of anxiety is an evidence-based treatment for PTSD, some people are either unwilling to undergo this treatment or are unable to visualize the traumatic stimuli because of anxiety avoidance [21]. For people living with PTSD, exposure to a controlled fictitious virtual environment might therefore be more tolerable than exposure by imagination [21]. Several randomized controlled trials (RCTs) have compared VRET with prolonged imaginal exposure therapy in active-duty US military personnel and veterans. The VRET sessions included imaginal exposure to traumatic war memories and computer-generated environments, tailored to participants' description of their trauma, and revealed that VRET was just as effective as extended exposure [22]. Although a large RCT comparing prolonged imaginal exposure with VRET in active-duty personnel found that prolonged exposure was more effective than VRET at 3-month and 6-month follow-up [23], Norr et al [24] found that relative to a waitlist control, both prolonged exposure and VRET led to decreased incidence of suicidal thoughts.

VR has also been used to assess mental health disorders, albeit less frequently. For instance, it has been used to simulate circumstances in which addiction-related behaviors might be triggered, so that the therapist can watch the client's response and use that knowledge to plan the patient's therapy [25].

Extensive literature shows that alcohol-related and drug-related cues (such as cigarettes, ashtrays, and smokers at a bar) contribute to the persistence of substance use disorders [26], and this extends to the virtual world. Moreover, peer pressure from virtual avatars is highly effective in inducing cravings [12,27]. Collectively, the findings from these studies show that immersing individuals in virtual worlds related to their specific addiction might be a valuable tool in both their assessment and treatment.

VR in the Treatment of Depression

In a scoping study conducted by Baghaei et al [14], it was argued that VR could be beneficial in helping the treatment of anxiety or depression in a variety of contexts, and its potential as a clinical tool was highlighted. However, despite this emerging literature about the use of VR for mental health conditions, relatively few studies so far have focused on depression, one of the most common conditions [28]. Of the limited number of studies that have tested the potential value of VR in treating depression, the first study used the standard nonindividualized approach. Li et al [29] investigated the impact of a VR-based restorative environment on the emotional and cognitive recovery of patients with mild to moderate depression and anxiety. The study enrolled 195 individuals with mild to moderate depression and anxiety. The study used multiple VR restorative contexts, including an urban and a park setting. Participants were divided into groups at random and each engaged with a different scene. According to the findings of the study, a VR-based restorative environment can improve the emotional well-being and cognitive recovery of people with mild to moderate anxiety and depression. Falconer et al [28] examined the effect of avatar embodiment on self-compassion in people diagnosed with clinical depression. In this study, 15 participants engaged in an 8-minute VR simulation in which they experienced delivering and receiving compassion in virtual bodies. This session was repeated 3 times, and at the conclusion of the trial, 9 participants reported improvement in their depressive symptoms. Our group later followed up this study, by investigating whether *individualization* of VR might increase the potential benefits of compassion for depressive symptoms [30]. An early version of an individualized VR (iVR) system was built and implemented and was evaluated [31]. In this study, participants were able to individualize their VR experience by selecting their preferred avatar, therapeutic environment, and treatment scenario.

Individualization of VR

The potential benefits of *personalizing* virtual experiences for mental health therapy, such as customizing the virtual environment to an individual's preferences and interests, has been the focus of limited studies. Most VR therapies have been predefined and predesigned, resulting in a uniform approach for all participants based on an immersion or presence-specific environment [32]. Although this approach has the advantage of being standardized, it does not account for important individual differences in characteristics, which may influence the treatment's efficacy [33]. iVR can also improve user experience and facilitate therapists' access to otherwise difficult-to-obtain clinical information [14,31].

Avatars are common across digital applications. By using avatars as self-representation, we engage in social activities, play, and conduct businesses. There has been an increasing interest in the potential effects of avatar customization on both user experience and performance indicators [34]. Customization of avatars has been found to positively influence outcomes in a variety of fields. In players who personalized their game characters, increased identification with their avatars was found to increase autonomy, invested effort, enjoyment, and immersion [35]. Waltemate et al [36] examined the effect of the degree of personalization and individualization of users' avatars and the level of immersion on typical psychophysical variables in embodied virtual environments. Personalized avatars in VR were found to enhance body ownership, sense of presence, and dominance in comparison with generic counterparts of comparable realism and quality of graphics. Such findings demonstrate the potential value of individualized avatars in VR applications that rely on body ownership and presence. As the architecture of iVR mainly relies on avatar embodiment [31,32], avatar customization should theoretically enhance the application. As therapists may also view the individual's choice of avatar in each session, the type of avatar characteristics selected also provides further unique insights into their client's mental state.

However, although the results from our previous study [31] provided important proof of concept in showing that individualization of the VR experience may have value in therapeutic contexts, this study did not examine the effects of iVR on self-compassion levels and depression symptoms specifically. Therefore, the central aim of this study was to report the design and implementation of a novel VR system that builds on but meaningfully improves the systems used in our early pilot study [31,32] and then to use this refined system to provide the first evaluation of the implemented iVR's impact on self-compassion and depressive symptoms.

iVR Design and Implementation

iVR Description

Our prototype iVR system mirrors the prototype used by Falconer et al [28] and the iVR system prototype used in our initial pilot study [30]. The VR application consists of 2 phases. First, participants are instructed to show or give compassion to a crying child in a virtual environment. This is then followed by the second stage, in which participants are given the opportunity to assume the role of the crying child and receive compassion. The broad literature on compassion predicts that both giving and receiving compassion should increase self-compassion.

In this study, individualization features were navigated via graphical user interface (GUI) panels added to the virtual environment throughout all stages. However, because qualitative feedback from participants indicated that participants wanted clear instructions about how to use the application, we then added a main menu or tutorial stage at the beginning of our iVR system. Therefore, the application includes three specific stages:

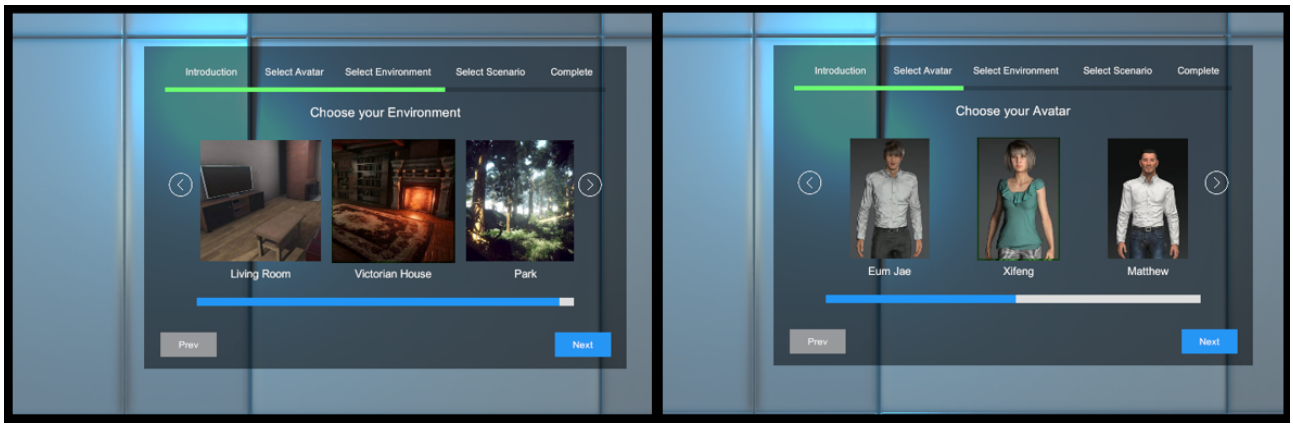
1. The main menu stage, which provides clear instructions about the application and enables participants to customize their experience
2. The delivering compassion stage, which allows participants to practice delivering compassion
3. The receiving compassion stage, which presents the opportunity to experience receiving compassion

Stage 1: Main Menu or Tutorial

In the main menu, participants are presented with a brief explanation of how the VR application works and are guided through the process of selecting their choice of avatar, environment, and scenario. To maintain consistency, all menu options in the main menu follow a template. The title of the menu is displayed at the top of the menu panel. All options are populated under the title banner.

The contents of the menu options vary between individualization categories. For instance, the avatar selection section might contain various virtual body selections, whereas the environment selection section might contain indoor and outdoor terrain selections. Figure 1 is a screenshot of the main menu panel as viewed by the participant. Interactions with the main menu occur via hand-held controllers. When participants have completed customizing their VR experience, they can then proceed to stage 2, where they can practice delivering compassion.

Figure 1. Main menu stage in the individualized virtual reality prototype.



Stage 2: Delivering Compassion

Overview

In stage 2 of the prototype, participants are given the opportunity to deliver compassion to a distressed in-game character. Participants embody a virtual avatar when delivering compassion in a scene generated based on their previous choices. Unlike the predetermined virtual body in the prototype used by Falconer et al [28], the appearance of the body is determined based on the avatar chosen from the main menu stage. The therapy environment and the in-game character's appearance and behavior can also be individualized. Participants were instructed to give compassion to a distressed youth closely following the 3-phase approach of *validation*, *redirection of attention*, and *memory activation*, which is the standard approach recommended for use in CFT when dealing with emotional situations [28].

Validation

The aim of this phase is to acknowledge that the other person is upset, that you do not judge them for this, and that it is perfectly acceptable for them to react in this way. The following is a sample dialogue: "It's not nice when things happen to us that we don't like. It has really upset you, hasn't it?"

Redirection of Attention

The aim of this phase is to direct the other person's attention toward something that is more positive, soothing, and comforting. The following is a sample dialogue: "Sometimes when we are sad it's helpful to think of someone who loves us or is kind to us."

Memory Activation

The aim of this phase is to suggest that the person could try to recall a memory of a person who loves or is kind to them. This memory is supposed to instill more positive feelings of warmth, comfort, and safety. The following is a sample dialogue: "Can you think of someone who loves you or is kind to you? What might they say to you now that would make you feel better?"

Hints and clear instructions were incorporated into the virtual environment to enhance user experience. When the stage is loaded, participants' speech is automatically recorded. After delivering compassion, they can go to the next stage by clicking the "Next Stage" button. Figure 2 depicts a stage-1 scene in which a participant has decided to interact with an upset young female character by embodying an Asian female avatar in an indoor setting.

Figure 2. A screenshot of a participant delivering compassion in the individualized virtual reality prototype.

Stage 3: Receiving Compassion

Stage 3 gives users the opportunity to experience receiving compassion from their chosen virtual self through the eyes of the game character as illustrated in Figure 3. During this stage, participants hear a virtual avatar reenacting the participant's speech that was recorded from the first stage to deliver compassion, but this time, they are on the receiving end of the speech. When the avatar has completed acting the scene, the scenario ends, and they are given the option of returning to the main menu or quitting the application.

Figure 3. A screenshot of participant receiving compassion in the individualized virtual reality prototype.

Individualized Components

Overview

The iVR application gives participants an opportunity to personalize their avatar, environment, and scenario. When used in a therapeutic environment, observing the client's decisions could potentially assist therapists in providing additional clinical insights, particularly for clients who are not willing or able to provide accurate information about their mental health.

Although the customization of avatars in our early pilot study [31] was well received, feedback from participants revealed a preference to interact with avatars that were more graphically realistic. Therefore, improving the appearance and feel of the avatar was one of the primary objectives in building the iVR prototype.

Reallusion software (Reallusion Inc) [37] was used to produce preexisting multilayered humanoid components for this project. The degree of realism and graphic quality of these assets are relatively good. However, these files use single meshes for the skin layer, making it difficult to programmatically alter the avatar's personal attributes, such as clothing and skin tone, based on participant's choice. Owing to this, the prototype's avatar customization was restricted to a selection of premade virtual bodies. As this prototype was intended for use in a New Zealand–based study, avatar options were made available for the region's 3 largest ethnic groups: Asian, European or Pakeha, and Māori [38]. The available choices of virtual bodies are as follows:

1. Asian female
2. Asian male
3. European female
4. European male
5. Māori female
6. Māori male

Virtual Environments

In the early prototype [31], users could choose from a variety of therapeutic environments. However, these environments did not depict the landscape in a visually accurate manner. Participants showed a tendency to favor locations with the most realistic appearance because this enhanced their sense of immersion [31]. Our current iVR prototype design focuses on enhancing the graphical quality of the selected environments to make them more realistic and immersive.

It is important to strike a balance between performance and the complexity of digital resources when developing the virtual environment. High-fidelity resources can produce a more realistic-looking environment but can be detrimental to the performance of the application because it creates a great burden on the software and hardware of the VR HMD to load the digital models. This was one of the difficulties encountered during the development of iVR environments.

To avoid performance issues, the iVR system was developed using photorealistic 3D digital models that were not so complex as to hinder the software's performance. In addition, we used occlusion culling to increase productivity by avoiding the rendering of items that are obscured from view. Figure 4 depicts the iVR prototype's enhanced assets in outdoor and indoor environments. We included both indoor and outdoor environments for the participants to choose from, including a living room, a study room in a Victorian house, and a park.

Figure 4. Improved (A) outdoor virtual environment and (B) indoor virtual environment used in the individualized virtual reality prototype.

Interaction Scenario

In addition to being able to choose their own in-game character, participants could select from two potential scenarios:

1. Delivering compassion to a crying person
2. Delivering compassion to an angry or upset person

Creating a convincingly realistic humanoid model can be difficult because humans are graphically complex objects, and the human eye is adept at identifying subtle humanoid object characteristics [39]. Overall, three distinct types of strategies can be used to animate human behaviors: (1) keyframing, (2) motion capture, and (3) simulation [40], each with specific advantages and limitations [41]. The keyframing technique requires the key positions of objects in each timeline to be manually specified. Using this technique, animators may control the motion of an object's minute elements. It also allows for a more seamless transition between stances. The disadvantage is

that it can be time consuming. However, although the motion capture approach is typically less time consuming than the keyframe method, it does not provide much control over the motion and does not necessarily have seamless transitions between scenarios. The third method is simulation, in which lifelike movements are generated. However, this may be prohibitively expensive and time consuming, making it impractical at this time. After assessing the advantages and disadvantages of each technique, keyframing was chosen to provide great control and smooth scenario transitions.

Design Considerations for VR Application

Immersive VR is fundamentally different from conventional digital products designed for flat 2D screens in that consumers receive a fully immersive 3D experience equivalent to the real world. Consequently, several aspects must be considered when designing components for a VR application.

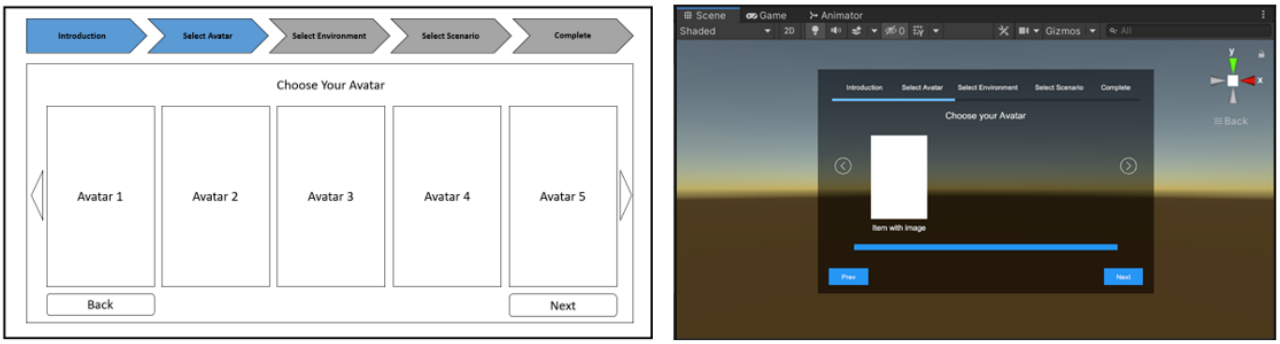
Vision

When designing virtual components, the comfort of the user was the primary concern. According to the recommendations by Meta [42], all menus and displays must be within a comfortable sight distance to avoid eye strain. For most end users, 1 m is the best seeing distance for GUIs and menus [42]. Following this advice, the main menu and GUI panels were rendered approximately 1 m away from the user, as measured in a real-world setting.

Another key factor considered was the end users' line of sight. Szauer [43] recommends that key gameplay items be placed within the immediate field of view of the user, as objects placed outside the user's field of view are commonly overlooked. In the iVR prototype, avatars and virtual bodies were therefore positioned exactly in front of the camera, within its typical range of view.

GUIs and menus must be drawn in front of the user's eyes and positioned relative to the camera, so that they are always within the user's field of view. All menu and user interface information must be displayed without requiring the user to move their head [43]. As we have various customizable settings, it was difficult to display all of them on a single panel. To address this issue, the menu options were divided into sections and subsequently shown using a paginated panel. The next or back button allows users to browse through the menu. Figure 5 illustrates a low-fidelity and high-fidelity prototype of the avatar selection menu panel.

Figure 5. (A) Low-fidelity and (B) high-fidelity prototype of the avatar selection menu panel.



For user input, this project used the Meta (old Oculus) Quest 2 controller input with laser pointers for GUI and menu interactions to provide participants with a more engaging and participatory experience.

Locomotion

In contrast to 2D displays, users can physically move around in a VR world. In VR development, the user's motions within the virtual environment are typically referred to as locomotive. Comfortable locomotion has a direct and meaningful effect on the user experience, making it essential for the success of any VR application [44].

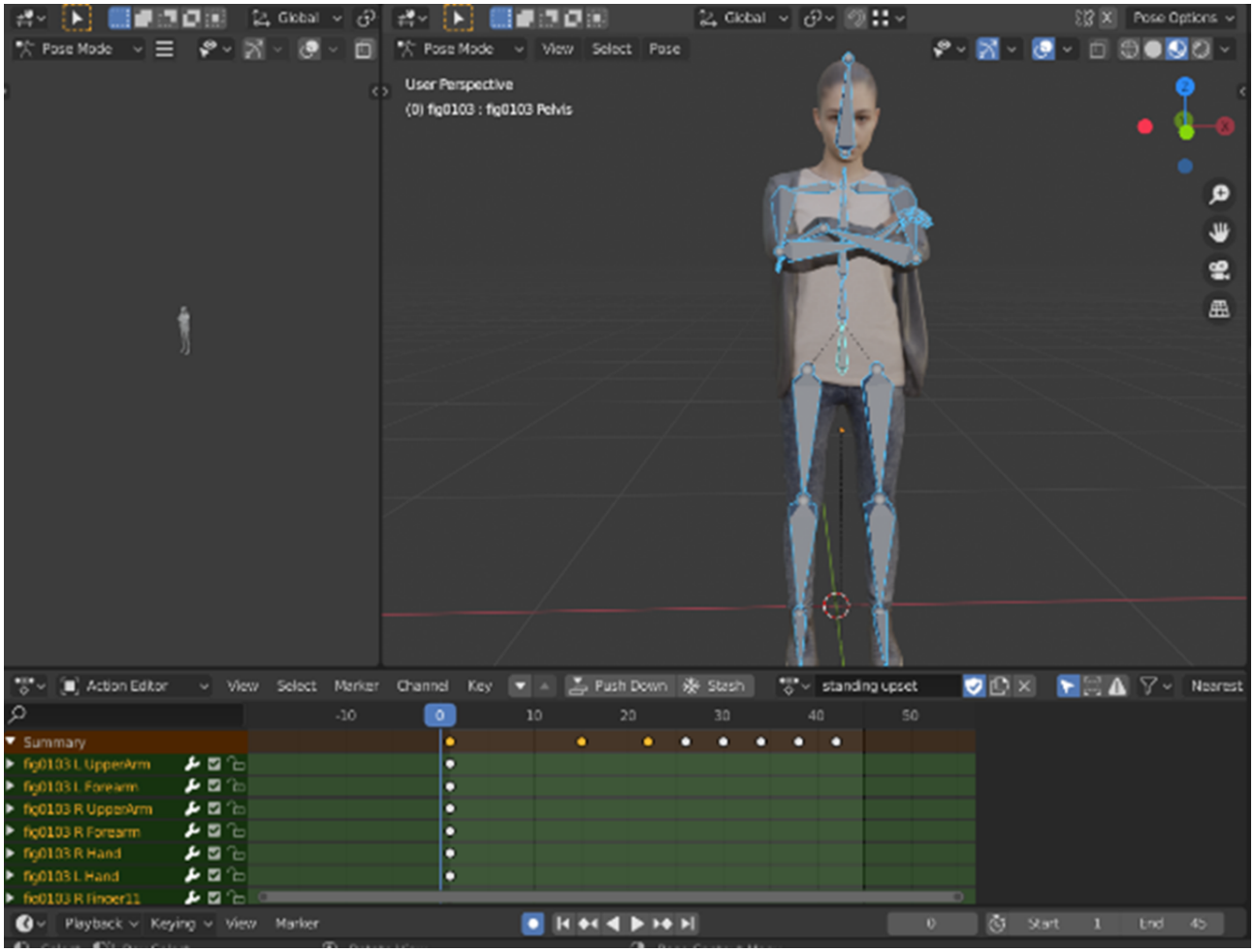
Physical and artificial locomotion are the 2 primary types of locomotion in the creation of VR [45]. Physical locomotion is when motions in the virtual world are synchronized with the user's movements in the real world. Artificial locomotion is when motions in the virtual world are controlled by external factors, such as button presses, as opposed to physical movement. Although this can be advantageous for navigating the virtual environment when physical space is limited, certain artificial actions, such as smooth turning, can be uncomfortable [45]. Therefore, caution should be exercised when developing an artificial locomotive in the virtual world. Teleportation may be used to handle artificial movement in the virtual environment [44], but it was not incorporated in the iVR prototype because participants might not be familiar with the VR HMD, and this functionality was susceptible to user mistakes. Therefore, the iVR for mental health application was designed to allow users to navigate the virtual world by walking in the real world.

Development Tools

The Unity [46] game engine was used in the development of the iVR prototype because it offers an intuitive user interface with a short learning curve. In addition, Unity provides flexible VR plugins that are compatible with Meta Quest 2.

Virtual avatars in the iVR prototype were animated using Mixamo [47] and Blender [48]. First, Mixamo was used to produce generic movement patterns for the models. Upon completion of the animation process in Mixamo, the file was converted into a FilmBoX file, so that it may be used in the Unity project (as depicted in Figure 6). Blender was then used to make any necessary animation adjustments, and the file was then imported into Unity. The animation transitions and position orders were then defined in Unity using animation controllers.

Figure 6. Keyframing animation process in Blender.



The iVR prototype records the audio input of the Meta Quest 2 headset during stage 1 and uses the recording as the audio input for stage 2. Participants were able to hear and observe their avatar reenacting stage 1 from a different perspective. To control the facial expressions of the avatar, Oculus VR Lipsync (Meta) was used. Oculus VR Lipsync [49] is a Unity add-on plugin that may be used to synchronize the lip and facial movement of a digital object with audio input.

Methods

The goal of the study was to examine whether iVR increases self-compassion and decreases symptoms of depression.

Participants

A total of 36 participants agreed to participate (n=17, 47% self-identified as men and n=19, 53% self-identified as women), who were on average aged 29.3 (SD 11.64) years. Participants were recruited via poster advertising, social media advertising, and word of mouth. Participants were informed that it was a study about depression, but there was no requirement for participants to have a depression diagnosis and that they did not need to disclose whether they were clinically depressed. Only participants aged >18 years were allowed to participate in the study because it dealt significantly with sensitive topics related to mental health. Participants were also required to reside in Auckland, New Zealand, because the experiment needed to be conducted in person at a specific location. Potential participants who expressed interest through email were sent a copy of the information sheet, and they provided written informed consent before completing the experimental sessions. All participants were reimbursed with a gift card worth NZD $20 (US $11.8) upon completion of the study.

Measures

Self-Compassion Scale

The Self-Compassion Scale (SCS) [50] was used to assess self-compassion. SCS is a tool created by Kristin Neff [51] to measure self-compassion. It is frequently used in academic settings [50]. SCS is a 26-item questionnaire that has previously been shown to be a reliable and valid indicator of self-compassion [52]. High scores on SCS are indicative of great self-compassion. SCS data were collected from participants 3 times throughout the study: at the beginning (baseline preintervention assessment), at the end of the first session (after interaction with iVR), and at the end of the follow-up session (after interaction with iVR). To determine whether the iVR session had any impact on participants' levels of self-compassion, the postintervention SCS scores were compared with the baseline score.

Patient Health Questionnaire

The 9-item Patient Health Questionnaire (PHQ-9) [53,54] and 8-item Patient Health Questionnaire (PHQ-8, which is an 8-item

subset of PHQ-9) [53], have been widely used in both research settings and therapeutic trials [54]. PHQ-9 is a short self-report questionnaire that combines The Diagnostic and Statistical Manual of Mental Disorders depression diagnostic criteria [55] with additional prevalent major depressive symptoms. High scores on PHQ-9 or PHQ-8 are indicative of high depressive symptoms. PHQ-8 omits the question about suicidal and self-harm ideas that is occasionally used to gauge the risk of suicide from PHQ-9 [54]. For the purpose of major depressive disorder screening, PHQ-8 was found to be just as helpful as PHQ-9 because it is a sensitive indicator of depression symptomatology [56]. As this study did not specifically assess the risk of suicide, depression symptoms were measured using the PHQ-8 scale. PHQ-8 scores were collected at the beginning of the trial (preintervention baseline), at the end of the first session (after interaction with iVR), and at the end of the follow-up session (after interaction with iVR). The PHQ-8 scores after the intervention were then compared with the baseline score to see if the iVR session had any effect on participants' depression symptoms.

System Usability Scale

The System Usability Scale (SUS) [57] is a 10-item, standardized questionnaire that was used to assess the perceived usability of VR systems by end users [57]. Previous studies have shown it is a valid or reliable indicator of VR system usability [58]. It has proven to be a robust tool, having been used to evaluate a wide range of interfaces, including websites, VR applications, GUI, and television user interfaces [57]. High scores on this measure are indicative of better usability.

User Experience Questionnaire

The User Experience Questionnaire (UEQ) [59] is a 26-item questionnaire that was used to evaluate the acceptability of the VR application. UEQ provides an index of a product's pragmatic and hedonic quality aspects by considering 6 scales: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty [59].

Qualitative Feedback

Information about participants' subjective preferences (what they liked and disliked about the iVR experience) and any potential recommendations they might have for improvement was also obtained upon the conclusion of the follow-up session. The following questions were asked: (1) What were the top three things that they liked about iVR? (2) What didn't they like about iVR? and (3) How do you think the next version can be improved? What other features would you like to see? Participants' responses were transcribed for qualitative analysis, using the thematic analysis method [60].

Procedures

Overview

We conducted 2 one-on-one sessions for each participant, completed a minimum of 2 weeks apart. The first session took approximately 40 to 45 minutes, whereas session 2 took approximately 20 to 30 minutes to complete. After providing informed consent, in session 1, participants were first asked to complete the questionnaires about depression and self-compassion (PHQ-8 and SCS). After completing the surveys, individuals were given the opportunity to become comfortable with the experimental setup and the VR technology. If they were unfamiliar with VR HMDs, participants were asked to watch a video example of the experimental process and received basic instructions about how to use the Meta Quest 2 VR HMD. In addition, they received an instruction sheet that explained the idea of CFT and the pertinent steps (validation, attentional redirection, and memory activation). They were then asked to interact with the iVR application for approximately 6 to 10 minutes. The participants interacted with the iVR application in 3 stages as previously outlined in the iVR Design and Implementation section.

Stage 1: Main Menu or Tutorial (Approximately 2 min)

This stage involves the following:

1. Interaction with GUI panels to choose the virtual avatar, environment, and scenario.

Stage 2 Delivering Compassion (Approximately 2 min)

This stage involves the following:

1. Visuomotor synchrony or embodiment with the chosen avatar through movement
2. Interaction with crying, angry child, or youth
3. Instruction: "React to the child/youth by using the sentences you have learned just now."

Stage 3 Receiving Compassion (Approximately 2 min)

This stage involves the following:

1. Visuomotor synchrony or embodiment with crying, angry child, or youth through movement
2. Real-time playback of recorded interaction part B from the perspective of crying, angry child, or youth
3. Instruction: "Stand/sit, look and listen."

Participants were allowed to explore the system without restriction, with researchers present in the room to assist if necessary. After the interaction, they were asked to complete the same measures of depression and self-compassion again, in addition to measures of their experience of engaging with the VR system (SUS and UEQ). In session 2, participants commenced the session by interacting with the iVR application and then were asked to complete the measures of depression and self-compassion and provide qualitative feedback about the iVR application.

Ethics Approval

The study was approved by Massey University's Human Research Ethics Committee (NOR 21/83).

Analyses

Although a total of 36 participants completed session 1, owing to COVID-19 constraints only 23 (64%) participants were able to complete session 2. A post hoc power analysis conducted using G*Power revealed that for the key comparisons of interest (whether there was any significant change in self-compassion and depression upon completion of session 1 relative to baseline and upon completion of session 2 relative to baseline), our within-participants design with 23 participants still had sufficient

power (>0.80) to detect a moderate to large effect difference (Cohen *d*=0.65) at Cronbach α=.05 (2 tailed), for each of these comparisons. Owing to the dependent nature of the samples, paired-samples 2-tailed *t* tests were used for these key contrasts.

Results

Mental Health Data Following the iVR Intervention

Descriptive statistics for both SCS and PHQ-8 are reported in Table 1.

Table 1. Descriptive statistics for the key measures of mental health (N=36).

Measure	Baseline		Session 1 (postintervention assessment)		Session 2 (postintervention assessment)	
	Score, mean (SD)	Participants, n (%)	Score, mean (SD)	Participants, n (%)	Score, mean (SD)	Participants, n (%)
SCS[a]	2.94 (0.68)	35 (97)	3.07 (0.73)	35 (97)	3.23 (0.80)	23 (64)
PHQ-8[b]	8.52 (6.32)	23 (64)	7.44 (6.15)	23 (64)	6.70 (4.97)	23 (64)

[a]SCS: Self-Compassion Scale.

[b]PHQ-8: 8-item Patient Health Questionnaire.

For both SCS and PHQ-8, skewness fell within an acceptable range to permit parametric data analysis [61]. A series of paired-sample 2-tailed *t* tests were therefore used to analyze the changes in SCS and PHQ-8 scores. Regarding SCS, these analyses revealed that, relative to baseline, there was a significant increase in participants' self-compassion, both at the end of session 1 (t_{34}=2.64; *P*=.01) and at the end of session 2 (t_{22}=2.37; *P*=.03). Regarding PHQ-8, the same analyses revealed that, relative to baseline, self-rated depression was overall unchanged at the end of session 1 (t_{35}=1.69; *P*=.10) but that there was a nonsignificant trend for depressive symptoms to be low at the end of session 2 (t_{22}=1.88; *P*=.07).

Associations Between Self-Compassion and Depressive Symptoms

The next step in the analyses was to calculate Pearson correlations [62] to test how participants' self-compassion and depressive symptoms related to one another across the different stages of the study. These correlations are reported in Table 2. There were significant negative associations between SCS and PHQ-8 scores at all time points assessed, and the magnitude of these associations was consistently large [63].

Table 2. Correlation analysis of the 8-item Patient Health Questionnaire (PHQ-8) and Self-Compassion Scale (SCS) scores.

Variables	SCS (baseline)	SCS (session 1)	SCS (session 2)	PHQ-8 (baseline)	PHQ-8 (session 1)	PHQ-8 (session 2)
SCS (baseline)						
r	1	0.949	0.747	–0.743	–0.699	–0.76
P value	—[a]	<.001	<.001	<.001	<.001	<.001
SCS (session 1)						
r	0.949	1	0.775	–0.741	–0.694	–0.727
P value	<.001	—	<.001	<.001	<.001	<.001
SCS (session 2)						
r	0.747	0.775	1	–0.581	–0.531	–0.735
P value	<.001	<.001	—	.004	.009	<.001
PHQ-8 (baseline)						
r	–0.743	–0.741	–0.581	1	0.955	0.685
P value	<.001	<.001	.004	—	<.001	<.001
PHQ-8 (session 1)						
r	–0.699	–0.694	–0.531	0.955	1	0.64
P value	<.001	<.001	.009	<.001	—	.001
PHQ-8 (session 2)						
r	–0.76	–0.727	–0.735	0.685	0.64	1
P value	<.001	<.001	<.001	<.001	.001	—

[a]Not applicable.

iVR System Usability and Acceptability

Data collected from 36 participants upon completion of session 1 revealed that the overall SUS score for the iVR application was 75.9 (SD 13.09). According to the adjective rating system defined by Bangor et al [57], this falls within the good to excellent range.

UEQ data were also gathered from 36 participants at the end of session 1. Figure 7 presents the ratings for the 6 UEQ scales.

Figure 8 shows the mean of the UEQ scales grouped by attractiveness, pragmatic quality (perspicuity, efficiency, and dependability), and hedonic quality (stimulation and originality).

The UEQ analysis tool [59] was used to compare the UEQ results with a benchmark data set containing data from 20,190 individuals from 452 studies of different products. Figure 9 presents the results of the comparison.

Figure 7. User Experience Questionnaire results.

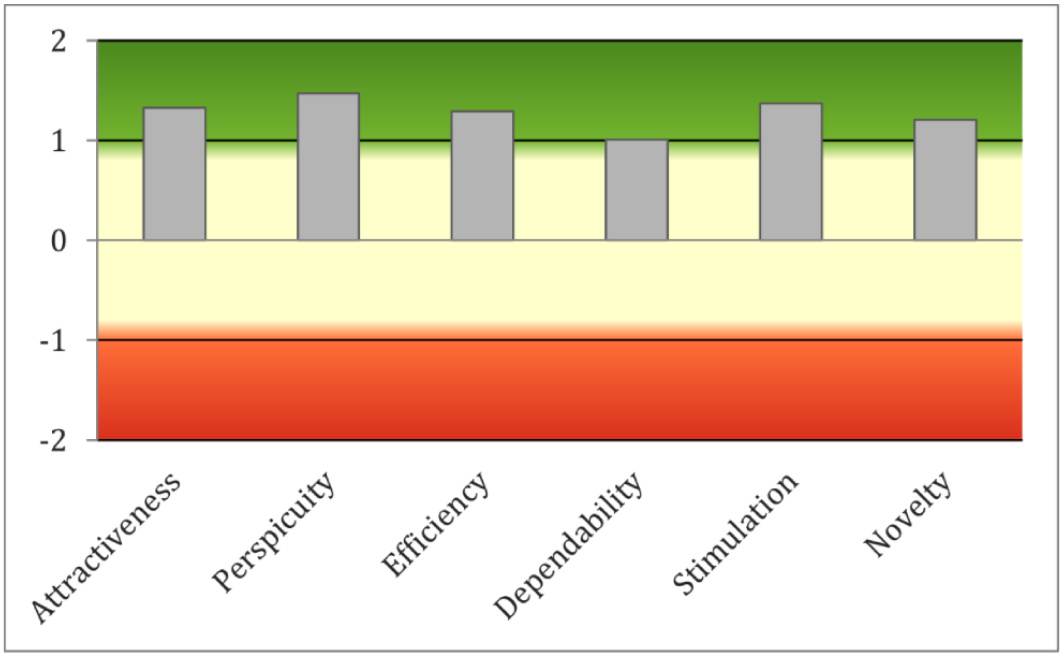


Figure 8. User Experience Questionnaire results grouped by attractiveness, pragmatic quality, and hedonic quality.

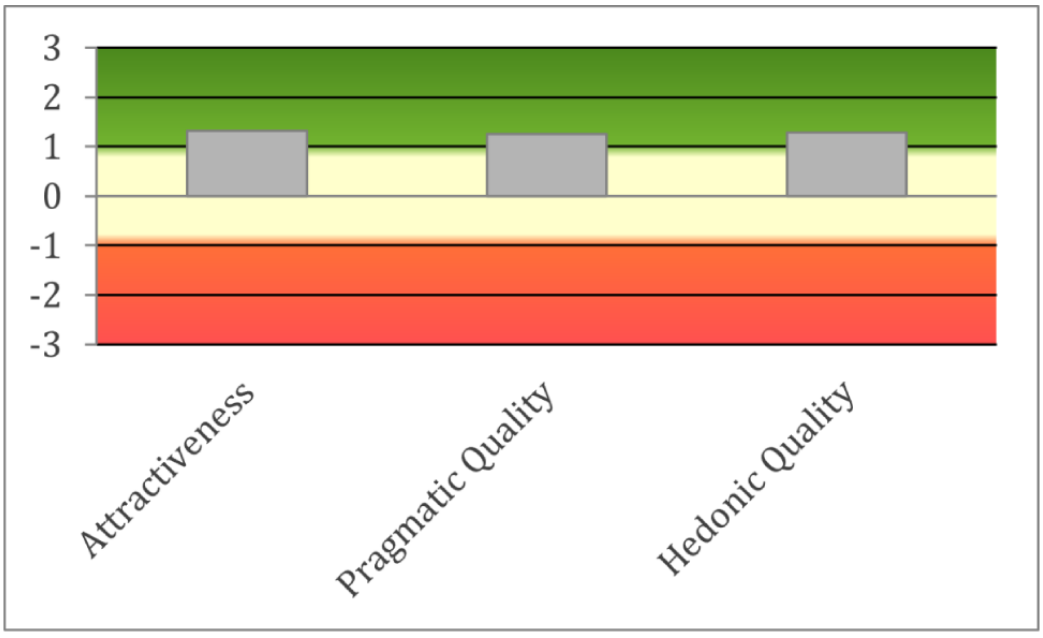
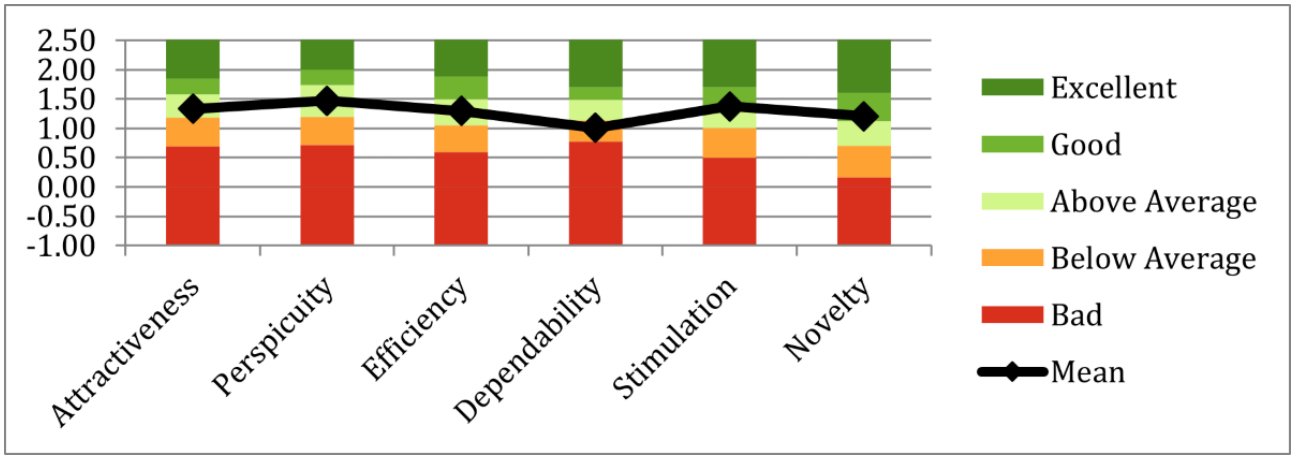


Figure 9. User Experience Questionnaire results in comparison with benchmark data.



Relative to the benchmark data, the iVR prototype's UEQ scores for attractiveness and efficiency lie within the "Above Average" range, whereas scores for perspicuity, novelty, and stimulation fall within the "Good" range. However, dependability falls within the "Below Average" area of UEQ measures.

Finally, qualitative feedback was gathered from 64% (23/36) of the participants at the end of session 2. Thematic analysis was used to interpret these responses [60]. This revealed that participants' responses to the question, "What were the top three things that they liked about iVR?" could be categorized according to the following themes:

1. The top element that participants liked was that it is user-friendly. Approximately half of the participants (17/36, 47%) liked how the app was "easy to use." Other quotes included the following: "easy instructions," "easy navigation," and "easy to use and has a positive vibe."
2. Participants liked how they could customize their VR experience. Quotes from participants included the following: "choose personal avatar," "I enjoyed customizing my situation," "how you got to choose your person and environment," and "I liked that you can pick different areas."
3. Participants liked how the experience felt real and immersive. Quotes included the following: "The immersive aspect really added an element of connection," "It feels like a person is there with you," and "Illusion of real experience."
4. Participants liked the use of VR technology. Quotes included the following: "I like virtual reality," "VR is very immersive," "suitable for VR because other systems are just a single screen," and "I just enjoy VR."
5. Some participants liked the virtual environment. Quotes included the following: "calming environment," "The background was calming," and "loved the scenery."
6. Some participants liked the user interface design. Quotes included the following: "Clean UI" and "I liked that the menu was simple."

For the question, "What didn't they like about iVR?" answers from participants could be grouped into the following:

1. Participants disliked the limited interaction with the virtual avatars. Quotes included the following: "Maybe there can get more operation to do," "Very dull, nothing really happened," and "a bit more interaction and maybe some more prompts."
2. Some participants found that the system was a little slow. Quotes included the following: "Laggy, sometimes unresponsive," "has some lags," "I found the VR menu was a bit slow," and "A little laggy going from one phase to the next."
3. Some participants felt that the scenarios were very simple. Quotes included the following: "The scenario could be more complex" and "I wish it had more scenarios."

The responses to the question, "How do you think the next version can be improved? What other features would you like to see?" can be categorized as follows:

1. Participants suggested adding more robust scenarios to choose from. Quotes included the following: "You could create different scenarios and paths for users in the next version," "Maybe a more interesting scenario, like an argument," and "Perhaps expanding on the first stage to include more situations."
2. Participants would also like to see more interactions with the virtual avatars and virtual environment. Quotes included the following: "more interaction with the environment," "More interaction, such as we can walk together, have a seat," "Better user interaction," and "The AI reacting to the user."
3. Some participants wanted more avatar options to be added. Quotes included the following: "more avatars and settings" and "More diverse ethnic avatars."
4. Some participants wanted to see improved graphics and a more natural avatar. Quotes included the following: "I'd like if the character models moved a bit more natural."
5. Some participants said that they wanted to see improved performance of iVR. Quotes included the following: "Improve load speed and fps."

Discussion

Principal Findings

The findings of this study provide important novel insights into the potential value of iVR as a tool to increase participants' self-compassion and reduce their depressive symptoms and

about the usability and acceptance of the iVR system in general. Although these findings are preliminary, it is hoped that they will help to motivate future studies in this area.

Changes in Self-Compassion Level and Depressive Symptoms

The most noteworthy finding was the significant increase in participants' self-compassion after interacting with the iVR prototype. There was a significant increase in participants' self-compassion after as little as a single session of iVR. These findings suggest that iVR could potentially be a helpful tool to improve self-compassion. As noted previously, the iVR session was designed to explicitly incorporate elements of CFT by encouraging participants to engage in compassionate behavior and compassion-based skills. Therefore, this study provides further support to the broad literature showing that self-compassion can be increased by training [7] and extends it meaningfully by showing that these benefits also emerge when this training is provided in the virtual world.

Regarding depression, participants' depressive symptoms were unchanged at the end of session 1 relative to baseline, but there was a nonsignificant trend for their symptoms to be reduced at the end of session 2. We consider these data to be encouraging, particularly considering the nonclinical nature of the cohort. This is because many participants reported relatively low levels of depressive symptoms at baseline. Thus, an important area of future studies is to use this novel iVR approach with participants who present with clinically significant levels of depressive symptomatology

Another important finding to emerge was that depressive symptoms and self-compassion were significantly correlated with one another at all stages of the study, and the magnitude of these associations was consistently large [63]. Although the design of this study clearly precludes inferences about causality, increased self-compassion should theoretically help decrease depression symptoms. Therefore, the finding of large, robust associations between measures of these 2 constructs broadly align with the broad literature that has established self-compassion as a protective factor against depression [5].

Usability and Acceptability of the iVR Prototype

In addition to showing that the iVR application was associated with significantly enhancing self-compassion, participants also generally found it to be highly usable. The mean SUS score for the iVR application was 75.9, which falls within the good to excellent range [57], indicating that participants mostly perceived that the system has good usability.

Comparing the UEQ ratings with the values in the benchmark data set was also valuable in highlighting the iVR prototype's specific strengths and weaknesses. Novelty and stimulation were rated as "Good" on UEQ, outperforming 75% of the systems within the benchmark. This suggests that participants viewed the iVR prototype as innovative and captivating. However, the dependability scale of UEQ was graded as "Below Average," indicating that participants may not experience a sense of control when interacting with the iVR system. Moreover, some participants believed the system to be slow. This suggests that there may be performance issues with the iVR prototype, and future iterations of iVR should optimize system performance and speed to improve the user experience.

Limitations and Future Directions

Despite the encouraging results of this study, there are several limitations that create opportunities for future studies. First, COVID-19–related travel restrictions and people testing positive for COVID-19 meant that many participants were unable to attend their follow-up appointments, and this made an RCT unfeasible.

Although our reduced sample size in session 2 still had adequate power to detect moderate to large within-group differences, obviously, such a small cohort is unlikely to be representative of the broad population. The absence of a control condition also means that we cannot rule out the possibility that the positive effects identified were related to nonspecific factors other than the iVR sessions. Such factors include expectation effects, natural resolution of depressive symptoms over time, and positive interactions with the research team. COVID-19 lockdowns and restrictions could also be a major influence on depressive symptoms, as shown in studies conducted during the pandemic [64,65]. Therefore, to cross-validate our results, the next important step is a large-scale RCT that includes participants presenting with clinically significant levels of depressive symptoms. It would also be very valuable to include a long-term follow-up to establish how long any observed gains persist.

Owing to the small sample size, this study was also underpowered to assess the potential influence of variables such as gender, culture, and ethnicity. Overt features, such as gender and ethnic and cultural disparities [33], can have a substantial impact on VR interventions and contribute to variance in results, even when the same VR simulation treatment is used. Therefore, consideration of such variables in iVR designs represents another important area for future studies.

Another limitation of our design was that we did not investigate VR variables, such as presence and embodiment. Future studies of iVR may benefit from considering these variables, as presence in the virtual environment and sense of embodiment toward the avatar may moderate the effects of VR interventions. For instance, according to Grassini et al [66], presence may influence training outcomes. It is noteworthy that, in this study, participants' qualitative data indicated that they enjoyed how immersive the VR environment was and how it made them feel as if they were there in person. It remains to be established whether presence was an important determinant of the efficacy of the intervention.

Comparison of the results of qualitative participant feedback about their recommendations for future iterations of iVR with those of previous studies [31] also revealed a few recurring themes that might be explored for future iterations of the iVR. For instance, more robust avatar interaction and additional avatar options should be added to potentially enhance the user experience. In a future study with a large sample size, it would also be interesting to examine the link between usability scores and depressive symptoms to establish if those with severe

depression symptoms have a different user experience from those with few symptoms.

Finally, it is also worth noting that the system only supported limited individualization. It would be of considerable value to test the impact of more advanced individualization (such as changing skin tone, hair color, and style) on participants' self-compassion and depressive symptoms. We were also unable to capture physiological cues to record participants' stress level or emotional states, and this also remains as an important avenue for future studies to build on the current findings.

Conclusions

In this paper, we have outlined the design and implementation of an iVR application to enhance self-compassion and improve symptoms of depression. The pilot user study revealed that, relative to baseline, self-compassion increased after both a single and a second iVR session and that there was a nonsignificant trend for depressive symptoms to decrease after a second iVR session. Importantly, most users thought that the product's usability was good or excellent.

Moving forward, we are evaluating potential improvements to the iVR application that would involve integrating more robust participant–virtual environment interaction, with emphasis on interactions with the avatars that are already present in the environment. We are considering including some interactive elements to boost user engagement. Adding artificial intelligence to the avatar is another option that is being considered. Our hope is that this study has the potential to pave the path for future, cost-effective, clinical applications and large-scale efficacy assessment of intelligent, iVR technology that might deliver mental health therapy in a safe, cost-effective, and engaging manner.

Conflicts of Interest

RP declares the use of computer software provided at no cost by SBT-Pro for research purposes and support from Servier and Lundbeck for travel to educational meetings.

References

1. Depressive disorder (depression). World Health Organization. 2023 Mar 31. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2023-07-25]
2. Wang PS, Aguilar-Gaxiola S, Alonso J, Angermeyer MC, Borges G, Bromet EJ, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 2007 Sep 08;370(9590):841-850 [FREE Full text] [doi: 10.1016/S0140-6736(07)61414-7] [Medline: 17826169]
3. López A, Sanderman R, Schroevers MJ. A close examination of the relationship between self-compassion and depressive symptoms. Mindfulness (N Y) 2018;9(5):1470-1478 [FREE Full text] [doi: 10.1007/s12671-018-0891-6] [Medline: 30294388]
4. Zhang H, Carr ER, Garcia-Williams AG, Siegelman AE, Berke D, Niles-Carnes LV, et al. Shame and depressive symptoms: self-compassion and contingent self-worth as mediators? J Clin Psychol Med Settings 2018 Dec;25(4):408-419. [doi: 10.1007/s10880-018-9548-9] [Medline: 29488038]
5. Ehret AM, Joormann J, Berking M. Examining risk and resilience factors for depression: the role of self-criticism and self-compassion. Cogn Emot 2015;29(8):1496-1504. [doi: 10.1080/02699931.2014.992394] [Medline: 25517734]
6. Körner A, Coroiu A, Copeland L, Gomez-Garibello C, Albani C, Zenger M, et al. The role of self-compassion in buffering symptoms of depression in the general population. PLoS One 2015 Oct 2;10(10):e0136598 [FREE Full text] [doi: 10.1371/journal.pone.0136598] [Medline: 26430893]
7. Jazaieri H, Jinpa GT, McGonigal K, Rosenberg EL, Finkelstein J, Simon-Thomas E, et al. Enhancing compassion: a randomized controlled trial of a compassion cultivation training program. J Happiness Stud 2013;14(4):1113-1126 [FREE Full text] [doi: 10.1007/s10902-012-9373-z]
8. Leaviss J, Uttley L. Psychotherapeutic benefits of compassion-focused therapy: an early systematic review. Psychol Med 2015 Apr;45(5):927-945 [FREE Full text] [doi: 10.1017/S0033291714002141] [Medline: 25215860]
9. Falconer CJ, King JA, Brewin CR. Demonstrating mood repair with a situation-based measure of self-compassion and self-criticism. Psychol Psychother 2015 Dec;88(4):351-365 [FREE Full text] [doi: 10.1111/papt.12056] [Medline: 25663161]
10. Tao G, Garrett B, Taverner T, Cordingley E, Sun C. Immersive virtual reality health games: a narrative review of game design. J Neuroeng Rehabil 2021 Feb 11;18(1):31 [FREE Full text] [doi: 10.1186/s12984-020-00801-3] [Medline: 33573684]
11. Xu W, Liang HN, Yu K, Wen S, Baghaei N, Tu H. Acceptance of virtual reality exergames among Chinese older adults. Int J Hum Comput Interact 2023;39(5):1134-1148 [FREE Full text] [doi: 10.1080/10447318.2022.2098559]
12. Meyerbröker K, Morina N, Emmelkamp PM. Enhancement of exposure therapy in participants with specific phobia: a randomized controlled trial comparing yohimbine, propranolol and placebo. J Anxiety Disord 2018 Jun;57:48-56. [doi: 10.1016/j.janxdis.2018.05.001] [Medline: 29804894]
13. Xu W, Liang HN, Baghaei N, Ma X, Yu K, Meng X, et al. Effects of an immersive virtual reality exergame on university students' anxiety, depression, and perceived stress: pilot feasibility and usability study. JMIR Serious Games 2021 Nov 22;9(4):e29330 [FREE Full text] [doi: 10.2196/29330] [Medline: 34813487]

14. Baghaei N, Chitale V, Hlasnik A, Stemmet L, Liang HN, Porter R. Virtual reality for supporting the treatment of depression and anxiety: scoping review. JMIR Ment Health 2021 Sep 23;8(9):e29681 [FREE Full text] [doi: 10.2196/29681] [Medline: 34554097]
15. Morina N, Ijntema H, Meyerbröker K, Emmelkamp PM. Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. Behav Res Ther 2015 Nov;74:18-24. [doi: 10.1016/j.brat.2015.08.010] [Medline: 26355646]
16. Rothbaum BO, Hodges L, Smith S, Lee JH, Price L. A controlled study of virtual reality exposure therapy for the fear of flying. J Consult Clin Psychol 2000 Dec;68(6):1020-1026. [doi: 10.1037//0022-006x.68.6.1020] [Medline: 11142535]
17. Riva G. Applications of virtual environments in medicine. Methods Inf Med 2003;42(5):524-534. [Medline: 14654887]
18. Krijn M, Emmelkamp PM, Olafsson RP, Schuemie MJ, van der Mast CA. Do self-statements enhance the effectiveness of virtual reality exposure therapy? A comparative evaluation in acrophobia. Cyberpsychol Behav 2007 Jun;10(3):362-370. [doi: 10.1089/cpb.2006.9943] [Medline: 17594260]
19. Donker T, Cornelisz I, van Klaveren C, van Straten A, Carlbring P, Cuijpers P, et al. Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized clinical trial. JAMA Psychiatry 2019 Jul 01;76(7):682-690 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0219] [Medline: 30892564]
20. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
21. Emmelkamp PM, Meyerbröker K. Virtual reality therapy in mental health. Annu Rev Clin Psychol 2021 May 07;17:495-519. [doi: 10.1146/annurev-clinpsy-081219-115923] [Medline: 33606946]
22. Rizzo A, Shilling R. Clinical virtual reality tools to advance the prevention, assessment, and treatment of PTSD. Eur J Psychotraumatol 2017 Jan 16;8(sup5):1414560 [FREE Full text] [doi: 10.1080/20008198.2017.1414560] [Medline: 29372007]
23. Reger GM, Koenen-Woods P, Zetocha K, Smolenski DJ, Holloway KM, Rothbaum BO, et al. Randomized controlled trial of prolonged exposure using imaginal exposure vs. virtual reality exposure in active duty soldiers with deployment-related posttraumatic stress disorder (PTSD). J Consult Clin Psychol 2016 Nov;84(11):946-959. [doi: 10.1037/ccp0000134] [Medline: 27606699]
24. Norr AM, Smolenski DJ, Reger GM. Effects of prolonged exposure and virtual reality exposure on suicidal ideation in active duty soldiers: an examination of potential mechanisms. J Psychiatr Res 2018 Aug;103:69-74. [doi: 10.1016/j.jpsychires.2018.05.009] [Medline: 29783077]
25. Srivastava K, Das RC, Chaudhury S. Virtual reality applications in mental health: challenges and perspectives. Ind Psychiatry J 2014 Jul;23(2):83-85 [FREE Full text] [doi: 10.4103/0972-6748.151666] [Medline: 25788795]
26. Emmelkamp PM, Vedel E. Evidence-Based Treatments for Alcohol and Drug Abuse: A Practitioner's Guide to Theory, Methods, and Practice. New York, NY: Routledge; 2006.
27. Ghiţă A, Gutiérrez-Maldonado J. Applications of virtual reality in individuals with alcohol misuse: a systematic review. Addict Behav 2018 Jun;81:1-11. [doi: 10.1016/j.addbeh.2018.01.036] [Medline: 29421343]
28. Falconer CJ, Rovira A, King JA, Gilbert P, Antley A, Fearon P, et al. Embodying self-compassion within virtual reality and its effects on patients with depression. BJPsych Open 2016 Feb 15;2(1):74-80 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002147] [Medline: 27703757]
29. Li H, Dong W, Wang Z, Chen N, Wu J, Wang G, et al. Effect of a virtual reality-based restorative environment on the emotional and cognitive recovery of individuals with mild-to-moderate anxiety and depression. Int J Environ Res Public Health 2021 Aug 27;18(17):9053 [FREE Full text] [doi: 10.3390/ijerph18179053] [Medline: 34501643]
30. Baghaei N, Stemmet L, Hlasnik A. Time to get personal: individualised virtual reality for mental health. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI EA '20; April 25-30, 2020; Honolulu, HI p. 1-9 URL: https://dl.acm.org/doi/abs/10.1145/3334480.3382932 [doi: 10.1145/3334480.3382932]
31. Baghaei N, Stemmet L, Khaliq I, Ahmadi A, Halim I, Liang H, et al. Designing individualised virtual reality applications for supporting depression: a feasibility study. In: Proceedings of the 2021 ACM SIGCHI Symposium on Engineering Interactive Computing Systems. 2021 Presented at: EICS '21; June 8-11, 2021; Virtual Event p. 6-11 URL: https://dl.acm.org/doi/10.1145/3459926.3464761 [doi: 10.1145/3459926.3464761]
32. Wilson CJ, Soranzo A. The use of virtual reality in psychology: a case study in visual perception. Comput Math Methods Med 2015;2015:151702 [FREE Full text] [doi: 10.1155/2015/151702] [Medline: 26339281]
33. Kritikos J, Alevizopoulos G, Koutsouris D. Personalized virtual reality human-computer interaction for psychiatric and neurological illnesses: a dynamically adaptive virtual reality environment that changes according to real-time feedback from electrophysiological signal responses. Front Hum Neurosci 2021 Feb 12;15:596980 [FREE Full text] [doi: 10.3389/fnhum.2021.596980] [Medline: 33643010]
34. McArthur V. The UX of avatar customization. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO p. 5029-5033 URL: https://dl.acm.org/doi/10.1145/3025453.3026020 [doi: 10.1145/3025453.3026020]

35. Kao D, Ratan R, Mousas C, Joshi A, Melcer EF. Audio matters too: how audial avatar customization enhances visual avatar customization. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA p. 1-27 URL: https://dl.acm.org/doi/abs/10.1145/3491102.3501848 [doi: 10.1145/3491102.3501848]
36. Waltemate T, Gall D, Roth D, Botsch M, Latoschik ME. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. IEEE Trans Vis Comput Graph 2018 Apr;24(4):1643-1652. [doi: 10.1109/TVCG.2018.2794629] [Medline: 29543180]
37. Reallusion. Reallusion Inc. URL: https://www.reallusion.com/iclone/?gclid=Cj0KCQjw94WZBhDtARIsAKxWG-9XZdXRHLE7fIPhDUKmzfr881P38ySxqqgYSx__TqraHY4hURDVTIIaAvktEALw_wcB [accessed 2022-09-01]
38. Ethnic group summaries reveal New Zealand's multicultural make-up. Stats NZ. 2020 Sep 03. URL: https://www.stats.govt.nz/news/ethnic-group-summaries-reveal-new-zealands-multicultural-make-up/ [accessed 2022-06-01]
39. Thalmann D, Shen J, Chauvineau E. Fast realistic human body deformations for animation and VR applications. In: Proceedings of the 1996 Computer Graphics International. 1996 Presented at: CGI '96; June 24-28, 1996; Pohang, South Korea p. 166-174 URL: https://ieeexplore.ieee.org/document/511798 [doi: 10.1109/cgi.1996.511798]
40. O'Brien JF, Hodgins JK. Animating fracture. Commun ACM 2000 Jul;43(7):68-75 [FREE Full text] [doi: 10.1145/341852.341865]
41. Vosinakis S, Panayiotopoulos T. A tool for constructing 3D environments with virtual agents. Multimed Tools Appl 2005 Feb;25(2):253-279 [FREE Full text] [doi: 10.1007/s11042-005-5607-y]
42. Vision. Meta Quest. URL: https://developer.oculus.com/learn/bp-vision/ [accessed 2021-12-01]
43. Szauer G. Unit 6 - Best UI practices for VR. Unity Learn. URL: https://learn.unity.com/tutorial/unit-6-best-ui-practices-for-vr?uv=2018.4&courseId=5d955b5dedbc2a319caab9a0#5d969338edbc2a2d9fcf7792 [accessed 2020-11-01]
44. Bozgeyikli E, Raij A, Katkoori S, Dubey R. Locomotion in virtual reality for room scale tracked areas. Int J Hum Comput Stud 2019 Feb;122:38-49 [FREE Full text] [doi: 10.1016/j.ijhcs.2018.08.002]
45. Locomotive. Meta Quest. URL: https://developer.oculus.com/learn/bp-locomotion/ [accessed 2021-08-01]
46. Home page. Unity. URL: https://unity.com/ [accessed 2022-09-01]
47. Mixamo. Adobe Systems Incorporated. URL: https://www.mixamo.com/#/ [accessed 2022-09-01]
48. Blender community. Blender. URL: https://www.blender.org/ [accessed 2022-09-01]
49. Oculus lipsync for Unity development. Meta Quest. URL: https://developer.oculus.com/documentation/unity/audio-ovrlipsync-unity/ [accessed 2022-09-01]
50. Muris P, Otgaar H. The process of science: a critical evaluation of more than 15 years of research on self-compassion with the self-compassion scale. Mindfulness 2020 Apr 07;11(6):1469-1482 [FREE Full text] [doi: 10.1007/s12671-020-01363-0]
51. Neff KD. The development and validation of a scale to measure self-compassion. Self Identity 2003 Jul;2(3):223-250 [FREE Full text] [doi: 10.1080/15298860309027]
52. Neff KD. The self-compassion scale is a valid and theoretically coherent measure of self-compassion. Mindfulness 2015 Dec 18;7(1):264-274 [FREE Full text] [doi: 10.1007/s12671-015-0479-3]
53. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
54. Thombs BD, Benedetti A, Kloda LA, Levis B, Nicolau I, Cuijpers P, et al. The diagnostic accuracy of the Patient Health Questionnaire-2 (PHQ-2), Patient Health Questionnaire-8 (PHQ-8), and Patient Health Questionnaire-9 (PHQ-9) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. Syst Rev 2014 Oct 27;3(1):124 [FREE Full text] [doi: 10.1186/2046-4053-3-124] [Medline: 25348422]
55. American Psychiatric Association. Diagnostic And Statistical Manual of Mental Disorders, 5th Edition. New York, NY: American Psychiatric Association; 2013.
56. Shin C, Lee SH, Han KM, Yoon HK, Han C. Comparison of the usefulness of the PHQ-8 and PHQ-9 for screening for major depressive disorder: analysis of psychiatric outpatient data. Psychiatry Investig 2019 Apr;16(4):300-305 [FREE Full text] [doi: 10.30773/pi.2019.02.01] [Medline: 31042692]
57. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
58. Lewis JR. The system usability scale: past, present, and future. Int J Hum Comput Interact 2018 Mar 30;34(7):577-590 [FREE Full text] [doi: 10.1080/10447318.2018.1455307]
59. Schrepp M, Hinderks A, Thomaschewski J. Construction of a benchmark for the User Experience Questionnaire (UEQ). Int J Interact Multimed Artif Intell 2017 Mar;4(4):40-44 [FREE Full text] [doi: 10.9781/ijimai.2017.445] [Medline: 34887180]
60. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
61. George D, Mallery P. SPSS for Windows Step by Step: A Simple Guide and Reference. Boston, MA: Allyn & Bacon; 2010.

62. Bonett DG, Wright TA. Sample size requirements for estimating pearson, kendall and spearman correlations. Psychometrika 2000 Mar;65(1):23-28 [FREE Full text] [doi: 10.1007/bf02294183]
63. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. New York, NY: Routledge; 1988.
64. Lakhan R, Agrawal A, Sharma M. Prevalence of depression, anxiety, and stress during COVID-19 pandemic. J Neurosci Rural Pract 2020 Oct;11(4):519-525 [FREE Full text] [doi: 10.1055/s-0040-1716442] [Medline: 33144785]
65. Mazza MG, De Lorenzo R, Conte C, Poletti S, Vai B, Bollettini I, COVID-19 BioB Outpatient Clinic Study group, et al. Anxiety and depression in COVID-19 survivors: role of inflammatory and clinical predictors. Brain Behav Immun 2020 Oct;89:594-600 [FREE Full text] [doi: 10.1016/j.bbi.2020.07.037] [Medline: 32738287]
66. Grassini S, Laumann K, Rasmussen Skogstad M. The use of virtual reality alone does not promote training performance (but sense of presence does). Front Psychol 2020 Jul 17;11:1743 [FREE Full text] [doi: 10.3389/fpsyg.2020.01743] [Medline: 32765384]

Abbreviations

CFT: compassion-focused therapy
GUI: graphical user interface
HMD: head-mounted display
iVR: individualized virtual reality
PHQ-8: 8-item Patient Health Questionnaire
PHQ-9: 9-item Patient Health Questionnaire
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
SCS: Self-Compassion Scale
SUS: System Usability Scale
UEQ: User Experience Questionnaire
VR: virtual reality
VRET: virtual reality exposure therapy

Edited by J Torous; submitted 27.03.23; peer-reviewed by N Mungoli, E Korshakova; comments to author 16.05.23; revised version received 10.07.23; accepted 11.07.23; published 02.10.23.

Please cite as:
Halim I, Stemmet L, Hach S, Porter R, Liang HN, Vaezipour A, Henry JD, Baghaei N
Individualized Virtual Reality for Increasing Self-Compassion: Evaluation Study
JMIR Ment Health 2023;10:e47617
URL: https://mental.jmir.org/2023/1/e47617
doi:10.2196/47617
PMID:37782537

Original Paper

The Impact of a Gamified Mobile Mental Health App (eQuoo) on Resilience and Mental Health in a Student Population: Large-Scale Randomized Controlled Trial

Silja Litvin[1], MSc; Rob Saunders[2], PhD; Philip Jefferies[3], PhD; Hayley Seely[4], MSc; Patrick Pössel[4], PD; Stefan Lüttke[5], Dipl-Psych

[1]Department Psychologie, Ludwig-Maximilians-Universität, München, Germany

[2]Division of Psychology and Language Sciences, Clinical Education and Health Psychology, University College London, London, United Kingdom

[3]Resilience Research Centre, Faculty of Health, Dalhousie University, Halifax, NS, Canada

[4]Department of Counseling and Human Development, College of Education and Human Development, University of Louisville, Kentucky, KY, United States

[5]Department für Klinische Psychologie und Psychotherapie, Institute für Psychologie, Universität Greifswald, Greifswald, Germany

Corresponding Author:
Silja Litvin, MSc
Department Psychologie
Ludwig-Maximilians-Universität
Geschwister-Scholl-Platz 1
München, 80539
Germany
Phone: 49 89 21800
Email: silja@psycapps.com

Abstract

Background: With many digital mental health interventions failing to engage clients for enough time to demonstrate substantive changes to their well-being and with only 2% of all digital solutions on app stores having undergone randomized controlled trials, the rising demand for mental health prevention and early intervention care is not being met. Young adults in particular struggle to find digital well-being apps that suit their needs.

Objective: This study explored the effects of eQuoo, an evidence-based mental health game that teaches psychological skills through gamification, on resilience, depression, anxiety, and attrition in a student population.

Methods: In total, 1165 students from 180 universities in the United Kingdom participated in a 5-week, 3-armed randomized controlled trial. Participants were randomly allocated into 1 of 3 groups: eQuoo users, users of a treatment-as-usual evidence-based cognitive behavioral health app called Sanvello, and a no-intervention waitlist. The Rugged Resilience Scale, Generalized Anxiety Disorder–7, and Patient Health Questionnaire–8 were administered to all participants at baseline and every 7 days until completion.

Results: A repeated measures–ANOVA revealed statistically significant increases in resilience scores in the test group (P<.001) compared with both control groups (Sanvello: P=.10 and waitlist: P=.82) over 5 weeks. The app also significantly decreased anxiety and depression scores (both P<.001). With 64.5% (251/389) adherence, the eQuoo group retained 42% more participants than the control groups.

Conclusions: Digital health interventions such as eQuoo are effective, scalable, and low-cost solutions for supporting young adults and are available on all leading mobile platforms. Further investigation could clarify the extent to which specific elements of the eQuoo app (including gamification) led to better outcomes.

Trial Registration: German Clinical Trials Register (DRKS) DRKS00027638; https://drks.de/search/en/trial/DRKS00027638

(JMIR Ment Health 2023;10:e47285) doi:10.2196/47285

KEYWORDS

mobile health; mHealth; gamification; resilience; randomized controlled trial; RCT; mental health; apps; mobile health; mobile game; mobile games; serious game; depression; anxiety; university; college; student; students; controlled trial; controlled trials; young adult; mobile phone

Introduction

Background

Digital health interventions (DHIs) have repeatedly been shown to be effective in improving mental well-being and relieving the symptoms of depression [1-4] and anxiety [1-3,5,6] in young people. They are efficacious in young people exhibiting elevated symptoms of anxiety and depression as well as in those who have been diagnosed with an anxiety or depressive disorder [2]. Randomized controlled trials (RCTs) have also revealed that DHIs can improve mood and promote emotional resilience in adolescents in the general populations [7,8].

DHIs may have other potential benefits for adolescents and young adults. For example, even when barriers to health systems are low (eg, health insurance coverage or free health care services) and individuals report elevated depression or anxiety symptoms, the rates of engagement with mental health support are low [9,10]. This suggests that there may be further obstacles, including attitudinal barriers, such as a preference for self-reliance and the desire to avoid perceived stigmatization [9-12], and off-putting aspects related to the delivery of the service, such as physical access issues [11], in addition to long waitlist times [12,13]. DHIs may help address these barriers, as research has shown that young people are drawn to digital health solutions [14,15], especially those on mobile phones [16].

DHIs delivered via mobile phones present a promising route for supporting the mental health of young people. For example, the ownership of mobile phones has grown enormously in recent years, with a national US survey finding that in 2019, approximately 90% of 16- to 18-year-olds possessed a smartphone, compared with 75% in 2015 [17]. There is also a high motivation to use mobile phone apps to improve mental health [18], which may stem from not only their convenience but also their control in terms of when and how much an individual chooses to engage with a DHI.

DHIs delivered by mobile phones typically take the form of “unguided” apps. Although numerous systematic reviews have found that DHIs for young people were effective when they incorporated some sort of human support [4,6,9,19], such as professional feedback or a live link to a professional, their superiority over unguided DHI approaches [20] has not been confirmed conclusively. For instance, in some studies, the effect in favor of guided DHIs was small [20] or not statistically significant when only studies with a low risk of bias were considered [19]. For smartphone-based interventions for depressive symptoms, a recent meta-analysis of RCTs by Firth et al [21] found that only apps without professional feedback produced moderate positive effects. Another meta-analysis found larger effect sizes on stress levels and quality of life when apps provided professional guidance [22].

Despite the convenience, appeal, and promise of DHIs delivered by mobile phones, evidence for their efficacy remains unsatisfactory. For example, review studies can neglect the modality of a DHI, which is important when the current bulk of studies involve comparisons of guided internet-based or computer-based cognitive behavioral therapy (CBT) with inactive controls [2,23]. Furthermore, in a recent review of 19 studies that investigated the use of serious games and virtual reality for the treatment of common mental health problems in children and young people, 10 (53%) had no comparison group, 3 (16%) used a waitlist condition, 4 (21%) used another DHI control group condition, and only 2 (11%; investigating the same DHI) used face-to-face interventions [3,4]. The effect of DHIs was largest when DHIs were compared with nonactive control conditions and was small or even nonexistent when there was an active comparator [4,22,23]. Recent meta-analyses that have specifically explored studies involving mobile phone apps for depression and anxiety have suggested that they have the potential to reduce symptoms when compared with inactive controls [21,24], and a recent meta-review by Goldberg et al [25] supports these findings. The authors noted that there are few studies involving active controls, which is important for establishing the effect of using a particular intervention.

A further limitation affecting recent studies on the efficacy of DHIs on mental health in young people (including DHIs delivered via mobile phones) was highlighted by Grist et al [4] in their meta-analysis of the efficacy of DHIs for depression and anxiety in children and adolescents. The authors found that many of the studies were underpowered. Insufficient sample sizes can promote favorable effects of particular interventions [1,2,5], reducing the likelihood that a statistically significant result reflects a true effect [26,27].

Aside from evaluation issues, a further issue troubles DHI research. In 2005, Eysenbach [28] noted that a considerable number of study participants stopped engaging with them (ie, nonuse attrition or nonadherence) or dropped out; therefore, they could not be followed up (dropout attrition). For example, Linardon et al [22] found that approximately 25% of participants using smartphone-delivered DHIs for common mental health problems dropped out before short-term follow-up (≤8 weeks) and up to 33% dropped out before long-term follow-ups. Similarly, Hollis [23] found that for DHIs targeting depression in young people, attrition ranged from 0% for a computerized in-house attention bias modification [29] to 42.5% for a synchronous chat intervention [30]. Although this is potentially problematic for evaluating the efficacy of DHIs, the lack of or diminishing engagement is troubling, given the benefits of engaging with such interventions. Developers of non–practitioner-led (unguided) DHIs face an overarching challenge: how to engage clients long enough for interventions to have an impact.

One approach to combating attrition and nonengagement in DHI design is the use of game design elements in nongaming contexts or “gamification” [31]. Although the gamification of software and technology in a behavior change context has gained traction over the past few years [32], it is still in the early stages in terms of methodology, classification, and implementation. In total, 2 relevant research streams have emerged: (1) persuasive systems and technology and (2) gamification as a tool that enables playful experiences and enhances engagement [33]. Persuasive systems are geared toward using software and technology to instigate behavioral change regarding preset goals, such as weight loss, developing beneficial psychological coping mechanisms, or even environmental sustainability actions [34].

Persuasive strategies can be implemented endogenously; software can be developed via game mechanics, which are intrinsically inherent to the gaming experience [35]. Endogenous games are designed to enhance the gaming experience to the highest degree possible; external goals of persuasive systems, where game mechanics rules are coupled with outcome goals, are likely to hinder or complicate the process. This is sometimes referred to as gamifying rather than gamification. The more commonly used gamification development procedure for persuasive games, where elements are layered on an existing framework of information exchange geared toward a predefined goal (such as weight loss), is exogenous [35]. Although it makes sense to implement evidence-based strategies, such as CBT, as the foundation of a persuasive system for a DHI, it can be at odds with the second research stream: playful experiences and enhanced engagement. Poorly implemented exogenous factors, such as badges or rewards that are not connected to the gameplay, have reportedly had the opposite effect on the playing population, leading to the rejection of the tool [36]. An example of an endogenous game would be eQuoo, where the intervention is woven into a story framework that is appealing in and of itself, whereas an exogenous element would be a badge given after a CBT exercise is successfully completed.

In addition to gamification, interventions aimed at building skills associated with improved mental health may help enhance mental health overall. As conceptualizations of mental health expand beyond the pathogenic to the salutogenic [37], researchers have started to investigate mental health in terms of both the capacity to manage difficulties and the absence of illness [38]. Resilience is widely considered to be a protective factor against negative mental health outcomes. Resilience is the capacity to overcome or adapt to adversity and thus stay mentally healthy or regain one's mental health following significant challenges [39]. Previously thought to reflect the presence of personality characteristics, such as *grit*, resilience is now more commonly understood as a process in which various modifiable protective factors are drawn upon. This leads to positive outcome trajectories [40]. Protective factors exist at various systemic levels, including psychological (eg, self-efficacy and motivation) and social (eg, supportive peer relationships and a sense of community belonging) [41]. The realization that many individuals experience significant adversity during their lifetime [42] has led to an increasing demand for resilience-building programs aimed at promoting mental health [34].

This Study

In summary, research has shown that DHIs can be effective in promoting mental well-being and reducing symptoms of depression and anxiety in young people. However, the evidence base is lacking, and given their unique qualities, there is a need to specifically test the advantages of using DHIs involving gamification and through sufficiently powered 3-armed RCTs (DHIs vs active controls vs nonactive controls) that also report on trial completion and attrition.

Primary Outcome Hypothesis—Resilience Levels

We hypothesized that resilience levels would increase significantly over the course of the intervention period in the gamified intervention group and that resilience would be significantly higher in the intervention group than in the active control and waitlist control groups.

Secondary Outcome Hypotheses—Depressive and Anxiety Symptoms and Attrition

We also hypothesized that depression and anxiety symptoms would decrease significantly over the course of the intervention period in the gamified intervention group and that depression and anxiety symptoms would be reduced significantly in the gamified intervention group compared with both the active control and waitlist control groups.

Finally, we hypothesized that rates of attrition would be significantly lower in the gamified intervention group at the last assessment than in the active control and waitlist control groups.

Methods

Inclusion and Exclusion

Participants had to be aged ≥18 years and enrolled as college or university students. They had to have access to a smartphone or tablet device and an app store (App Store or Google Play) to install the eQuoo app. No further exclusion criteria were applied.

Sampling Procedure and Participant Consent

Recruitment took place in 2 waves between March and April 2021 using the UNiDAYS subscriber database, a global student discount platform. To qualify for UNiDAYS, people must provide a valid college or university email address. In the first wave, 9000 people who were then registered on the platform with their University College London affiliation were emailed an invitation to participate in a trial exploring the effects of a mental health app on resilience with a link to the trial landing page; 443 people signed up for the trial and completed the baseline assessment. This figure represented 8.6% of those emailed. On the landing page, the study goal was explained briefly, the study protocol was shared, and a consent box needed to be clicked that read "I have understood the trial terms and I consent." This checkbox needed to be checked for participants to join the trial; without it, they could not gain access to the questionnaires and information on what to test, thus ensuring consent. In the second wave, 100,000 randomly selected UNiDAYS subscribers were emailed the same invitation; 724 participants completed the baseline assessment. This figure represented 1.6% of those emailed. The trial was completed and closed, and all participants were debriefed in June 2021.

Ethics Approval

The study (an RCT) was approved by the University College London Ethics Committee (0501/001), and the authors confirm that all ongoing and related trials are registered with the German Clinical Trials Register (DRKS00027638). This study was conducted in accordance with the principles of the Declaration of Helsinki.

Sample Size, Power, and Precision

An a priori power analysis using G*Power 3.1.9.4 revealed that a sample size of 207 participants was required to detect a within-between interaction for the primary outcome (input

parameters: repeated measures [RM]–ANOVA, effect size *f*=0.10; Cronbach α=.05; power=0.95; number of groups=3; number of measurements=6; RM intercorrelation=0.50; nonsphericity correction=1). However, given the high attrition rates reported in previous research (eg, 40%-60% [28]), we aimed to enroll a minimum of 850 participants.

Measures and Instruments

Primary Outcome: Resilience

As eQuoo exists primarily in the digital space of prevention and early intervention, resilience was considered the primary outcome. To assess resilience, we used the Rugged Resilience Measure (RRM). The RRM is a 10-item self-report questionnaire designed to measure key psychologically protective factors that foster resilience [43]. Participants respond to the items on a 5-point Likert-type scale (1=*not at all*, 2=*a little*, 3=*somewhat*, 4=*quite a bit*, and 5=*a lot*). The questionnaire was initially validated with a sample of young adults (aged 16-30 years), which matched the population of the study. As the understanding of resilience has shifted from a fixed trait to a process encompassing the development and application of skills and resources that support positive outcomes despite the experience of distress [44,45], the RRM taps the key internal resources necessary to initiate said development. In this study, the internal consistency of the RRM was .87 (Cronbach α).

Secondary Outcomes: Anxiety, Depression, and Attrition

The secondary outcomes in this study were mental health (the level of anxiety and depression symptoms in particular) and attrition. The generalized anxiety disorder–7 (GAD-7)-item scale was used to assess anxiety. This is a widely applied 7-item measure of generalized anxiety symptoms [46]. Each item was scored from 0 to 3 (0=*notat all*, 1=*several days*, 2=*more than half the days*, and 3=*nearly every day*). The internal consistency of the GAD-7 was high (Cronbach α=.83). We also used the Patient Health Questionnaire–8 (PHQ-8). The PHQ-8 is a well-established screening measure for depressive symptoms in large clinical studies [47] and encompasses the American Psychological Association's *Diagnostic and Statistical Manual of Mental Disorders* (5th edition) criteria for a depressive episode (with the exception of the item regarding self-harm). It was preferred to the more widely known PHQ-9 [27] because the study design did not allow us to intervene in the event of reporting self-injurious behavior. In the PHQ-8, participants are asked to rate how far they have been bothered by each symptom over the previous 2 weeks on a 4-point Likert-type scale (0=*not at all*, 1=*several days*, 2=*more than half the days*, and 3=*nearly every day*). The internal consistency of the PHQ-8 in this study was high (Cronbach α=.88). Finally, it was hypothesized that attrition would be significantly lower for the gamified group intervention at the last assessment (t5) than for other groups. In addition to measures of resilience, anxiety, and depression, participants provided demographic information regarding their age, sex, living situation, and level of study.

Design

A mixed factorial 3 (condition) × 6 (time) RM design was used. Participants were randomly assigned to a condition (eQuoo vs active control vs waitlist group, the details of which are described in the subsequent section) using a randomization generator provided by random.org [48], which randomizes based on atmospheric noise [49]. Across the study period, measurements were taken at the beginning (t0, baseline assessment), week 1 (t1), week 2 (t2), week 3 (t3), week 4 (t4), and week 5 (t5, end point).

Gamified Group Intervention (eQuoo)

The original version of the emotional fitness app eQuoo was a 5-week mental health game that presented psychoeducational material and psychological exercises based on the principles of CBT, systemic psychology, and positive psychology [50]. It combined a mix of endogenous and exogenous design features designed to maximize engagement. The game was played on mobile phone only and was available internationally on the Google Play store and the Apple App Store. It was designed to be a prevention and early intervention (preclinical threshold) tool. In a recent trial, the app was found to significantly reduce anxiety while improving resilience, perceived growth skills, and interpersonal relationships [51]. After examining feedback from thousands of users through clinical trials, focus groups, case studies, and players offering their opinions on the app stores and via email, the developers of eQuoo sought to revise the app to enhance enjoyment, immersion, and retention. This resulted in eQuoo, the Next Generation [50] (hereafter, eQuoo), and it is the version of the app that was tested in this study.

The app begins by explaining the player's role as a Lodestar. Each page has a picture with characters in it and a speech bubble with a maximum of 160 characters per screen. The player moves to the next page by tapping on the screen. They can also click a back button to reread the previous screen. The type of introduction used in eQuoo is commonly known as game lore, a game-specific mythology, or the so-called backstory of the general narrative within a game. Lore in video games has been proven to motivate players to read and learn more, making it more likely that they will engage, read, and therefore learn the skills presented [52]. In eQuoo, Lodestars travel through time and space to fight against the Quavering by growing their inner light. The following is the onboarding text:

> *For centuries, the Lodestars have watched over this world, and they would love for YOU to join them. Get ready for the ultimate adventure.*
>
> *You'll journey to different times in history, become friends with people from all walks of life. You'll learn skills that will set you on a path towards personal growth. And you'll help to counter a massive threat.*
>
> *This threat is called The Quavering. A force created from all the greed and negativity in the world. Whenever someone's inner light grows dim, The Quavering grows stronger. You may have experienced this? If you join the Lodestars, you will help fight the battle against The Quavering.*
>
> *So, here you are. You have been given a chance to grow your inner light. And to shine that light on others. Welcome... to the Lodestars.*

The game consists of multiple books of different genres (eg, fantasy, historical drama, and teen drama). Each book consists

of 8 to 10 chapters where the player can learn and repractice up to 4 skills. Before each new chapter of a book in the game, they are led through a gamified skill tutorial by the game's guide, Joy, who is introduced to have been a player themselves and is now a Lodestar, thanks to having completed the game. Once they have successfully learned the skills, they continue into the book where they—playing themselves via an avatar that they customized at the beginning of their journey—meet the characters of the game and are thrown into various situations where they need to correctly use the psychological skills that they learned in the tutorial. To facilitate between-session learning and incorporation of the necessary skills, users must wait 7 days until the next chapter unlocks. This allows them to practice the skills in real-life settings, which has a positive impact on the therapeutic outcomes [53]. The weekly lock is also to protect players from addictive patterns and not flood them with too much information that they have retained after only one session.

Of the 18 gamification elements named by Cheng et al [54], 11 have been incorporated into eQuoo:

1. Levels
2. Points—in the form of gem shards
3. Rewards—in the form of unlocking levels and completing gems
4. Narratives
5. Personalisation—in the form of the story choices
6. Customisation—in the form of the avatar
7. Mini games
8. Quests and challenges—in the form of stories
9. Badges—in the form of personality types
10. Artificial assistance—in the form of the guide, Joy
11. Unlockable content—in the form of a free trial

The intervention group was instructed to download eQuoo via the Apple or Google Play Store and install it on their digital device. Participants were informed at the start of the study that they could stop using the app at any time. After starting the game, the player is introduced to the game's lore. Participants are asked to design an avatar that resembles themselves as a virtual person in the game. The content of the application is divided into multiple multigenre stories that consist of 8 to 10 chapters presented as levels. In the chapters, users are first presented with 1 to 4 lessons that teach them psychological concepts, such as emotional bids [55], generalization [56], catastrophization [57], beliefs [58], and 52 psychological skills commonly used in therapeutic sessions for anxiety and depression (as well as prevention programs designed to increase resilience). After each lesson, the players can test their mastery of skills using a simple multiple-choice scenario. They are then either debriefed on why their choice was not the most beneficial (and are invited to choose another answer) or allowed to enter an interactive adventure story where they play themselves while practicing the skills in a low-cost environment. A low-cost environment means that the failure to succeed comes at a low cost, such as having to replay a level. All prompts during the psychoeducational part of the game to check whether the player has understood the skill are divided into 3 responses: (1) beneficial—the skill has been implemented by the player in a way that is beneficial for the player's mental health; (2) neutral—the skill has been ignored and not used, and the counter indication was not chosen; and (3) unbeneficial—the player chose an answer that is considered unbeneficial for the player's mental health.

The onboarding of the game consists of 3 levels of introduction to the lore; the building of the avatar; and the first in-app baseline assessment of the RRM, GAD-7, and PHQ-8 in the form of a pop-up chatbot where the player can chat with the game's guide, Joy, and fill out the questionnaires. After acquiring their second skill, they play the first chapter of the first story and hit the level lock until the next week. This ensures that they play only 1 level per week for 5 weeks of the clinical trial. Each survey includes a question that players can answer only if they have completed the level for the week. Biweekly in-app nudges and weekly emails pull the participants back into the game.

Active Control Group Intervention (Sanvello)

Participants in the active control group were instructed to download an evidence-based mental health app called *Sanvello*, which was used as treatment-as-usual or active control group. It was chosen specifically to explore the secondary hypothesis of attrition, as the app has already been tested in another RCT [59] and is associated with a reduction in depression [60] and anxiety [61]. Sanvello is based on CBT and includes psychoeducation, CBT exercises, notifications, and a diary. It includes free access to multiple modules, which was sufficient to cover the 5-week trial. We requested that the participants use it for a minimum of 10 minutes per week. Participants were informed at the start of the study that they could stop using the app at any time.

Waitlist Control Group

The waitlist group received no intervention but completed the questionnaires at the same time points as the control and intervention groups. After completing the trial, they were debriefed on the results and provided with a link to both the eQuoo and the Sanvello apps. The study information was available via the link.

Data Collection

Participants were reminded via weekly emails to complete the questionnaires at t1 to t5. In addition to the questionnaires, the participants in the eQuoo group were asked a question to prove that they had completed their level, and the participants in the active control group were asked if they had spent 10 minutes on the Sanvello app. Data were collected using LimeSurvey, a widely used secure open-source tool.

Statistical Analysis

Differences in participant characteristics between each intervention arm were first compared using chi-square tests of independence (for categorical variables) and one-way ANOVAs (for continuous variables).

Attrition rates and the number of participants who completed the levels at all time points were compared across groups. Attrition was defined as not completing the assessments past t0. This is consistent with previous studies that defined attrition as the failure to complete the study protocol associated with the

intervention [22]. To compare the likelihood of attrition between intervention arms, logistic regression models were constructed, with both unadjusted and adjusted (for sex, age, living situation, and baseline measures) odds ratios (ORs) and 95% CIs reported.

A standardized outcome was then created using the last available measure, even if it was only the participant's baseline, and this was carried forward (*intention to treat*) [62]. Initially, paired sample 2-tailed *t* tests were conducted within each intervention arm to assess statistically significant changes in the primary and secondary outcomes. Pre-post effect sizes (Cohen *d*) were calculated using 0.2, 0.5, and 0.8, which were used as thresholds to signify small, medium, and large effects, respectively. Differences between intervention arms in end point (t5) scores were explored using linear regression models (for each outcome) with a baseline measure, and age, sex, living situation, and study level (ie, undergraduate or postgraduate) were entered as covariates. Adjusted end point means were estimated, and the magnitude of between-group differences was explored by calculating Cohen *d*.

To assess differences between interventions over time points, initial analyses used RM-ANOVAs with time (6 levels) entered alongside the intervention group (3 levels). To assess the impact of listwise deletion on these models, further analysis was conducted using mixed effects models exploring changes in the primary and secondary outcomes over time, entering the intervention arm as an independent variable and age, sex, and living situation as covariates. These mixed effects models, using restricted maximum likelihood estimation, used all available data at each time point. The survey was programmed so that all fields were mandatory, that is, participants either filled out the entire questionnaire or did not participate. The data were published using the Open Science Framework [63].

Results

Attrition and Characteristics by Group

Of the 1167 individuals who were recruited for the study, 2 individuals reported being aged <18 years and were excluded; therefore, the final sample comprised 1165 participants. These individuals were then randomly allocated to the study groups: 389 (33.39%) were placed in the eQuoo group, 384 (32.96%) were placed in the Sanvello group, and 392 (33.65%) were placed in the waitlist group. Figure 1 shows the participant flow diagram and the proportion of participants providing data at each time point. All participants (except 1 individual in the eQuoo group) completed baseline (t0) measures, and end point (t5) measures were available for 251 (64.5%) out of 389 participants in the eQuoo group, 77 (20%) out of 384 participants in the Sanvello group, and 101 (25.8%) out of 392 participants in the waitlist group. Of the eQuoo participants, 349 completed at least one measure after baseline (but potentially not t5), and 123 of the Sanvello participants completed at least one nonbaseline measure, as did 211 of the waitlist group.

The likelihood of attrition (defined as completing only the baseline measures) was compared for the intervention conditions. Although only 10.3% (40/389) of the participants in the eQuoo group met the criteria for attrition, the rates for the Sanvello and waitlist groups were 67.9% (261/384) and 46.2% (181/392), respectively. The odds of attrition were significantly higher in the Sanvello group than in the eQuoo group (OR 18.51, 95% CI 12.52-27.38) as well as in the waitlist group (OR 7.48, 95% CI 5.10-10.97). After adjusting for age, sex, living situation, study level, and all 3 baseline measures, Sanvello was compared with eQuoo (OR 19.27, 95% CI 12.91-28.77), as was the waitlist (OR 7.34, 95% CI 4.96-10.85).

Table 1 presents differences in participant characteristics between intervention arms. The comparative statistics suggest the groups were balanced in terms of age, sex, and baseline resilience scores but not living situation, study level, or initial depression and generalized anxiety symptom scores.

Figure 1. Study procedure and number of completers at each stage of the assessment.

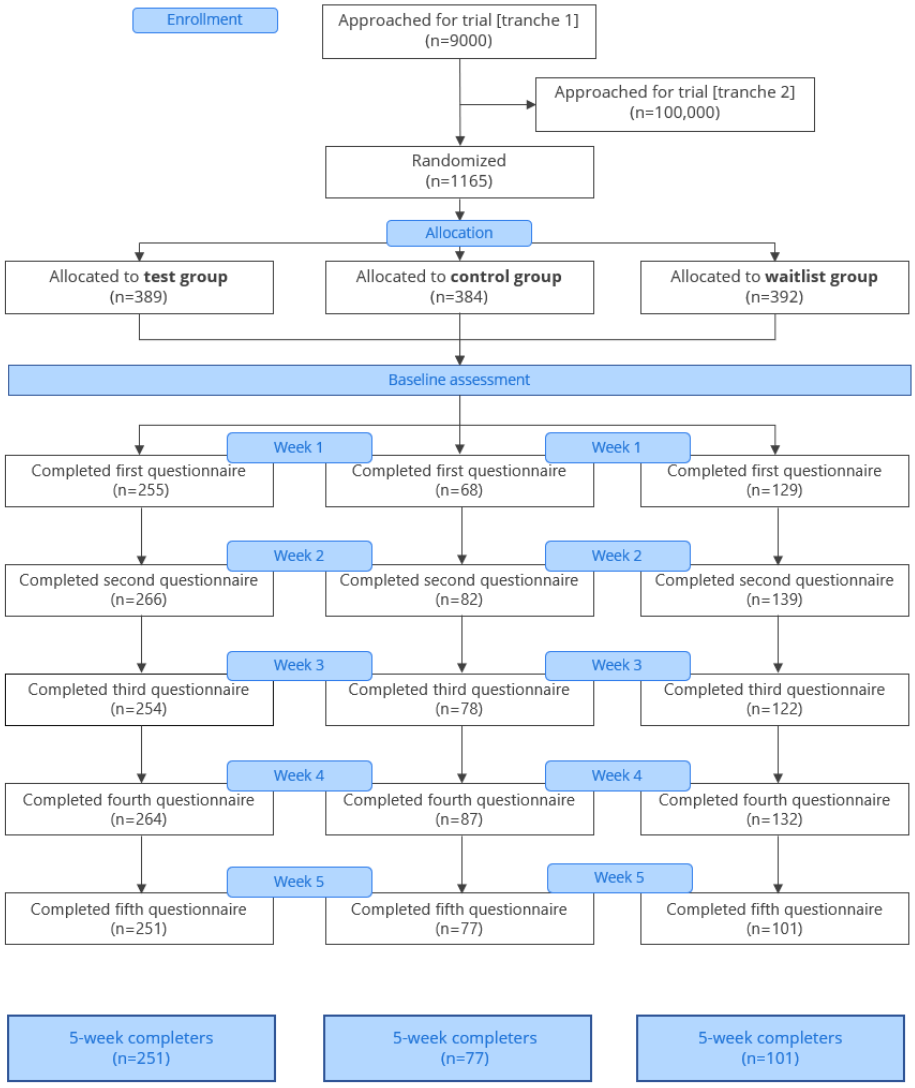


Table 1. Participant characteristics between the intervention arms.

Characteristic	Total (n=1165)	eQuoo (n=389)	Sanvello (n=384)	Waitlist (n=392)	*P* value
Wave, n (%)					.91
Wave 1	443 (38)	145 (37.3)	149 (38.8)	149 (38)	
Wave 2	722 (62)	244 (62.7)	235 (61.2)	243 (62)	
Sex, n (%)					.43
Female	891 (76.5)	285 (73.3)	297 (77.3)	309 (78.8)	
Male	251 (21.5)	95 (24.4)	79 (20.6)	77 (19.6)	
Other	18 (1.5)	6 (1.5)	6 (1.6)	6 (1.5)	
Missing	5 (0.4)	3 (0.8)	2 (0.5)	0 (0)	
Living situation, n (%)					.03
Close family or relatives	471 (40.4)	142 (36.5)	170 (44.3)	159 (40.6)	
Student apartment	234 (20.1)	83 (21.3)	75 (19.5)	76 (19.4)	
Own apartment	329 (28.2)	126 (32.4)	102 (26.6)	101 (25.8)	
Other	124 (10.6)	36 (9.3)	32 (8.3)	56 (14.3)	
Missing	7 (.6)	2 (0.5)	5 (1.3)	0 (0)	
Level of education, n (%)					.02
Undergraduate	717 (61.5)	217 (55.8)	246 (64.1)	254 (64.8)	
Postgraduate	446 (38.3)	170 (43.7)	138 (35.9)	138 (35.2)	

Effects of Treatment on Primary Outcome (Resilience)

Pre- and posttest scores for resilience are presented in Table 2. There was a significant improvement (medium effect size) in resilience among eQuoo participants (t_{387}=18.35; *P*<.001; Cohen *d*=0.58) but not among Sanvello or waitlist participants (*P*=.10 and .82, respectively).

Differences in end point scores were then compared using linear regression models, controlling for baseline resilience score, age, sex, living situation, and study level. Significantly higher resilience scores were observed in the eQuoo condition compared with Sanvello (*b*=−4.41, 95% CI −5.09 to −3.74; *P*<.001) and in eQuoo compared with the waitlist condition (*b*=−4.78, 95% CI −5.46 to −4.10; *P*<.001).

RM-ANOVA models were then used to compare resilience scores at each time point between the intervention arms, and mixed effects models were constructed to explore the change between the intervention arms. A total of 174 participants (n=78, 44.8% eQuoo; n=37, 21.3% Sanvello; and n=59, 33.9% waitlist) completed the measures at all time points and were included in the RM-ANOVA. A significant effect was observed for time ($F_{5,855}$=18.51; *P*<.001), intervention arm ($F_{2,171}$=8.68; *P*<.001), and intervention interaction ($F_{10,855}$=9.69; *P*<.001). Mixed effects models, including age, sex, living situation, and study level as covariates led to a significant intervention-in-time interaction: scores were significantly lower over time for participants in the Sanvello arm (*b*=−0.86, 95% CI −1.06 to −0.66; *P*<.001) and waitlist arm (*b*=−1.25, 95% CI 1.42 to −1.08; *P*<.001), compared with the eQuoo arm (Figure 2).

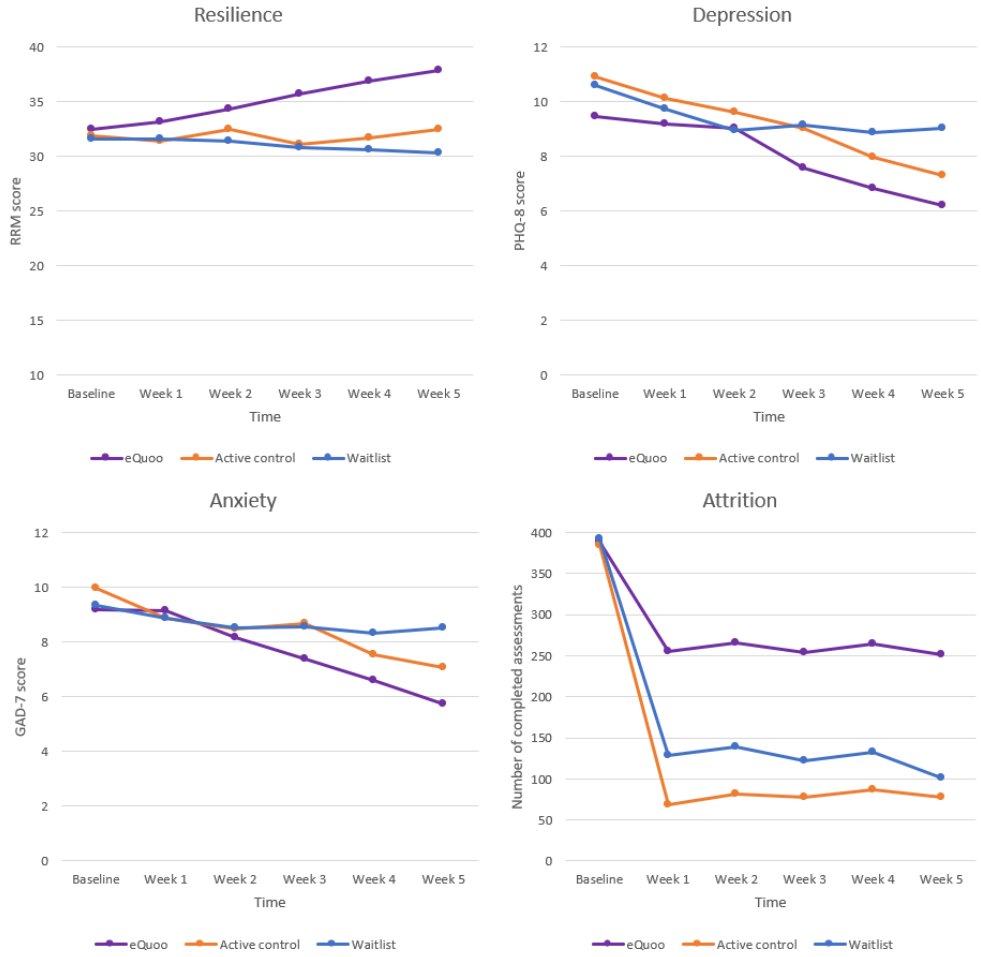
Table 2. Pre- and postoutcome scores between intervention arms.

Outcome or intervention	Sample, n (%)	Baseline, mean (SD)	End point[a], mean (SD)	*t* test (*df*)	*P* value
Resilience					
eQuoo[b]	388 (99.74)	32.47 (8.35)	37.88 (7.17)	−18.35 (387)	<.001
Sanvello	384 (100)	31.86 (7.66)	32.42 (7.94)	−1.65 (383)	.10
Waitlist	392 (100)	31.60 (8.67)	30.34 (9.01)	0.22 (391)	.82
Anxiety					
eQuoo	388 (99.74)	9.17 (6.91)	5.75 (4.26)	12.09 (387)	<.001
Sanvello	384 (100)	9.98 (5.62)	7.06 (5.49)	5.77 (383)	<.001
Waitlist	392 (100)	9.33 (5.68)	8.51 (5.82)	1.97 (391)	.049
Depression					
eQuoo	388 (99.74)	9.44 (5.61)	6.18 (4.59)	20.91 (387)	<.001
Sanvello	384 (100)	10.89 (5.61)	7.30 (5.39)	5.45 (383)	<.001
Waitlist	392 (100)	10.59 (5.78)	9.02 (5.75)	3.45 (391)	.001

[a]End point scores included last observation carried forward for participants who did not complete an end point measure.

[b]One individual did not complete a baseline assessment.

Figure 2. Average (unadjusted) weekly outcome scores by week (95% CIs) for resilience, depression, and anxiety. All available scores are included. GAD-7: generalized anxiety disorder–7; PHQ-8: Patient Health Questionnaire–8; RRM: Rugged Resilience Measure.



Effects of Treatment on Secondary Outcomes

Generalized Anxiety

In the pre-post analysis, GAD-7 scores significantly decreased within all 3 conditions (eQuoo: t_{386}=22.86, *P*<.001; Cohen *d*=0.60; Sanvello: t_{347}=5.77, *P*<.001; Cohen *d*=0.16; and waitlist: t_{391}=1.97, *P*=.049; Cohen *d*=0.08; Table 2). Linear regression models demonstrated significantly lower GAD-7 end point scores in the eQuoo condition compared with the Sanvello (*b*=2.28, 95% CI 1.83-2.73; *P*<.001) and waitlist conditions (*b*=2.54, 95% CI 2.08-2.99; *P*<.001).

The RM-ANOVAs, including only those participants who completed every time point, showed that although there was a significant main effect of time ($F_{5, 855}$=29.337; *P*<.001), there was no main effect of condition (*P*=.35) or a condition-by-time interaction (*P*=.10). Further analysis performed using mixed effects models (which included the covariates listed earlier) indicated a significant time-by-condition interaction; over time, GAD-7 scores were lower in the eQuoo group compared with the waitlist group (*b*=0.56, 95% CI 0.40-0.71; *P*<.001). The difference between the eQuoo and Sanvello conditions was not significant (*b*=0.12, 95% CI 0.06-0.30; *P*=.20; Figure 2).

Depression

In the pre-post analysis, the PHQ-8 scores significantly decreased in all 3 groups (*P*≤.001). However, although changes in depression scores produced a medium effect (Cohen *d*=0.58), they were small for the Sanvello and waitlist groups (Cohen *d*=0.14 and 0.13, respectively; Table 2).

As with the GAD-7 results, the linear regression demonstrated significantly lower PHQ-8 end point scores in the eQuoo condition compared with the Sanvello group (*b*=2.47, 95% CI 2.91-2.93; *P*<.001) and in the eQuoo condition compared with the waitlist group (*b*=2.46, 95% CI 2.00-2.91; *P*=.15).

The RM-ANOVAs for PHQ-8 scores, including only those individuals who completed every time point, indicated that although there was a significant main effect of time ($F_{5,855}$=55.392; *P*<.001), there was no main effect of condition (*P*=.22) or a condition-by-time interaction (*P*=.11). The mixed effects models (which included the covariates listed earlier) indicated a significant time-by-condition interaction, showing that over time, PHQ-8 scores remained lower for the eQuoo group compared with waitlist group (*b*=0.40, 95% CI 0.25-0.55; *P*<.001). The difference between the eQuoo and Sanvello conditions was not significant (*b*=0.09, 95% CI −0.09 to 0.27; *P*=.31).

Discussion

Principal Findings

The results suggest that using the gamified mental health app, eQuoo is an effective pathway for improving mental health and resilience. In particular, the use of eQuoo was related to increased resilience scores over time; this was not the case for the nongamified or waitlist groups. Participants using eQuoo also reported lower depression and anxiety scores at the end of the intervention compared with the other 2 groups and were significantly less likely to drop out of treatment. The findings indicated that both apps were effective in reducing anxiety and depression symptoms, although the larger effect sizes associated with the improvements suggest that eQuoo was more effective in this regard.

Mental health issues are pervasive and particularly impactful in student populations [64,65]. Unfortunately, the gap between mental health treatment needs and access to care continues to grow [66,67]. Although DHIs have been identified as meaningful alternatives to face-to-face therapy in student populations [9,11], they have significant shortcomings [68-70] in terms of motivation, interest, and engagement [69]. As the prevalence of mental health issues [69] and the demand for digital mental health solutions, such as telehealth and internet therapy [71] have increased, research investigating digital, app-based interventions is especially timely. This study contributes to the quest for solutions by presenting eQuoo as a gamified DHI that can improve mental health by reducing depression and anxiety symptoms while building resilience and compares favorably with a well-established, nongamified mental health app.

Implications

Gamification using mobile mental health apps has been suggested as a way to increase engagement in mental health services [72-74]. This study supports this suggestion by highlighting the benefits of DHIs incorporating gamification and resilience training (particularly when eQuoo was used). Although further research that distinguishes the effects of gamification and includes more diverse populations is necessary, the results of this study suggest that mobile mental health apps can assist in mental health treatment and resilience building.

Furthermore, our findings are particularly relevant, given the current events. Global factors, such as the 2019 COVID-19 pandemic and the accompanying restrictions, have been linked to an increased prevalence of mental health issues and difficulties in accessing support [75]. According to a recent World Health Organization survey, the pandemic has disrupted critical mental health services in 93% of countries worldwide while increasing demand for them [50]. Our findings suggest that DHIs may be a meaningful option in lieu of traditional mental health services for those seeking help.

The finding that resilience scores improved for eQuoo users is also promising, as this suggests that the impact and lingering effects of future adversity on mental health may be less for those with higher scores [74,76], among other benefits [77]. In other words, the inclusion of resilience training elements in DHIs such as eQuoo can facilitate a reduction in psychopathology symptoms such as depression and anxiety but may also serve a protective function against these symptoms for users when encountering difficulties later on. This would be especially important for young adults who experience many stressful transitions in different life domains [78]. Further longitudinal research is needed to explore this empirically, capturing both the long-term impact of eQuoo and the impact of subsequent stressors.

With continued research, DHIs that incorporate gamification and resilience training may play a pivotal role in improving access to mental health services and preventing mental health issues. They could be integrated into standard face-to-face treatment as part of the client's homework, offered as a pretreatment option, or used as a mental health resource for all college students as part of an intervention and prevention strategy.

Strengths and Limitations

This study has several strengths. It is one of only a few follow-up RCTs involving a gamified mental health app [51]. Its large participant pool enables calculations to be performed with a large power size, thus ensuring the reliability of the outcome. This study is based on past research on the efficacy of the same gamified app that has been tested and shown to increase resilience and positive relations with others and decrease anxiety symptoms in a very different sample (employed adults) [51]. Given this (and the similarities in findings between the present and previous studies), the benefits of eQuoo may be generalized to different populations. Finally, the study included 3 treatment arms, including a treatment-as-usual active control condition using a nongamified app that has been tested [59] and shown to be efficacious [59,61]. Including this treatment-as-usual active control condition allowed the gamified app and nongamified app to be usefully compared. In addition, unlike previous studies, we have explicitly reported on dropout and nonuse using engagement data.

This study has some limitations. Although including a treatment-as-usual active control condition was important to evaluate eQuoo, the 2 apps were not structurally equivalent interventions. Ordinarily, structurally equivalent interventions are identical in terms of the number and duration of sessions, settings (group vs individual), level of therapists' experience, and adaptability of the therapy to the client [79]. The 2 apps used in this study have different purposes and use different interventions. Although the eQuoo is designed to teach psychological skills to decrease depression and anxiety and increase resilience, and many of these skills are based on CBT principles that overlap with the structure of Sanvello, the latter relies exclusively on CBT psychoeducation and exercises [59]. In other words, there were meaningful differences between the 2 apps, in addition to gamification. A meta-analysis comparing in-person psychotherapies found no differences between programs when they had structural equivalence [79]. Thus, it may be that the differences between eQuoo and Sanvello were not solely caused by gamification. In light of this, future studies might compare structurally equivalent gamified and nongamified apps, although a robust comparison would involve a version of the experimental app with gamification features removed, to ensure that there were no other explanatory variables, such as the look or interface of the app.

Another limitation of this study is that it did not record the duration of participants' use. Hence, it was not possible to examine whether the level of use was linked to attrition, an important phenomenon discussed in eHealth research [28]. In other words, although the attrition among participants using the gamified app was lower than that in the nongamified app and waitlist groups, it was not possible to examine the influence of active engagement. This could have been accomplished by including time-of-use data from the apps.

A further limitation was the lack of follow-up; therefore, we could not draw conclusions about the stability of the observed effects. As such, future researchers should investigate the impact of gamification on nonuse attrition, engagement, long-term attrition, and mental health.

Similarly, it was unclear which aspects of the intervention resulted in the observed effects. As the eQuoo was created as a skills training app, it may have been that the development of resilience and reduction in mental health symptomatology was the result of skills acquisition. However, specific skill acquisition was not examined. Studies in which individual skills were trained were excluded to investigate the mechanisms of the observed effects. Future researchers might, therefore, want to include pre- and postmeasures of trained skills.

The participants in this study were self-selecting; therefore, a self-selection bias may have influenced our findings. In addition, the study was fully reliant on self-report, which may have also affected the results through both response bias [80] and shared variance [81].

Given these sample-specific and method-specific limitations, future researchers might want to investigate the effects of gamification using more diverse samples and additional measures of efficacy, such as sleep duration and quality (objective measures related to depression) and clinical interviews. In addition, the authors are aware that using the traditional "gold standard" of RCTs may not be the best method for evaluating the effectiveness of an app such as eQuoo. Such studies can take years from inception to publication, and developers will have real-time feedback and data that would influence app improvements more quickly than they could be tested [82]. The authors are investigating current research possibilities that would help address this issue and will be implementing them in the coming years while documenting the results for peer review.

Conclusions

This study aimed to investigate the effect of the gamified mobile mental health game eQuoo on levels of resilience, anxiety, depression, and attrition (one of digital mental health's Achilles heels [83]) in a student population. Compared with the active control and waitlist groups, we found a significant increase in resilience scores, a decrease in depression and anxiety scores, and a significantly lower attrition rate. The results suggest that eQuoo is an engaging and effective means to support students' mental health and build their resilience.

Acknowledgments

The study received no financial funding or sources of financial support externally or internally that might have influenced its outcome. S Litvin could not have undertaken this journey without the support and guidance of Professor Steve Pilling of the University College London, whose deep interest in student well-being made this trial possible. S Litvin is also thankful to Manuel Holihan, who helped shape this paper.

Conflicts of Interest

S Litvin is a shareholder of the company PsycApps Limited, which developed the intervention tested in this trial (eQuoo). S Lüttke, along with a group of other scientists, has received public funding for a project on an app-based aftercare for adolescents and young adults with depression. He has received consultancy fees from companies for offering advice on study design and interventions as well as payments for lectures from psychotherapy associations in the context of e–mental health matters.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3273 KB - mental_v10i1e47285_app1.pdf]

References

1. Ebert DD, Zarski AC, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015 Mar 18;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
2. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
3. Halldorsson B, Hill C, Waite P, Partridge K, Freeman D, Creswell C. Annual research review: immersive virtual reality and digital applied gaming interventions for the treatment of mental health problems in children and young people: the need for rigorous treatment development and clinical evaluation. J Child Psychol Psychiatry 2021 May;62(5):584-605. [doi: 10.1111/jcpp.13400] [Medline: 33655534]
4. Grist R, Croker A, Denne M, Stallard P. Technology delivered interventions for depression and anxiety in children and adolescents: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2019 Jun;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]
5. Ye X, Bapuji SB, Winters SE, Struthers A, Raynard M, Metge C, et al. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis. BMC Health Serv Res 2014 Jul 18;14:313 [FREE Full text] [doi: 10.1186/1472-6963-14-313] [Medline: 25037951]
6. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: a systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
7. Calear AL, Christensen H, Mackinnon A, Griffiths KM, O'Kearney R. The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009 Dec;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]
8. David OA, Cardo RA, Matu S. Changes in irrational beliefs are responsible for the efficacy of the REThink therapeutic game in preventing emotional disorders in children and adolescents: mechanisms of change analysis of a randomized clinical trial. Eur Child Adolesc Psychiatry 2019 Mar;28(3):307-318. [doi: 10.1007/s00787-018-1195-z] [Medline: 29987559]
9. Ebert D, Mortier P, Kaehlke F, Bruffaerts R, Baumeister H, Auerbach RP, WHO World Mental Health-International College Student Initiative collaborators. Barriers of mental health treatment utilization among first-year college students: first cross-national results from the WHO World Mental Health International College student initiative. Int J Methods Psychiatr Res 2019 Jun;28(2):e1782 [FREE Full text] [doi: 10.1002/mpr.1782] [Medline: 31069905]
10. Eisenberg D, Golberstein E, Gollust SE. Help-seeking and access to mental health care in a university student population. Med Care 2007 Jul;45(7):594-601. [doi: 10.1097/MLR.0b013e31803bb4c1] [Medline: 17571007]
11. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
12. Hidden waits force more than three quarters of mental health patients to seek help from emergency services. Royal College of Psychiatrists. 2022. URL: https://www.rcpsych.ac.uk/news-and-features/latest-news/detail/2022/10/10/hidden-waits-force-more-than-three-quarters-of-mental-health-patients-to-seek-help-from-emergency-services [accessed 2023-06-16]
13. Improving Access to Psychological Therapies (IAPT) waiting times guidance and FAQ's. National Health Service England. 2015. URL: https://www.england.nhs.uk/wp-content/uploads/2015/02/iapt-wait-times-guid.pdf [accessed 2023-06-28]

14. Pretorius C, Coyle D. Young people's use of digital tools to support their mental health during COVID-19 restrictions. Front Digit Health 2021 Dec 01;3:763876 [FREE Full text] [doi: 10.3389/fdgth.2021.763876] [Medline: 34927133]
15. Lupton D. Young people's use of digital health technologies in the global north: narrative review. J Med Internet Res 2021 Jan 11;23(1):e18286 [FREE Full text] [doi: 10.2196/18286] [Medline: 33427684]
16. Rich E, Miah A, Lupton D, Lewis SC. Digital health generation? Young people's use of 'healthy lifestyle' technologies. University of Bath. URL: https://www.researchgate.net/profile/Andy-Miah/publication/343335783_Digital_Health_Generation_Young_People's_Use_of_'Healthy_Lifestyle'_Technologies/links/5f23e5ac458515b729f5ed49/Digital-Health-Generation-Young-Peoples-Use-of-Healthy-Lifestyle-Technologies.pdf [accessed 2023-03-07]
17. The common sense census: media use by tweens and teens, 2021. Common Sense Media. 2019. URL: https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens-2021 [accessed 2023-03-07]
18. Beard C, Silverman AL, Forgeard M, Wilmer MT, Torous J, Björgvinsson T. Smartphone, social media, and mental health app use in an acute transdiagnostic psychiatric sample. JMIR Mhealth Uhealth 2019 Jun 07;7(6):e13364 [FREE Full text] [doi: 10.2196/13364] [Medline: 31199338]
19. Bennett S, Cuijpers P, Ebert DD, McKenzie Smith M, Coughtrey AE, Heyman I, et al. Practitioner review: unguided and guided self-help interventions for common mental health disorders in children and adolescents: a systematic review and meta-analysis. J Child Psychol Psychiatry 2019 Aug;60(8):828-847. [doi: 10.1111/jcpp.13010] [Medline: 30775782]
20. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215 [FREE Full text] [doi: 10.1016/j.invent.2014.08.003]
21. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
22. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
23. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
24. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
25. Goldberg SB, Lam SU, Simonsson O, Torous J, Sun S. Mobile phone-based interventions for mental health: a systematic meta-review of 14 meta-analyses of randomized controlled trials. PLOS Digit Health 2022;1(1):e0000002 [FREE Full text] [doi: 10.1371/journal.pdig.0000002] [Medline: 35224559]
26. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013 May;14(5):365-376. [doi: 10.1038/nrn3475] [Medline: 23571845]
27. Ioannidis JP. Why most discovered true associations are inflated. Epidemiology 2008 Sep;19(5):640-648. [doi: 10.1097/EDE.0b013e31818131e7] [Medline: 18633328]
28. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
29. Yang W, Ding Z, Dai T, Peng F, Zhang JX. Attention bias modification training in individuals with depressive symptoms: a randomized controlled trial. J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):101-111 [FREE Full text] [doi: 10.1016/j.jbtep.2014.08.005] [Medline: 25245928]
30. Kramer J, Conijn B, Oijevaar P, Riper H. Effectiveness of a web-based solution-focused brief chat treatment for depressed adolescents and young adults: randomized controlled trial. J Med Internet Res 2014 May 29;16(5):e141 [FREE Full text] [doi: 10.2196/jmir.3261] [Medline: 24874006]
31. Deterding S, Dixon DH, Khaled R, Nacke LE. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11; September 28-30, 2011; Tampere, Finland p. 9-15 URL: https://dl.acm.org/doi/10.1145/2181037.2181040 [doi: 10.1145/2181037.2181040]
32. Koivisto J, Hamari J. The rise of motivational information systems: a review of gamification research. Int J Inf Manage 2019 Apr;45:191-210 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2018.10.013]
33. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. Burlington, MA, USA: Morgan Kaufmann; Dec 2002:2.
34. Detoni AM, Lourenço JE, Pacheco LP, Gomes LA, de Brito TC, da Silveira Madalena T. Prevenção e intervenção em saúde mental: resiliência e psicologia positiva. ANALECTA-Centro Universitário Academia 2022;7(2) [FREE Full text]
35. Floryan MR, Ritterband LM, Chow PI. Principles of gamification for internet interventions. Transl Behav Med 2019 Nov 25;9(6):1131-1138. [doi: 10.1093/tbm/ibz041] [Medline: 30944933]
36. Nicholson S. A RECIPE for meaningful gamification. In: Reiners T, Wood LC, editors. Gamification in Education and Business. Cham, Switzerland: Springer; 2015:1-20.

37. Mental health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response [accessed 2022-06-17]
38. Wren-Lewis S, Alexandrova A. Mental health without well-being. J Med Philos 2021 Dec 02;46(6):684-703 [FREE Full text] [doi: 10.1093/jmp/jhab032] [Medline: 34668012]
39. Ungar M. Resilience, trauma, context, and culture. Trauma Violence Abuse 2013 Jul;14(3):255-266. [doi: 10.1177/1524838013487805] [Medline: 23645297]
40. Egeland B, Carlson E, Sroufe LA. Resilience as process. Dev Psychopathol 2009 Sep 09;5(4):517-528 [FREE Full text] [doi: 10.1017/s0954579400006131]
41. Ungar M, Theron L. Resilience and mental health: how multisystemic processes contribute to positive outcomes. Lancet Psychiatry 2020 May;7(5):441-448. [doi: 10.1016/S2215-0366(19)30434-1] [Medline: 31806473]
42. Bellis MA, Hughes K, Leckenby N, Perkins C, Lowey H. National household survey of adverse childhood experiences and their relationship with resilience to health-harming behaviors in England. BMC Med 2014 May 02;12:72 [FREE Full text] [doi: 10.1186/1741-7015-12-72] [Medline: 24886026]
43. Jefferies P, Vanstone R, Ungar M. The rugged resilience measure: development and preliminary validation of a brief measure of personal resilience. Appl Res Qual Life 2021 May 06;17(2):985-1000 [FREE Full text] [doi: 10.1007/s11482-021-09953-3]
44. Ungar M. The social ecology of resilience: addressing contextual and cultural ambiguity of a nascent construct. Am J Orthopsychiatry 2011 Jan;81(1):1-17. [doi: 10.1111/j.1939-0025.2010.01067.x] [Medline: 21219271]
45. Van Breda AD. A critical review of resilience theory and its relevance for social work. Soc Work 2018 Apr;54(1):1-18 [FREE Full text] [doi: 10.15270/54-1-611]
46. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
47. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
48. Random.org: True Random Number Service. RANDOM.ORG. URL: https://www.random.org/ [accessed 2023-06-28]
49. Haahr M. Introduction to randomness and random numbers. Random. 2023. URL: https://www.random.org/randomness/ [accessed 2023-06-16]
50. Litvin S. A review on digital mental health applications types, PsycappsE and eQuoo. Neurol Neurosurg 2019;2(1):1-4 [FREE Full text] [doi: 10.15761/nns.1000119]
51. Litvin S, Saunders R, Maier MA, Lüttke S. Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: a randomized controlled trial. PLoS One 2020 Sep 02;15(9):e0237220 [FREE Full text] [doi: 10.1371/journal.pone.0237220] [Medline: 32877425]
52. Cantallops MM, Sicilia MA. Motivations to read and learn in videogame lore: the case of league of legends. In: Proceedings of the 4th International Conference on Technological Ecosystems for Enhancing Multiculturality. 2016 Presented at: TEEM '16; November 2-4, 2016; Salamanca, Spain p. 585-591 URL: https://dl.acm.org/doi/10.1145/3012430.3012578 [doi: 10.1145/3012430.3012578]
53. Haller E, Watzke B. The role of homework engagement, homework-related therapist behaviors, and their association with depressive symptoms in telephone-based CBT for depression. Cognit Ther Res 2021;45(1):224-235 [FREE Full text] [doi: 10.1007/s10608-020-10136-x]
54. Cheng VW, Davenport T, Johnson D, Vella K, Hickie IB. Gamification in apps and technologies for improving mental health and well-being: systematic review. JMIR Ment Health 2019 Jun 26;6(6):e13717 [FREE Full text] [doi: 10.2196/13717] [Medline: 31244479]
55. Rajaei A, Daneshpour M, Robertson J. The effectiveness of couples therapy based on the Gottman method among Iranian couples with conflicts: a quasi-experimental study. J Couple Relatsh Ther 2019;18(3):223-240 [FREE Full text] [doi: 10.1080/15332691.2019.1567174]
56. Laufer O, Israeli D, Paz R. Behavioral and neural mechanisms of overgeneralization in anxiety. Curr Biol 2016 Mar 21;26(6):713-722 [FREE Full text] [doi: 10.1016/j.cub.2016.01.023] [Medline: 26948881]
57. Sullivan MJ, D'Eon JL. Relation between catastrophizing and depression in chronic pain patients. J Abnorm Psychol 1990 Aug;99(3):260-263. [doi: 10.1037//0021-843x.99.3.260] [Medline: 2145334]
58. Syafryadin S, Nurkamto J, Linggar DA, Mujiyanto J. Active, belief, and consequence (ABC) model technique in minimizing students' speech anxiety. In: Proceedings of the 5th International Seminar on English Language and Teaching. 2017 Presented at: ISELT-5 '17; August 17, 2017; Sumatra, Indonesia p. 320-324 URL: https://ejournal.unp.ac.id/index.php/selt/article/viewFile/8021/6123
59. Petrinec A, Wilk C, Hughes JW, Zullo MD, Chen YJ, Palmieri PA. Delivering cognitive behavioral therapy for post-intensive care syndrome-family via a mobile health app. Am J Crit Care 2021 Nov 01;30(6):451-458. [doi: 10.4037/ajcc2021962] [Medline: 34719716]
60. Kaltenthaler E, Parry G, Beverley C, Ferriter M. Computerised cognitive-behavioural therapy for depression: systematic review. Br J Psychiatry 2008 Sep;193(3):181-184. [doi: 10.1192/bjp.bp.106.025981] [Medline: 18757972]

61. Hedman E, Furmark T, Carlbring P, Ljótsson B, Rück C, Lindefors N, et al. A 5-Year follow-up of internet-based cognitive behavior therapy for social anxiety disorder. J Med Internet Res 2011 Jun 15;13(2):e39 [FREE Full text] [doi: 10.2196/jmir.1776] [Medline: 21676694]
62. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
63. Litvin S. Gamification as an approach to improve mental well-being and reduce attrition in mobile mental health interventions: a randomized controlled trial. Open Science Framework. 2020. URL: https://dx.plos.org/10.1371/journal.pone.0237220 [accessed 2023-06-06]
64. Blanco C, Okuda M, Wright C, Hasin DS, Grant BF, Liu SM, et al. Mental health of college students and their non-college-attending peers: results from the national epidemiologic study on alcohol and related conditions. Arch Gen Psychiatry 2008 Dec;65(12):1429-1437 [FREE Full text] [doi: 10.1001/archpsyc.65.12.1429] [Medline: 19047530]
65. Barnett P, Arundell LL, Saunders R, Matthews H, Pilling S. The efficacy of psychological interventions for the prevention and treatment of mental health disorders in university students: a systematic review and meta-analysis. J Affect Disord 2021 Feb 01;280(Pt A):381-406. [doi: 10.1016/j.jad.2020.10.060] [Medline: 33227669]
66. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
67. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
68. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
69. Greenwood K, Alford K, O'Leary I, Peters E, Hardy A, Cavanagh K, et al. The U and I study: study protocol for a feasibility randomised controlled trial of a pre-cognitive behavioural therapy digital 'informed choice' intervention to improve attitudes towards uptake and implementation of CBT for psychosis. Trials 2018 Nov 20;19(1):644 [FREE Full text] [doi: 10.1186/s13063-018-3023-7] [Medline: 30458850]
70. Lippke S, Gao L, Keller FM, Becker P, Dahmen A. Adherence with online therapy vs face-to-face therapy and with online therapy vs care as usual: secondary analysis of two randomized controlled trials. J Med Internet Res 2021 Nov 03;23(11):e31274 [FREE Full text] [doi: 10.2196/31274] [Medline: 34730541]
71. Titov N, Staples L, Kayrouz R, Cross S, Karin E, Ryan K, et al. Rapid report: early demand, profiles and concerns of mental health users during the coronavirus (COVID-19) pandemic. Internet Interv 2020 Sep;21:100327 [FREE Full text] [doi: 10.1016/j.invent.2020.100327] [Medline: 32537424]
72. Zhu J, Dallal DH, Gray RC, Villareale J, Ontañón S, Forman EM, et al. Personalization paradox in behavior change apps: lessons from a social comparison-based personalized app for physical activity. Proc ACM Hum Comput Interact 2021 Apr 22;5(CSCW1):1-21 [FREE Full text] [doi: 10.1145/3449190]
73. Grua EM, Hoogendoorn M. Exploring clustering techniques for effective reinforcement learning based personalization for health and wellbeing. In: Proceedings of the 2018 Symposium Series on Computational Intelligence. 2018 Presented at: SSCI '18; November 18-21, 2018; Bangalore, India p. 813-820 URL: https://ieeexplore.ieee.org/document/8628621 [doi: 10.1109/ssci.2018.8628621]
74. Alsawaier RS. The effect of gamification on motivation and engagement. Int J Inf Learn Technol 2018 Jan 02;35(1):56-79 [FREE Full text] [doi: 10.1108/ijilt-02-2017-0009]
75. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The role of telehealth in reducing the mental health burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
76. Davydov DM, Stewart R, Ritchie K, Chaudieu I. Resilience and mental health. Clin Psychol Rev 2010 Jul;30(5):479-495 [FREE Full text] [doi: 10.1016/j.cpr.2010.03.003] [Medline: 20395025]
77. Sarkar M, Fletcher D. How resilience training can enhance wellbeing and performance. In: Crane MF, editor. Managing for Resilience: A Practical Guide for Employee Wellbeing and Organizational Performance. London, UK: Routledge; 2017:227-237.
78. Arnett JJ, Žukauskienė R, Sugimura K. The new life stage of emerging adulthood at ages 18-29 years: implications for mental health. Lancet Psychiatry 2014 Dec;1(7):569-576. [doi: 10.1016/S2215-0366(14)00080-7] [Medline: 26361316]
79. Baskin TW, Tierney SC, Minami T, Wampold BE. Establishing specificity in psychotherapy: a meta-analysis of structural equivalence of placebo controls. J Consult Clin Psychol 2003 Dec;71(6):973-979. [doi: 10.1037/0022-006X.71.6.973] [Medline: 14622072]
80. Furnham A. Response bias, social desirability and dissimulation. Pers Individ Dif 1986;7(3):385-400 [FREE Full text] [doi: 10.1016/0191-8869(86)90014-0]
81. Kormos C, Gifford R. The validity of self-report measures of proenvironmental behavior: a meta-analytic review. J Environ Psychol 2014 Dec;40:359-371 [FREE Full text] [doi: 10.1016/j.jenvp.2014.09.003]

82. Marshall JM, Dunstan DA, Bartik W. Smartphone psychology: new approaches towards safe and efficacious mobile mental health apps. Prof Psychol Res Pr 2020 Jun;51(3):214-222. [doi: 10.1037/pro0000278]
83. Nwosu A, Boardman S, Husain MM, Doraiswamy PM. Digital therapeutics for mental health: is attrition the Achilles heel? Front Psychiatry 2022 Aug 05;13:900615 [FREE Full text] [doi: 10.3389/fpsyt.2022.900615] [Medline: 35982936]

Abbreviations

CBT: cognitive behavioral therapy
DHI: digital health intervention
GAD-7: generalized anxiety disorder–7
OR: odds ratio
PHQ-8: Patient Health Questionnaire–8
RCT: randomized controlled trial
RM: repeated measures
RRM: Rugged Resilience Measure

Edited by J Torous; submitted 14.03.23; peer-reviewed by S Dix, J Marshall; comments to author 06.04.23; revised version received 02.06.23; accepted 20.06.23; published 21.07.23.

Please cite as:
Litvin S, Saunders R, Jefferies P, Seely H, Pössel P, Lüttke S
The Impact of a Gamified Mobile Mental Health App (eQuoo) on Resilience and Mental Health in a Student Population: Large-Scale Randomized Controlled Trial
JMIR Ment Health 2023;10:e47285
URL: https://mental.jmir.org/2023/1/e47285
doi:10.2196/47285
PMID:37477955

Original Paper

Suicide Prevention Using Google Ads: Randomized Controlled Trial Measuring Engagement

Sandersan Onie[1], PhD; Patrick Berlinquette[2]; Sarah Holland[1]; Nicola Livingstone[1]; Coco Finemore[1]; Nyree Gale[1], BPsych (Hons), MPsych (Clin); Emma Elder[1]; George Laggis[1]; Cassandra Heffernan[1]; Susanne Oliver Armstrong[1], MPhil, PhD; Adam Theobald[1], BPsych (Hons), MClinPsych; Natasha Josifovski[1], PhD; Michelle Torok[1], PhD; Fiona Shand[1], PhD; Mark Larsen[1], PhD

[1]Black Dog Institute, Randwick, Australia
[2]BERLIN SEM, New York, NY, United States

Corresponding Author:
Sandersan Onie, PhD
Black Dog Institute
1 Hospital Rd
Randwick, 2031
Australia
Phone: 61 (02) 9382 4530
Email: s.onie@blackdog.org.au

Abstract

Background: Studies have shown that individuals may search for suicide-related terms on the internet prior to an attempt.

Objective: Thus, across 2 studies, we investigated engagement with an advertisement campaign designed to reach individuals contemplating suicide.

Methods: First, we designed the campaign to focus on crisis, running a campaign for 16 days in which crisis-related keywords would trigger an ad and landing page to help individuals find the national suicide hotline number. Second, we expanded the campaign to also help individuals contemplating suicide, running the campaign for 19 days with a wider range of keywords through a co-designed website with a wider range of offerings (eg, lived experience stories).

Results: In the first study, the ad was shown 16,505 times and was clicked 664 times (4.02% click rate). There were 101 calls to the hotline. In the second study, the ad was shown 120,881 times and clicked 6227 times (5.15% click rate); of these 6227 clicks, there were 1419 (22.79%) engagements with the site, a substantially higher rate than the industry average of 3%. The number of clicks on the ad was high despite a suicide hotline banner likely being present.

Conclusions: Search advertisements are a quick, far-reaching, and cost-efficient way of reaching those contemplating suicide and are needed despite suicide hotline banners being present.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12623000084684; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=385209

(JMIR Ment Health 2023;10:e42316) doi:10.2196/42316

KEYWORDS

suicide prevention; suicide; suicidal; self harm; digital advertising; Google Ads; search; suicide hotline; advertise; advertising; campaign; mental health; prevention; digital intervention; online intervention

Introduction

Overview

What Is Already Known on This Topic

Individuals may search for suicide-related terms on the internet prior to an attempt. These search terms may reflect their current cognitive state pertaining to suicide. However, there is currently no intervention targeting search pages.

What This Study Adds

This study provides evidence for the high level of reach and engagement of a Google AdWords campaign despite a hotline banner likely being present on the search page. Further, this

study provides evidence that this intervention is effective in promoting help seeking, with study 1 showing a high number of people calling the national hotline from the landing page. This study also provides information on the feasibility, reach, speed, and cost of this type of intervention.

How This Study Might Affect Research, Practice, or Policy

Although a hotline banner appears when suicide terms are searched, the results suggest that this is not enough as many individuals still engage with the advertisement below the hotline banner. Internet ads may be a rapid, far-reaching, and cost-effective way to reach out to individuals for suicide and across a range of health issues.

Background

Previous studies have shown that individuals may search for suicide-related terms on the internet prior to an attempt. A recent study showed that in the 60 days prior to a suicide attempt, individuals had searched for terms expressing suicidal ideation and suicide means but also relevant keywords not directly associated with crisis (eg, feeling empty, divorce, and alcohol use) [1]. In addition, studies have found that the increases in the volume of internet searches for suicide-related terms predicted subsequent increases in national suicide rates [2-4]. Thus, targeting internet search engines to intervene in such searches, and connect individuals to help, may be an important avenue for intervention.

Internet search ads can be used to reach out to individuals contemplating suicide, or in crisis, by identifying what keywords people are likely to search, presenting an advertisement on the search page when the keywords are searched, and linking to a landing page with appropriate resources and help-seeking information. The reach and effectiveness of these campaigns can be assessed by measuring the impressions (how many times the advertisement is shown), clicks or click rate (clicks on the advertisement), and conversions (specific behaviors performed on the website) [5]. Note that the party placing the advertisements must determine what behaviors they would like visitors to do on the website, counting these behaviors as conversions.

One reason to use advertisements over organic (or nonadvertisement) searches is that when using organic search, it is not guaranteed that a link will appear first on the search result page. A study found that on average, a link on the search page is twice as likely to be clicked on versus the link directly below it [6], suggesting that individuals are most likely to click and engage with the search result presented first. Further, organic search relies on similarities of the keyword and the page itself; thus, a person who is searching for terms pertaining to "loneliness" may not be shown a link to a page with suicide help. Finally, internet searches are a part of daily life, with 93% of browsing sessions starting on the search page [7]; thus, individuals may be more likely to engage with help if the process by which they access help is a part of their day-to-day behavior.

In some search engines, a hotline number will appear in a prominently displayed banner at the top of the search results if certain suicide-related terms are searched. Reports from the United States have shown that in certain instances, the banners have increased calls to the hotline by 10% [8]. However, there is no public information on what keywords will trigger the banner and whether past search history increases or decreases the likelihood of this banner appearing. Further, a Google spokesperson has said and past research has shown that not all relevant keywords will trigger this hotline banner [8,9]. Research has also shown that not everyone in a suicidal crisis would like to call a hotline, suggesting a need for more diverse offerings [10]. Further, a phenomenon called "banner blindness" is often observed, in which eye-tracking data suggest that individuals will tend to ignore elements on a page in the form of a banner [11].

Objectives

The overall objective of this research was to investigate patterns of engagement with advertisements and landing pages—that is, the first page shown after a link is clicked—designed for those in crisis and contemplating suicide, which was examined in 2 studies. The first study aimed to assess whether the presence of a Google Ads campaign promoting a crisis line would encourage connection with the crisis service. The aim of the second study was to conduct a 2-arm trial investigating the additional benefits of age tailoring while extending the campaign to include individuals at the precrisis or contemplation phase and assess the different engagement levels of different types of search keywords. Both studies are reported in CONSORT (Consolidated Standards of Reporting Trials)–compliant formats.

Study 1

Methods

Internet Advertisement Platform

For this study, we used Google Ads, Google's proprietary search advertisement platform, to promote a crisis line to individuals searching for specific suicide-related terms. Google Ads was chosen as the search engine with the largest market share [12] and thus offers maximum reach. Ads are charged only if the advertisement is clicked, not just for displaying the advertisement. Cost per conversion is the average amount of money spent for a single conversion; there is no further charge per conversion.

Google Ads has several ethical protections in place. The data are only available in aggregate format, and it is not possible to identify individual people who have seen the ad, clicked on the ad, or browsed the landing page website. Although Google Ads may use past browsing history to present advertisements that the individual is likely to engage with, the user can erase all data Google has collected [13] or opt out of personalized advertising [14]. Furthermore, personal information, such as email, is never collected or shared without express permission [15].

Ethical Considerations

Ethics approval was not required for this study as only deidentified, pre-existing data are reported in aggregate.

Participants

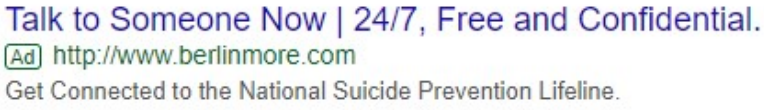
Individuals residing in the United States were eligible to participate.

Intervention

An ad campaign was configured to run across the United States for a total of 16 days in 2019 (from October 1 to 3 and from November 15 to 27). The keywords used in this study consisted of 28 commonly searched suicide-related terms that were compiled by a Google Ads agent, an individual who coordinates and runs Google Ads campaigns for organizations. Of the 28 terms, 13 keywords indicated overt suicidality; 7 keywords featured a method or location; 4 pertained to help seeking; and 4 were related to, but not explicitly mentioning, suicide. Note that Google Ads will trigger the ad when a keyword or combination of keywords are semantically similar to the keywords; thus, a single keyword such as “kill myself” may capture a wide range of terms (eg, “end my life” and “take my life”), and a short list may cover a much wider range of keywords. Figure 1 shows the advertisement triggered during the campaign.

If an individual clicked on the link in the ad, it would lead them to a simple landing page consisting of a link to the National Suicide Prevention Hotline for the United States. The “conversion” for this campaign (ie, the desired behavior after an individual had clicked through from the ad to the landing page) was calling the hotline number.

Figure 1. The advertisement that is triggered during the Google Ads campaign in study 1.



Outcomes

Data on impressions, clicks, click rate (clicks/impressions), conversions, conversion rate (conversion/click), cost per click, and cost per conversion were extracted from Google Ads. Total conversion rate was manually calculated (conversions/impressions). The primary outcomes for this trial to measure engagement were click rate (engagement with the ad), conversion rate (engagement with the landing page), and total conversion rate (total engagement with campaign).

Statistical Analysis

For any analyses in which we compare rates (eg, click-through rates or conversion rates), we used the MedCalc software (MedCalc Software Ltd), which uses a chi-square test to test a significant incidence rate difference (IRD) [16], set to a significance rate of .05. All rate comparisons in this paper use this method.

Results

Within the 16-day duration of the ad campaign, the advertisement was shown 16,505 times and was clicked 664 times, yielding a 4.02% click rate. From those who clicked the ad (n=664), there were 101 calls to the hotline, yielding a 15.21% conversion rate with an average cost of US $13.57 per helpline call. The campaign had an overall conversion rate (conversion/impressions) of 0.61% (101/16,505).

To investigate how our campaign performed against industry standards, we conducted a rate comparison test to compare our campaign data versus publicly available industry data. The industry data were drawn from approximately 985,804,416 impressions, 31,250,000 clicks, and 1,171,875 conversions, with a 3.17% click rate and 3.75% conversion rate, collected from the Google Ads agency [17]. The analysis suggested that our campaign overall performed better than the industry standard (IRD 0.004931, 95% CI 0.004405-0.0005457; P<.001; ie, 0.49 percentage points higher than the industry standard;).

Discussion

In study 1, we investigated the efficacy of a Google Ads intervention in encouraging help seeking by calling a helpline by running a campaign across the United States. The results revealed high click-through rates and conversion rates when compared with the industry average across all sectors (click rate: 4.02% vs 3.17% industry average; conversion rate: 15.21% vs 3.75% industry average; overall conversion rate: 0.61% vs. 0.12% industry average), having a 5.08 times greater total conversion rate than the industry average.

As information about when the banner was or was not triggered is limited, it is not possible to provide a further breakdown of the ad engagement based on whether the banner was triggered. Nevertheless, as good engagement was observed with an ad that promoted the national hotline, 2 general conclusions can be drawn. First, if the hotline banner *was* triggered, this suggests that some individuals skipped over the banner but did click on the ad to call the hotline. Second, if the banner was *not* triggered, then the ad led to an individual seeking help that otherwise would not have. Overall, the data provide evidence for the utility of an intervention using Google Ads.

Study 2

Background and Objectives

In study 2, we developed a Google Ads campaign designed to be helpful for individuals who may be contemplating suicide but are not in immediate crisis [18]. In this context, we use the term contemplation phase to capture any stages that may lead to a suicide attempt but prior to immediate crisis. Simply providing a link to the hotline may not be the most appropriate response in the contemplation stage, and therefore, we undertook a co-design process to expand what was offered on the landing page. Further, given that the keywords searched may represent different cognitive states, we investigated whether different categories of keywords led to different levels of engagement. Finally, we used Google Ads’ targeting features and investigated

whether individuals show higher engagement with landing pages tailored to specific age groups compared to a general all-ages landing page.

Methods

Trial Design

The study used a 2-arm experimental design (landing page: general vs age-tailored—18-24, 25-44, or 45+ years) with 4 initial pathways (individuals searching for different types of keywords: low risk, high risk, help seeking, and means specific). Participants were allocated equally to the 2 arms.

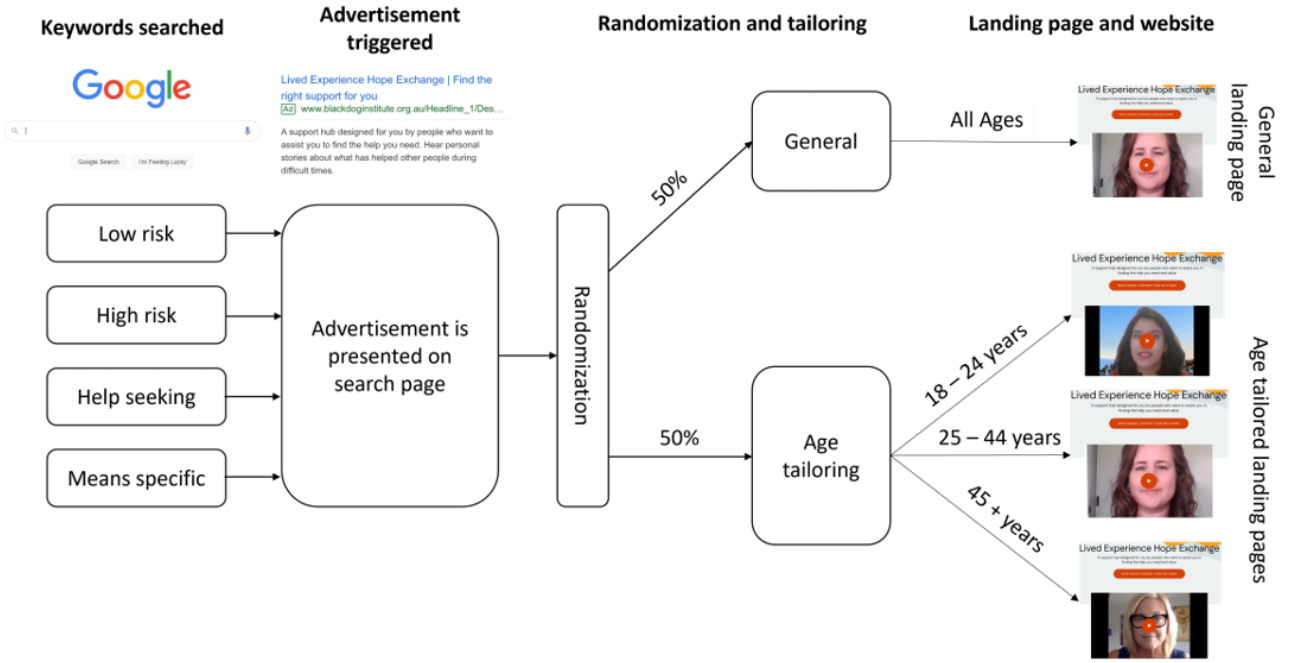
Participants

Participants aged >18 years, whose ages could be inferred by Google from their past browsing history and are currently residing in Australia, were eligible to join the trial.

Intervention

As with study 1, data were downloaded in an aggregate, deidentified form. The campaign was launched on March 2, 2022, and ran until the prespecified budget was exhausted (March 21, 2022). Full details of the keyword generation, advertisement and landing page co-design process, and content of the landing pages and linked pages can be found elsewhere [10]. A schematic of the campaign can be found in Figure 2.

Figure 2. A schematic of the Google AdWords campaign used in this study. As individuals would search for a keyword listed in 1 of 4 keyword lists (low risk, high risk, help seeking or means specific), the advertisement would trigger on the Google search page. If clicked, individuals would be randomly allocated into 1 of 2 study conditions, that is, general or age-tailored landing pages. If in the general condition, individuals would then be presented with the same landing page, whereas if in the age-tailored condition, individuals would be presented with a landing page tailored for their age group.



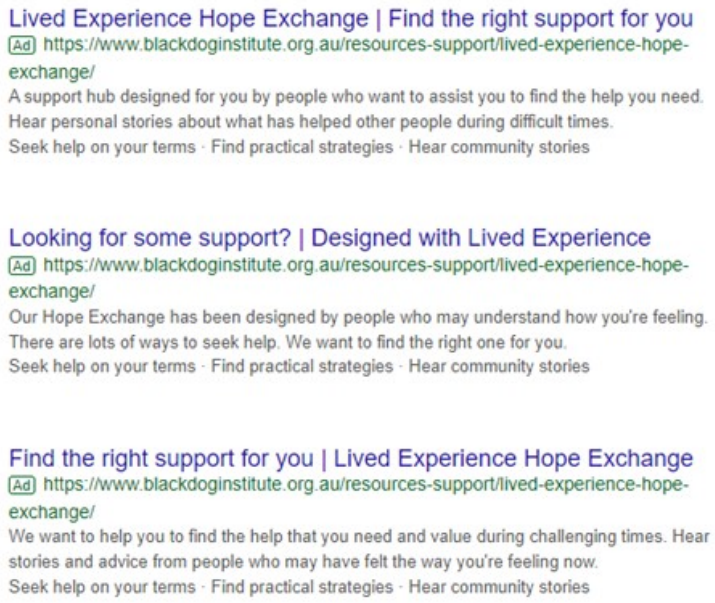
Keywords

Together with lived experience advisors, researchers, and a Google Ads agent, we generated 4 lists of keywords: *low risk* keywords, which included keywords people are likely to search when in distress or situations associated with suicide, without explicitly mentioning suicide (eg, “Feeling so alone” and “debt”); *high risk* keywords, which included keywords explicitly communicating suicidal ideation or intent (eg, “I want to die”); *help seeking* keywords, which included keywords explicitly searching for help for suicidal thoughts (eg, “suicide help”); and *means specific* keywords, which relate to searching or using specific means.

Advertisements

Ads were shown to users independently of which category of keywords were searched. The co-design process yielded 3 similar advertisements. When the advertisement triggered, 1 of 3 advertisements would be randomly shown, resulting in equal presentations across the study. The 3 advertisements are shown in Figure 3.

Figure 3. Advertisements shown on the Google search page if the advertisement is triggered in study 2.



Landing Page

In collaboration with lived experience advisors, we co-designed a series of landing pages with a primary focus on the contemplation stage, rather than crisis, which was entitled the "Lived Experience Hope Exchange." The pages contained lived experience stories, calming and distracting activities, and links to support services and hotlines with descriptions of what the individual is likely to experience when engaging these services. In all, 4 different versions of the Hope Exchange were developed: a general version and 3 age-tailored versions (18-24, 25-44, and 45+ years). Details of the pages can be found elsewhere [10].

Randomization

Google Experiments, Google's A/B testing feature, was used to randomize individuals clicking the ad link to either the general version of the Hope Exchange (50%) or the versions tailored to their estimated age (50%). We excluded any individual whose age could not be determined or who were aged <18 years. This randomization was implemented using Google Experiments; both participants and researchers were blinded.

Outcomes

The primary outcomes for study 2 were identical to study 1. However, given the much wider range of resources on the website compared to study 1, we included a wider range of conversions (drawing from engagement metrics in the advertising field) to include behaviors that the investigators, lived experience advisors, and collaborative team considered positive. Triggering any of these conditions was considered a conversion, including:

- Clicking the "get help" button to see available support services
- Downloading any file
- Clicking on a link to call a support service
- Spending more than 2 minutes on the website

Statistical Analysis

Data collection and analysis methods for study 2 were the same as for Study 1. Primary analyses included comparing the total conversion rates of studies 1 and 2 and the industry average; click rate by keywords searched; conversions by keywords searched; and conversions by tailoring condition.

Ethics Approval

This study was approved by the University of New South Wales Human Research Ethics Committee (HC210827), including a formal waiver of consent as it was not possible to obtain consent prior to participants searching for relevant keywords or clicking through on the displayed ads. The project did not meet the committee's definition of a clinical trial, as it did not evaluate the effect of an intervention on health outcomes; as such, the study was not prospectively registered with the Australian New Zealand Clinical Trials Registry (ANZCTR).

Results

The advertisements ran from March 2 to 21, 2022, in Australia with a total of 120,881 impressions, 6227 clicks (a 5.15% click-through rate), and 1419 conversions, with a conversion rate of 22.79% (1419/6227) and a total conversion rate of 1.17% (1419/120,881). The breakdown per keyword group is shown in Table 1. Note that due to the low number of the clicks related to the means-related keywords (n=4, with no conversions), these keywords were excluded from the analysis.

We investigated whether the study-2 total conversion rate was higher than those of study 1 or the industry standard. The results revealed that the study-2 campaign had a higher total conversion rate than the study-1 total conversion rate (IRD 0.005619, 95% CI 0.004984-0.007436; *P*<.001) and industry standard total conversion rate (IRD 0.01055, 95% CI 0.01036-0.01191; *P*<.001), with the study-2 total conversion rate being 1.91 times greater than the total conversion rate in study 1 and 9.75 times greater than the industry standard.

Table 1. Metrics from the study-2 Google Ads campaign.

Metric, campaign	Keyword type			
	Low risk	High risk	Help seeking	Means specific
Impressions, n				
Total	108,219	3369	9079	223
General	45,646	1800	5279	78
Tailored (all)	62,573	1569	3800	145
Tailored (18-24 years)	8837	311	683	26
Tailored (25-44 years)	25,429	577	1274	72
Tailored (45+ years)	28,307	681	1843	47
Clicks, n				
Total	5530	118	575	4
General	2440	56	340	2
Tailored (all)	3093	62	235	2
Tailored (18-24 years)	514	7	25	0
Tailored (25-44 years)	1032	23	66	1
Tailored (45+ years)	1547	32	144	1
Click rate, n/N (%)				
Total	5530/108,219 (5.11)	118/3369 (3.5)	575/9079 (6.33)	4/223 (1.79)
General	2440/45,646 (5.35)	56/1800 (3.11)	340/5279 (6.44)	2/78 (2.56)
Tailored (all)	3093/62,573 (4.94)	62/1569 (3.95)	235/3800 (6.18)	2/145 (1.38)
Tailored (18-24 years)	514/8837 (5.82)	7/311 (2.25)	25/683 (3.66)	0/26 (0)
Tailored (25-44 years)	1032/25,429 (4.05)	23/577 (3.99)	66/1274 (5.18)	1/72 (1.39)
Tailored (45+ years)	1547/28,307 (5.46)	32/681 (4.7)	144/1843 (7.81)	1/47 (2.13)
Conversions, n				
Total	1209	37	171	0
General	583	20	104	0
Tailored (all)	626	17	67	0
Tailored (18-24 years)	47	3	11	0
Tailored (25-44 years)	185	7	19	0
Tailored (45+ years)	394	7	37	0
Conversion rate, n/N (%)				
Total	1209/5530 (21.86)	37/118 (31.36)	171/575 (29.74)	0/4 (0)
General	583/2440 (23.9)	20/56 (35.71)	104/340 (28.51)	0/2 (0)
Tailored (all)	626/3093 (20.25)	17/62 (27.42)	67/235 (30.59)	0/2 (0)
Tailored (18-24 years)	47/514 (9.14)	3/7 (42.86)	11/25 (44)	0/0 (0)
Tailored (25-44 years)	185/1032 (17.94)	7/23 (30.43)	19/66 (28.79)	0/1 (0)
Tailored (45+ years)	394/1547 (25.49)	7/32 (21.88)	37/144 (25.69)	0/1 (0)
Total conversion rate, n/N (%)				
Total	1209/108,219 (1.12)	37/3369 (1.1)	171/9079 (1.88)	0/223 (0)
General	583/45,646 (1.28)	20/1800 (1.11)	104/5279 (1.97)	0/78 (0)
Tailored (all)	626/62,573 (1)	17/1569 (1.08)	67/3800 (1.76)	0/145 (0)
Tailored (18-24 years)	47/8837 (0.53)	3/311 (0.96)	11/683 (1.61)	0/26 (0)
Tailored (25-44 years)	185/25,429 (0.73)	7/577 (1.21)	19/1274 (1.49)	0/72 (0)

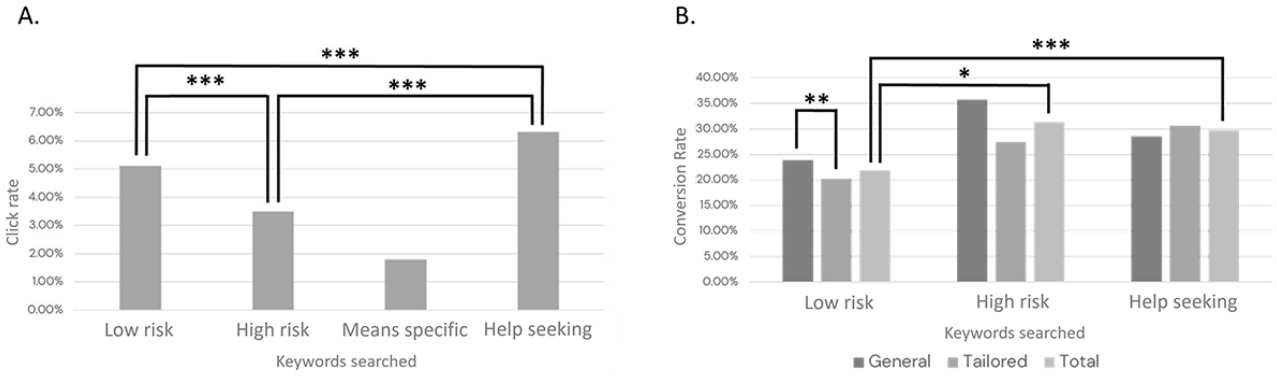
Metric, campaign	Keyword type			
	Low risk	High risk	Help seeking	Means specific
Tailored (45+ years)	394/28,307 (1.39)	7/681 (1.03)	37/1843 (2.01)	0/47 (0)

We then investigated whether the click-through rate differed as a function of keywords searched. The results revealed that the click rates for individuals searching *low risk* keywords, *high risk* keywords, and *help seeking* keywords were all significantly different from one another, in which individuals searching for *help seeking* keywords had the highest click rate, followed by individuals searching for *low risk* keywords, followed by individuals searching for *high risk* keywords. Results are shown in Figure 4A (low risk vs high risk: IRD 0.01608, 95% CI 0.00836-0.04194; *P*<.001; low risk vs help seeking: IRD 0.01223, 95% CI 0.001711-0.00734; *P*<.001; help seeking vs high risk: IRD 0.02831, 95% CI 0.01898-0.03764; *P*<.001).

We then investigated whether conversion rates differed by keyword type. The rate comparison analysis found a significant difference between *low risk* and *high risk* keywords (IRD 0.6972, 95% CI 0.00929-0.18058; *P*=.009), in which there was a higher conversion rate for people who searched for *high risk* keywords, and a significant difference between *low risk* and *help seeking* keywords (IRD 0.07877, 95% CI 0.03793-0.1196; *P*<.001), in which there was a higher conversion rate for individuals who are seeking help. However, there was not enough evidence to suggest that there was a difference in conversion rates between *high risk* and *help seeking* keywords (IRD 0.01617, 95% CI 0.09235-0.12469; *P*=.77). A graphical representation of these results can be found in Figure 4B.

Finally, we investigated whether age tailoring had a significant effect on conversions. The analysis revealed a significant difference between tailored and general landing pages for *low risk* keywords, in which general landing pages had a higher conversion rate than tailored landing pages (IRD 0.03635, 95% CI 0.06116-0.01153; *P*=.004). However, there was no significant difference for the general versus tailored landing pages for *high risk* (IRD 0.08295, 95% CI –0.11938 to 0.28528; *P*=.42) and *help seeking* keywords (IRD 0.02078, 95% CI 0.11145-0.0699; *P*=.65; see Figure 4B).

Figure 4. (A) Click through-rate by keywords searched. The x-axis shows the group of keywords searched. (B) Conversion rate by keywords searched. The x-axis shows the group of keywords searched, whereas different column shades indicate the trial tailoring condition. Total column is calculated using the following formula: (general click count + tailored click count) / (general conversion count + tailored conversion count). * Indicates *P*<.05, ** indicates *P*<.01, and *** indicates *P*<.001.



Discussion

In this study, we codeveloped a series of search advertisement campaigns targeting individuals contemplating suicide and evaluated them for levels of reach and engagement. As per study 1, we observed high engagement relative to the industry standards. Within 21 days, we were able to reach individuals searching for suicide-related terms 120,881 times, with an average cost of US $13 each time a person engaged with a behavior designed to help them. Note that cost per conversion does not differ from one conversion to another; thus, it is the role of the Google Ads client—in this case, the research team—to set conversions that are meaningful.

Given that the advertisement was designed for individuals contemplating suicide, it is consistent with the finding that there was a significantly higher click-through rate for individuals searching for *low risk* than *high risk* keywords. The *help seeking* keywords yielded the highest click-through rate, which is consistent with the idea that individuals responded to the ad that featured words communicating explicitly that help could be found. This is consistent with the notion that the better the advertisement matches the search term, the more likely an individual is to click the advertisement. However, the analysis found a higher conversion rate for *high risk* compared to *low risk* keywords, which may stem from the fact that not all individuals searching for *low risk* keywords are searching for or need immediate help.

One finding was that there were markedly lower impressions, clicks, and conversions for individuals searching for *means specific* keywords, suggesting that overall, there were fewer people who were searching for *means specific* keywords. Given the low number of impressions and clicks, we were unlikely to see any conversions.

An unexpected finding was that there was a higher conversion rate for general landing pages compared to tailored landing

pages for *low risk* keywords, suggesting that our tailoring was not effective in application and rather reduced engagement. This may indicate that despite there being clear, mutually exclusive preferences indicated by different age groups in the co-design process [10], these preferences may have limited generalizability or how we operationalized these preferences may be limited despite having approval from the co-design team. Other types of tailoring have yet to be explored.

General Discussion

Principal Findings

Across 2 studies, we investigated whether using Google Ads would allow us to reach individual searching for suicide-related terms for both crisis, in which we presented a suicide helpline, and contemplation, in which we presented a landing page offering help beyond presenting a helpline, such as hearing lived experience stories and calming exercises. Across both studies, we have observed exceptionally high click-through and conversion rates compared to the industry standard, with study-2 total conversion rate being 9.75 times higher than the industry standard. Overall, the data show that individuals are engaging well with both the advertisement and the pages themselves.

Given that for much of the time, it is likely that the hotline banners were present in both studies, future studies should investigate why individuals were still engaging with a search result if the helpline number was present on the search page. One possibility in study 1 is that within the few lines of text, the advertisement was able to provide more information on the hotline compared to the banner—that is, the hotline is anonymous and available 24/7. Lived experience advisors noted that it would be helpful to include descriptions on what it is like to call a hotline or engage in other services to help remove apprehension and promote help seeking [10]. In the second study, the *help seeking* keywords yielded the highest click-through rate—keywords that are likely to trigger a hotline banner given they explicitly indicate help seeking for suicide—further suggesting that there are individuals who are seeking help for suicide who do not want to immediately call a hotline, which is consistent with findings from the co-design study [10].

Future studies should investigate whether other advertising services can be used for suicide prevention. Although Google has a majority market share [12] for search engine advertisements, the advertising industry spans across multiple platforms. For example, there is data collection and advertising on social media (such as Twitter, Instagram, and TikTok) and streaming sites (eg, YouTube) and on advertising banners on websites (eg, Google AdSense). Future studies should investigate whether the findings from this study generalize to other platforms and whether using more than one platform for data collection and advertising increases reach and promotes help seeking. We propose, given the wealth of research showing web-based behaviors reflecting suicidality outside of search (eg, social media [19]), that we should see similarly fruitful findings outside of search engine advertisements.

Future work should also investigate how to integrate digital advertising such as Google Ads into routine practice. For example, given that individuals may discover existing resources through internet searches, partnerships with local health providers could help to place ads to link individuals to their local area's most validated and well-resourced services. To implement this into routine practice, we would need to ascertain (1) the cost and cost-effectiveness of integrating digital advertising into existing health systems, (2) what services could be highlight on the landing pages, and (3) the funding requirements to sustain this approach.

Limitations

Our investigation has several limitations. One limitation is that we did not record data on whether persons clicking on the ads were experiencing suicidal ideation or had engaged in suicidal behaviors, nor did we assess change in these outcomes. Future studies could investigate whether help-promoting web pages, such as those designed for this study, increase the number of individuals who seek help and reduce the suicide rate within a specific geographic region using Google Ads' geographic targeting. Another limitation is that we do not know when the helpline was also triggered. Thus, we are unable to investigate how individuals behave when there is both the helpline and the advertisement, compared with just the advertisement.

Another limitation is that in study 2, we only included individuals whose ages could be ascertained by Google as 18 years or older. This was to ensure that the general versus age-tailored groups were comparable. However, we cannot determine what proportion of the population were excluded due to their ages being undetermined from past browsing history. Thus, future studies should also include individuals whose ages cannot be determined, where the analysis permits. Another key limitation in study 2 is that none of the advertisements explicitly used the word "suicide." Although this is the result of the rigorous co-design process [10], there were advisors who suggested that in some settings, explicit use of the word "suicide" may lead to higher engagement. Thus, future studies should compare engagement with the campaign when advertisements use and do not use explicit suicide wording.

Conclusion

Although the true effect of this intervention remains to be seen, the engagement metrics suggest that using internet ads may be useful in reaching out to individuals searching for suicide-related terms as a service delivery platform. Beyond suicidality, using internet ads may be helpful to reach out to individuals across a wide range of different health applications. Given that internet search is perhaps one of the most primary methods of information seeking and the ranking on search page can shape behavior, health-based organizations can use internet ads to ensure that what is presented is not only what is the most engaging but the most authoritative and valid.

Acknowledgments

Study 1 was self-funded by BERLIN SEM. Study 2 was supported by the Australian Government Department of Health–funded National Suicide Prevention Research Fund, managed by Suicide Prevention Australia; and the NHMRC Centre of Research Excellence in Suicide Prevention (APP1152952).

Authors' Contributions

For the first study, PB was responsible for the conceptualization, design, and data collection, and SOA was responsible for formal analyses. For the second study, SOA, ML, MT, FS, SA, GL, and CH were responsible for the conceptualization and design; SH, NL, CF, EE, and NG were responsible for developing the materials; PB and NL were responsible for data collection; and SOA was responsible for formal analysis and writing the original draft. AT and NJ were responsible for project administration. All authors wrote and reviewed the final draft. Study 1 was funded by PB. Funding for study 2 was obtained by SOA, ML, MT, and FS.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was only retrospectively registered. The authors explained that the trial was not prospectively registered as the discussion on whether it was a clinical trial was still tenuous. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 96 KB - mental_v10i1e42316_app1.pdf]

References

1. Areán PA, Pratap A, Hsin H, Huppert T, Hendricks K, Heagerty P, et al. Perceived Utility and Characterization of Personal Google Search Histories to Detect Data Patterns Proximal to a Suicide Attempt in Individuals Who Previously Attempted Suicide: Pilot Cohort Study. J Med Internet Res 2021 May 06;23(5):e27918 [FREE Full text] [doi: 10.2196/27918] [Medline: 33955838]
2. Lee JY. Search trends preceding increases in suicide: A cross-correlation study of monthly Google search volume and suicide rate using transfer function models. J Affect Disord 2020 Feb 01;262:155-164. [doi: 10.1016/j.jad.2019.11.014] [Medline: 31733460]
3. Sueki H. Does the volume of Internet searches using suicide-related search terms influence the suicide death rate: data from 2004 to 2009 in Japan. Psychiatry Clin Neurosci 2011 Jun;65(4):392-394 [FREE Full text] [doi: 10.1111/j.1440-1819.2011.02216.x] [Medline: 21569178]
4. Yang AC, Tsai S, Huang NE, Peng C. Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004-2009. J Affect Disord 2011 Jul;132(1-2):179-184. [doi: 10.1016/j.jad.2011.01.019] [Medline: 21371755]
5. Nicholas M. Conversion rate vs. click-through rate: what's the difference? Ladder. 2021 Jul 19. URL: https://tinyurl.com/3r6nfw94 [accessed 2023-01-27]
6. Chaffey D. 2022 comparison of Google organic clickthrough rates (SEO CTR) by ranking position. Smart Insights. 2022 Aug 29. URL: https://www.smartinsights.com/search-engine-optimisation-seo/seo-analytics/comparison-of-google-click through-rates-by-position/ [accessed 2023-01-27]
7. Keats M. Australian Internet Statistics 2021. Prosperity Media. 2021 Oct 11. URL: https://tinyurl.com/y92x9k8y [accessed 2021-12-31]
8. Heussner KM. Google 'Suicide' search feature offers lifeline. ABC News. 2010 Apr 7. URL: https://abcnews.go.com/Technology/google-suicide-search-feature-offers-lifeline/story?id=10313064 [accessed 2023-01-27]
9. Berlinquette P. I used Google Ads for social engineering. it worked. The New York Times. 2019 Jul 7. URL: https://www.nytimes.com/2019/07/07/opinion/google-ads.html [accessed 2023-01-27]
10. Onie S, Elder E, Holland S, Gale N, Finemore C, Livingstone N, et al. Co-design and development of an internet search intervention using Google AdWords for individuals contemplating suicide: the Lived Experience Hope Exchange. PsyArXiv Preprint posted online on March 24, 2023. [doi: 10.31234/osf.io/kzhgw]
11. Pernice K. Banner blindness revisited: users dodge ads on mobile and desktop. Nielsen Norman Group. 2018 Apr 22. URL: https://www.nngroup.com/articles/banner-blindness-old-and-new-findings/ [accessed 2023-01-27]
12. Statcounter (. Search engine market share worldwide. Statcounter. URL: https://gs.statcounter.com/search-engine-market-share/all/australia [accessed 2023-01-27]

13. Delete your activity. Google. URL: https://support.google.com/accounts/answer/465?hl=en&co=GENIE.Platform%3DDesktop [accessed 2023-01-27]
14. Google (. Google. URL: https://policies.google.com/technologies/partner-sites?hl=en-US [accessed 2023-01-27]
15. Google privacy policy. Google. URL: https://support.google.com/adspolicy/answer/54817?hl=en-AU [accessed 2023-01-27]
16. Sahai H, Khurshid A. Statistics in Epidemiology: Methods, Techniques, and Applications. Boca Raton, FL: CRC Press, Inc; 1996.
17. Irvine M. Google Ads benchmarks for your industry [updated!]. WorldStream. 2022 Dec 14. URL: https://www.wordstream.com/blog/ws/2016/02/29/google-adwords-industry-benchmarks [accessed 2023-01-27]
18. McAuliffe CM. Suicidal Ideation as an Articulation of Intent: A Focus for Suicide Prevention? Archives of Suicide Research 2002 Nov;6(4):325-338 [FREE Full text] [doi: 10.1080/13811110214524]
19. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. Early Interv Psychiatry 2016 Apr;10(2):103-121. [doi: 10.1111/eip.12229] [Medline: 25702826]

Abbreviations

ANZCTR: Australian New Zealand Clinical Trials Registry
CONSORT: Consolidated Standards of Reporting Trials
IRD: incidence rate difference

Edited by J Torous, G Eysenbach; submitted 31.08.22; peer-reviewed by A AL-Asadi, J Torous; comments to author 09.12.22; revised version received 19.12.22; accepted 19.12.22; published 20.04.23.

Please cite as:
Onie S, Berlinquette P, Holland S, Livingstone N, Finemore C, Gale N, Elder E, Laggis G, Heffernan C, Armstrong SO, Theobald A, Josifovski N, Torok M, Shand F, Larsen M
Suicide Prevention Using Google Ads: Randomized Controlled Trial Measuring Engagement
JMIR Ment Health 2023;10:e42316
URL: https://mental.jmir.org/2023/1/e42316
doi:10.2196/42316
PMID:37079348

Original Paper

The Impacts of a Psychoeducational Alcohol Resource During Internet-Delivered Cognitive Behavioral Therapy for Depression and Anxiety: Observational Study

Vanessa Peynenburg[1], MA; Ram P Sapkota[1], PhD; Tristen Lozinski[1], BSc; Christopher Sundström[2], PhD; Andrew Wilhelms[1], BA; Nickolai Titov[3], PhD; Blake Dear[3], PhD; Heather Hadjistavropoulos[1], PhD

[1]University of Regina, Regina, SK, Canada

[2]Karolinska Institutet, Stockholm Health Care Services, Stockholm, Sweden

[3]Macquarie University, Sydney, Australia

Corresponding Author:
Heather Hadjistavropoulos, PhD
University of Regina
3737 Wascana Parkway
Regina, SK
Canada
Phone: 1 306 585 5133
Email: hadjista@uregina.ca

Abstract

Background: Problematic alcohol use is common among clients seeking transdiagnostic internet-delivered cognitive behavioral therapy (ICBT) for depression or anxiety but is not often addressed in these treatment programs. The benefits of offering clients a psychoeducational resource focused on alcohol use during ICBT for depression or anxiety are unknown.

Objective: This observational study aimed to elucidate the impacts of addressing comorbid alcohol use in ICBT for depression and anxiety.

Methods: All patients (N=1333) who started an 8-week transdiagnostic ICBT course for depression and anxiety received access to a resource containing information, worksheets, and strategies for reducing alcohol use, including psychoeducation, reasons for change, identifying risk situations, goal setting, replacing drinking with positive activities, and information on relapse prevention. We assessed clients' use and perceptions of the resource; client characteristics associated with reviewing the resource; and whether reviewing the resource was associated with decreases in clients' alcohol use, depression, and anxiety at posttreatment and 3-month follow-up among clients dichotomized into *low-risk* and *hazardous* drinking categories based on pretreatment Alcohol Use Disorders Identification Test (AUDIT) scores.

Results: During the 8-week course, 10.8% (144/1333) of clients reviewed the resource, and those who reviewed the resource provided positive feedback (eg, 127/144, 88.2% of resource reviewers found it worth their time). Furthermore, 18.15% (242/1333) of clients exhibited hazardous drinking, with 14.9% (36/242) of these clients reviewing the resources. Compared with nonreviewers, resource reviewers were typically older (P=.004) and separated, divorced, or widowed (P<.001). Reviewers also consumed more weekly drinks (P<.001), scored higher on the AUDIT (P<.001), and were more likely to exhibit hazardous drinking (P<.001). Regardless of their drinking level (ie, low risk vs hazardous), all clients showed a reduction in AUDIT-Consumption scores (P=.004), depression (P<.001), and anxiety (P<.001) over time; in contrast, there was no change in clients' drinks per week over time (P=.81). Reviewing alcohol resources did not predict changes in AUDIT-Consumption scores or drinks per week.

Conclusions: Overall, ICBT appeared to be associated with a reduction in alcohol consumption scores, but this reduction was not greater among alcohol resource reviewers. Although there was some evidence that the resource was more likely to be used by clients with greater alcohol-related difficulties, the results suggest that further attention should be given to ensuring that those who could benefit from the resource review it to adequately assess the benefits of the resource.

(JMIR Ment Health 2023;10:e44722) doi:10.2196/44722

KEYWORDS

internet-delivered cognitive behavioral therapy; transdiagnostic; depression; anxiety; alcohol; drinking

Introduction

Background

Over the last few decades, internet-delivered cognitive behavioral therapy (ICBT) has been established as an effective treatment option for a variety of mental health concerns [1]. The content of ICBT often mirrors that of face-to-face cognitive behavioral therapy (eg, cognitive restructuring, behavioral activation, breathing strategies, and relapse prevention) but is offered in a web-based format, often in the form of a course. In particular, transdiagnostic ICBT has been developed to reduce clients' needs to engage in multiple courses of therapy to address comorbidities [2] and has been found to be similarly effective to disorder-specific treatment programs in some studies [3,4].

In transdiagnostic ICBT for depression or anxiety, it is relatively uncommon to address comorbid alcohol use difficulties. This represents a missed opportunity, as previous research suggests that heavy drinking days are common among ICBT clients, with 56.8% (514/905) of clients endorsing drinking ≥6 drinks on 1 occasion during the past year [5]. Moreover, in another study on ICBT for depression, panic disorder, and social anxiety, 24.1% (381/1581) of clients reported drinking difficulties [6]. Interestingly, alcohol use difficulties have not been found to predict ICBT completion or outcomes in terms of improved depression or anxiety, which supports the use of ICBT for depression and anxiety among individuals with comorbid alcohol use difficulties [5]. The extent to which transdiagnostic ICBT is associated with reduced alcohol use or problems over time remains unknown.

In the literature, we identified only 1 study in which participants completing a self-guided ICBT program for depression received a brief intervention related to alcohol use [7]. In this study, participants were randomized to receive either a self-guided ICBT program for depression (MoodGYM; Australian National University) or the same program with a normative feedback intervention for alcohol use (Check Your Drinking). The Check Your Drinking screener was administered at baseline, and participants received a report summarizing their drinking compared with others of the same age, sex, and country (ie, Canada), which could be accessed at any point while clients completed the ICBT program [7]. Adding a brief feedback intervention did not predict drinking or depression outcomes. However, the Check Your Drinking intervention did not provide strategies for managing alcohol use. Thus, more research is needed to understand how alcohol-related treatment may be beneficial in ICBT for depression and anxiety.

Objectives

This study included an evaluation of data from clients enrolled in an ICBT course offered in a Canadian province (Saskatchewan) over a span of 1 year (January to December 2021). This observational study aimed to add to the sparse literature and explore whether an additional alcohol resource available to clients at any point during transdiagnostic ICBT would be used and positively evaluated by clients. Furthermore, this study sought to explore whether alcohol use improved over time among clients in ICBT and, more specifically, among those who reviewed resources with low-risk or hazardous alcohol consumption.

We aimed to explore the following research questions: (1) What percentage of clients review the alcohol resource? (2) What client characteristics are associated with reviewing the alcohol resource? (3) Is the use of transdiagnostic ICBT generally associated with improvements in alcohol consumption over time? (4) Compared with clients who do not review the alcohol resource, do those who review resources show greater improvements in alcohol consumption, depression, and anxiety over time than those who do not? and (5) How will clients who review the alcohol resource evaluate it?

Given the limited nature of previous research in this area and that this was an exploratory analysis, the only hypothesis was that clients endorsing alcohol use problems at pretreatment would be more likely to review alcohol resources.

Methods

Design

This study was an uncontrolled observational trial conducted within the Online Therapy Unit, which is a Saskatchewan government–funded ICBT clinic that accepts clients for treatment on an ongoing basis. This study included data from the Online Therapy Unit's regular service delivery.

Ethics Approval

The study was approved by the institutional research ethics board of the University of Regina (approval number 2019-197).

Participants

Recruitment

Prospective clients learned about the services of the Online Therapy Unit through a variety of sources (ie, family physicians, other medical professionals, community mental health clinics, web-based searches, word of mouth, media, and posters or cards).

Sample Size

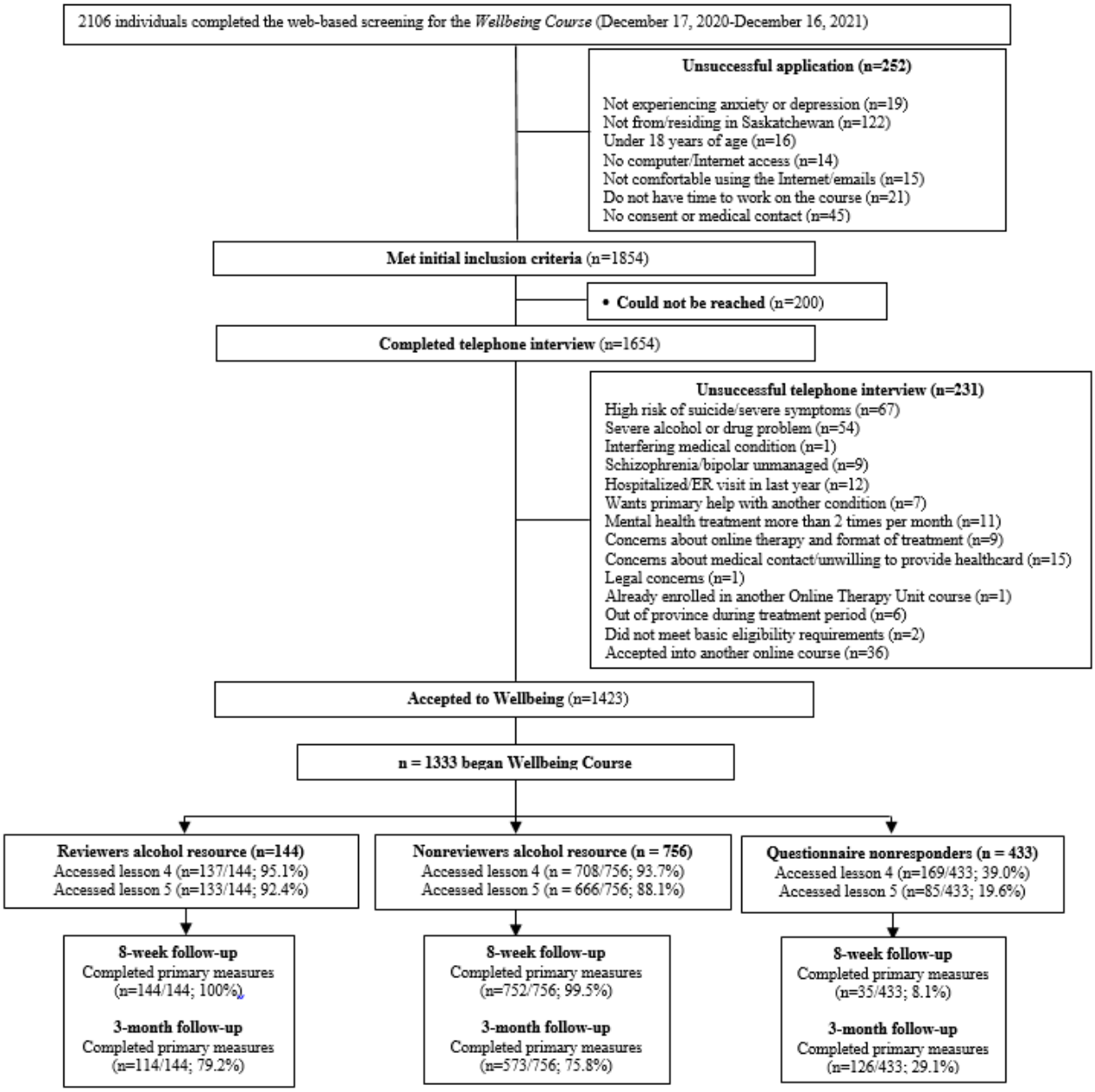
This study included all clients (N=1333) who started ICBT for anxiety and depression at the Online Therapy Unit during 2021, allowing for the analysis of a full year of service delivery.

Eligibility Criteria

All clients first completed a web-based screening via the Online Therapy Unit website, after which a telephone screening call was made. Prospective clients were eligible for the Online Therapy Unit's services and to be a part of this study if they endorsed (1) being aged at least 18 years, (2) experiencing a minimum of mild depression or anxiety symptoms as their primary concern or concerns, (3) residing in Saskatchewan for the duration of treatment, (4) having access to and comfort using a computer and the internet, (5) a willingness to provide emergency medical contact (eg, family physician and psychiatrist), and (6) consenting to and beginning ICBT. Furthermore, prospective clients were excluded from this study if they reported or were assessed as (1) exhibiting unmanaged psychosis or mania, (2) demonstrating high suicide risk, (3)

receiving mental health support from another provider more than twice per month, and (4) experiencing severe difficulties with alcohol use (ie, scoring ≥20 on the Alcohol Use Disorder Identification Test [AUDIT]) [8] or other substance use (ie, scoring ≥25 on the Drug Use Disorder Identification Test [DUDIT]) [9]. Figure 1 shows a client flowchart.

Figure 1. Client flow from screening to 3-month follow-up. ER: emergency room.



Measures

Overview

All primary and secondary outcome measures were administered at pretreatment, posttreatment (8 weeks after enrollment), and 3-month follow-up. Clients completed treatment satisfaction and resource evaluation questions at posttreatment. Although clients responded to additional questions about their symptoms over the course of the 8-week treatment period, the measures listed below were the focus of this study. Clients also completed questionnaires assessing insomnia, panic, social anxiety, mental health-related disability, treatment experiences, mental health service use, and pandemic-related anxiety but were not used as part of this observational study.

Baseline Screening Measures

AUDIT Screening Tool

AUDIT [8] is a well-standardized 10-item screening measure for alcohol use difficulties. The total scores range from 0 to 40, with higher scores indicating greater alcohol-related difficulties. Scores from 6 to 14 (for women) and 8 to 14 (for men) indicate hazardous alcohol use, scores from 15 to 19 indicate harmful alcohol use, and scores ≥20 indicate possible alcohol dependence [10]. In this trial, we referred to all scores ≥6 (for women) and

≥8 (for men or other) as indicative of hazardous alcohol use. In this study, the Cronbach α for the AUDIT was .75.

DUDIT Screening Tool

The DUDIT [9] is a well-standardized 11-item screening tool for substance use difficulties. The total scores ranged from 0 to 44. Higher scores indicated greater difficulties in substance use. Scores from 2 to 24 (for women) and 6 to 24 (for men) indicated difficulties with substance use, and scores ≥25 indicated possible substance dependence. In this study, the Cronbach α for the DUDIT was .77.

Demographics

Clients responded to questions regarding their age, gender, relationship status, race and ethnicity, location, education, and employment status during the web-based screening. Furthermore, clients were asked if they had taken psychotropic medications within the past 3 months.

Primary Outcome Measures

AUDIT-Consumption

AUDIT-Consumption (AUDIT-C) [11] consists of the first 3 consumption items of the full AUDIT [12]. Previous studies have suggested that the AUDIT-C has similar sensitivity and specificity indices as the full AUDIT [13]. Cronbach α for the AUDIT-C in this study was .62 and .77.

Total Drinks in Previous Week

Clients were asked to indicate how many standard drinks of alcohol they had drank in the last 7 days. This question is commonly used in ICBT trials of alcohol misuse [14,15].

Secondary Outcome Measures

Patient Health Questionnaire 9-Item

The Patient Health Questionnaire 9-item (PHQ-9) [16] is a screening measure for depression, with total scores ranging from 0 to 27. Scores <5 indicated minimal depression, and scores ≥10 were used to identify probable cases of major depressive disorder [17]. Cronbach α in this study was between .84 and .89.

Generalized Anxiety Disorder 7-Item

The Generalized Anxiety Disorder 7-item (GAD-7) [18] is a measure to screen for generalized anxiety disorders. Total scores ranged from 0 to 21, with scores <5 indicating minimal symptoms of anxiety and scores ≥10 indicating clinically significant symptoms of generalized anxiety [18]. Cronbach α in this study was between .87 and .91.

Treatment Engagement

Clients' treatment engagement was captured based on whether they completed all 5 lessons as well as their total number of website log-ins over the 8-week treatment period. The platform does not track how long clients are logged in because this is a biased estimate impacted by whether clients fail to log out.

Treatment Satisfaction

Clients were asked yes-or-no questions about whether the course was worth their time and whether they would recommend the course to a friend. On 5-point scales, they were also asked to rate their overall satisfaction with the course (1=*very dissatisfied* to 5=*very satisfied*).

Alcohol Resource Evaluation Survey

Clients were asked a yes-or-no question about whether they reviewed the alcohol resource. If they reported reviewing the resource, clients were asked to rate their level of effort dedicated to reviewing the resource from 1=*none at all* to 7=*a great deal*. Furthermore, from 1=*not at all* to 7=*very*, clients were asked (1) how understandable the resource was, (2) if they learned something new from the resource, and (3) how helpful the resource was. Clients then responded to 3 open-ended questions asking them to describe what they liked about the resource, what they disliked about the resource, and the nature of any changes they made to their drinking because of reviewing the resource.

Intervention

The Well-being Course

All clients were offered a therapist-assisted well-being course, which is an 8-week transdiagnostic ICBT course for depression and anxiety [3]. The course was developed at the eCentreClinic at Macquarie University and is licensed for use by the Online Therapy Unit. During the course, clients read five web-based lessons based on key components of cognitive behavioral therapy: (1) psychoeducation about symptoms and the cognitive behavioral model; (2) thought monitoring and challenging; (3) physical symptoms of depression or anxiety, de-arousal strategies, and pleasant activity scheduling; (4) graded exposure; and (5) relapse prevention. Clients read materials presented as slides, case stories, frequently asked questions, and downloadable guides, with homework activities and lesson summaries. All materials were presented in the English language. Lessons were released gradually over 8 weeks, and clients needed to complete each lesson before proceeding to the next lesson. Clients received automated reminder emails as upcoming lessons became available. Furthermore, all clients were assigned to a therapist for the duration of the 8-week ICBT program. Clients received either optional therapist support (ie, support provided at client requests) or once-weekly support. These 2 approaches for providing therapist support in ICBT have been found to be effective [19].

In addition to the 5 core lessons, clients can access additional downloadable resources at any time. The resources addressed a wide range of topics, namely, anger, alcohol use, assertiveness, beliefs, chronic conditions, chronic pain, communication skills, grief, health anxiety, mental skills, motivation, new motherhood, panic, posttraumatic stress disorder, sleep, workplace mental health, and worry. In this trial, therapists informed clients about the availability of resources during their first message to clients and made tailored recommendations based on the clients' presenting concerns. Clients can also self-select the resources to review. During week 5, therapists asked all clients whether they had questions about any additional resources that they had reviewed.

Alcohol Resource

The alcohol resource was based on content included in an ICBT program for alcohol use called the Alcohol Change Course [14], which was developed in collaboration with a patient-oriented research steering committee comprising researchers, clinicians, stakeholders, trainees, and patient partners with lived experiences. The Alcohol Change Course and the alcohol resource use a relapse prevention model to address alcohol use [20]. The resource consisted of 17 pages and was in a downloadable format. Content wise, the resource started with a section on the relationship between alcohol use and mental health and provided information about the stimulating and inhibiting effects of alcohol, reasons for drinking alcohol, how alcohol use difficulties can vary in severity, and Canada's low-risk drinking guidelines (enacted in 2011) that recommended consuming ≤10 (for women) and ≤15 (for men) weekly drinks [21]. The next sections focused on how alcohol affects physical health; the connection between alcohol consumption, depression, and anxiety; and the impact of alcohol on sleep. The remaining sections highlighted strategies for changing one's drinking habits. Clients were prompted to consider their reasons for drinking and were provided worksheets to note the pros and cons of drinking, how drinking does or does not align with their values, and their reasons for change. A list of questions was provided to assist the clients in setting goals related to alcohol use. Thereafter, the resource included sections on identifying personal strengths and supports, changing the availability of alcohol in one's day-to-day life, a worksheet on identifying risk situations and positive activities to replace drinking, information on "slips," and a summary of the resource.

Therapist Support

Clients who scored in the clinical range (≥10) on either the PHQ-9 or GAD-7 were offered optional or regular once-weekly therapist support. Clients who scored in the nonclinical range (<10) on both PHQ-9 and GAD-7 were offered optional weekly support. In regular once-weekly therapist support, therapists reviewed symptom measures and sent clients a brief, tailored message once a week. Clients could also be contacted via phone if they had not logged in during the past week, if their PHQ-9 or GAD-7 scores increased by 5 or more points, or if there was an indication of elevated suicide risk. In optional support, therapists would only contact clients if they initiated contact that week, if their PHQ-9 or GAD-7 score increased by 5 or more points, or if there was an indication of elevated suicide risk. Previous research has shown that both approaches are similarly effective, and approximately 25% of clients prefer optional support [19,22].

Data Analyses

Overview

Analyses were conducted using SPSS (version 28.0.0.0; IBM Corp) [23]. In an initial review of the data, the following 3 client groups were identified: clients who reported reviewing the alcohol resource ("reviewers"; n=144), clients who reported not reviewing the alcohol resource ("nonreviewers"; n=756), and those who did not respond to the Alcohol Resource Evaluation Survey ("questionnaire nonresponders" [QNRs]; n=433). Descriptive statistics were used to describe the pretreatment characteristics of the 3 groups. Furthermore, ANOVA, χ^2 analyses, and *t* tests (2-tailed) were used to assess group differences on all pretreatment variables. A significance level of *P*=.01 was used as a partial control for multiple comparisons.

There were no missing data for the clients' baseline screening variables. Data for the primary and secondary outcome variables were missing mainly because of client dropout (eg, 4/896, 0.4% to 213/687, 23.7% at posttreatment and follow-up, respectively). The missing data were determined to be missing completely at random (MCAR) via Little MCAR test (χ^2_{44}=56.8; *P*=.09) [24].

To assess whether there was a significant change in clients' AUDIT-C scores and previous weekly drinks over time, as well as to determine if pretreatment, posttreatment, and follow-up scores differed across clients and between groups (ie, reviewers vs nonreviewers), a series of mixed models were computed using the maximum likelihood estimation method with 3 assessment points (ie, pretreatment, posttreatment, and follow-up). As only 900 clients reported whether they reviewed the alcohol resource, mixed model analyses were performed with this subsample (*group*=reviewers vs nonreviewers) to assess the rate of change in these clients' AUDIT-C scores and previous weekly drinks as well as their depression and anxiety. These analyses were also conducted with clients dichotomized based on their pretreatment AUDIT scores into *low-risk* drinking (ie, scores <6 for women and <8 for men or other) and *hazardous* drinking (ie, scores ≥6 for women and ≥8 for men or other).

Missing data were not imputed because the data were assumed to be MCAR, and linear mixed model analysis can handle missing data [25]. Fixed effects for *time*, *group*, and their interactions (ie, *time* × *group*) and random effects for the intercept and *time* variables were tested and included in the model to account for the correlated nature of the data. Intraclass correlation coefficients were calculated to identify the proportion of variance across clients and to determine if mixed model analyses were appropriate [26]. To select the model that best fit the data, various within-individual and between-individual covariance structures (eg, scaled identity, diagonal, unstructured, and autoregressive) were tested. Models with the smallest Akaike information criterion and Bayesian information criterion were retained for the final analysis. The repeated-measure indicator variable, *time*, was recoded as 0, 2, and 5 to reflect the actual assessment months (ie, 0=pretreatment, 2=posttreatment at 2 months [8 weeks], and 5=follow-up at 5 months).

Treatment Engagement and Satisfaction

Treatment engagement and satisfaction were compared between alcohol resource reviewers and nonreviewers through a combination of ANOVA and χ^2 analyses.

Qualitative Data Analysis

Client responses to the Alcohol Resource Evaluation Survey were analyzed using conventional qualitative content analysis

[27]. The literature on including a resource for managing alcohol use in transdiagnostic ICBT is limited; therefore, conventional qualitative content analysis can be helpful for identifying quantifiable response categories [28].

The coding process consisted of the following steps:

1. VP reviewed all client responses to questions regarding what clients liked about the alcohol resource, disliked about the resource, and any changes they made to their drinking because of the resource. A codebook was created during the review. The codebook consisted of identified codes, a description of each code, and an example quote for each code. A decision to review all client responses to ensure saturation was made during the creation of the codebook.
2. AW and TL used the codebook to independently code all client responses. The 2 coders were able to assign more than 1 code to a client response where appropriate (eg, if clients identified making more than 1 change after reviewing the resources).
3. VP reviewed the responses of AW and TL to identify and resolve instances of disagreement.

Results

Client Characteristics

The pretreatment client characteristics of the overall sample are presented in Table 1. Most clients reported identifying as a woman (1036/1333, 77.72%), being in a married or common law relationship (1210/1333, 90.77%), identifying as White (1158/1333, 86.87%), living in a large city (778/1333, 58.36%), being educated beyond high school (1058/1333, 79.37%), and being employed either part time or full time (699/1333, 52.44%). In terms of clinical characteristics, more than half of the clients (732/1333, 54.91%) reported taking psychotropic medication or medications in the past 3 months, and most of the sample exhibited clinically remarkable symptoms of depression (955/1333, 71.64%) and anxiety (949/1333, 71.19%). The rates of alcohol consumption per week and hazardous alcohol use are summarized in Table 1. Of note, while 49.67% (662/1333) of the clients did not consume any previous weekly drinks at pretreatment, 6.8% (45/662) of these clients reviewed the alcohol resource. Similarly, 24.46% (326/1333) of clients reported never drinking alcohol by scoring 0 on the pretreatment AUDIT; however, 5.8% (19/326) of these clients reviewed the alcohol resource.

Table 1 displays the pretreatment characteristics of resource reviewers, nonreviewers, and QNRs. In terms of significant differences between resource reviewers and nonreviewers, resource reviewers were older (t_{898}=2.90; P=.004), more likely to be men (χ^2_1=16.2; P<.001), as well as more often separated, divorced, or widowed (χ^2_2=14.2; P<.001). Furthermore, compared with nonreviewers, reviewers consumed more weekly drinks (t_{898}=6.08; P<.001); scored significantly higher on the AUDIT (t_{898}=6.81; P<.001); and were more likely to score above the cutoff, indicating hazardous drinking (χ^2_1=14.6; P<.001). Most clients received standard once-weekly therapist support (reviewers: 84/144, 58.3%; and nonreviewers: 448/756, 59.2%).

In terms of significant differences between reviewers and QNRs, QNRs were typically younger (t_{575}=7.81; P<.001), were more often single or never married (χ^2_2=20.3; P<.001), and had a lower education level compared with resource reviewers (χ^2_2=11.1; P=.004). In addition, QNRs were more likely to be women (χ^2_1=7.9; P=.005), consumed fewer weekly drinks (t_{575}=4.11; P<.001), and had higher depression symptoms (t_{575}=3.14; P=.002).

Table 1. Client characteristics at pretreatment.

Variable	All clients (N=1333)	Reviewers (n=144)	Nonreviewers (n=756)	Questionnaire nonresponders (n=433)	Significance	
					Test	*P* value
Age (years), mean (SD; range)	38.04 (13.78; 18-86)	43.17 (13.70; 20-77)	39.46 (14.16; 18-86)	33.85 (11.95; 18-77)	$F_{2,1330}$=35.79	<.001
Gender, n (%)					χ^2_2=16.2	<.001
Woman	1036 (77.72)	94 (65.3)	608 (80.4)	334 (77.1)		
Man or other	297 (22.28)	50 (34.7)	148 (19.6)	99 (22.9)		
Relationship status, n (%)					χ^2_4=34.7	<.001
Single or never married	408 (30.61)	29 (20.1)	211 (27.9)	168 (38.8)		
Married or common law	802 (60.17)	90 (62.5)	485 (64.2)	227 (52.4)		
Separated, divorced, or widowed	123 (9.22)	25 (17.4)	60 (7.9)	38 (8.8)		
Ethnicity, n (%)					χ^2_4=5.8	.22
Indigenous	81 (6.08)	9 (6.2)	38 (5)	34 (7.8)		
Other	94 (7)	9 (6.2)	49 (6.5)	36 (8.3)		
White	1158 (86.87)	126 (87.5)	669 (88.5)	363 (83.8)		
Location, n (%)					χ^2_4=6.7	.15
Large city (>100,000)	778 (58.36)	89 (61.8)	437 (57.8)	252 (58.2)		
Small to medium city	190 (14.25)	13 (9)	104 (13.8)	73 (16.8)		
Small rural location (<7000)	365 (27.38)	42 (29.2)	215 (28.4)	108 (24.9)		
Education, n (%)					χ^2_4=23.5	<.001
High school or less	275 (20.63)	20 (13.9)	143 (18.9)	112 (25.9)		
More than high school or less than university	589 (44.19)	69 (47.9)	317 (41.9)	203 (46.9)		
University education	469 (35.18)	55 (38.2)	296 (39.2)	118 (27.2)		
Employment status, n (%)					χ^2_4=9.8	.04
Employed part time or full time	699 (52.44)	75 (52.1)	380 (50.3)	244 (56.4)		
Unemployed or disability	250 (18.75)	27 (18.8)	135 (17.8)	88 (20.3)		
Homemaker, student, or retired	384 (28.81)	42 (29.2)	241 (31.9)	101 (23.3)		
Pretreatment scores						
Value, n (%)						
Psychotropic medication in the past 3 months	732 (54.91)	77 (53.5)	405 (53.6)	250 (57.7)	χ^2_2=2.1	.36
AUDIT[a] ≥6 for women and ≥8 for men or other	242 (18.15)	36 (25)	96 (12.7)	110 (25.4)	χ^2_2=35.0	<.001
PHQ-9[b] ≥10	955 (71.64)	102 (70.8)	506 (66.9)	347 (80.1)	χ^2_2=23.7	<.001
GAD-7[c] ≥10	949 (71.19)	104 (72.2)	512 (67.7)	333 (76.9)	χ^2_2=11.4	.003
Value, mean (SD)						
Drinks per week	2.89 (6.78)	6.35 (15.63)	2.23 (4.42)	2.87 (4.69)	$F_{2,1330}$=23.00	<.001
AUDIT	3.39 (3.76)	4.75 (4.17)	2.72 (3.07)	4.11 (4.11)	$F_{2,1330}$=30.37	<.001
PHQ-9	13.50 (5.74)	13.34 (5.61)	12.65 (5.69)	15.03 (5.58)	$F_{2,1330}$=24.48	<.001
GAD-7	12.64 (5.09)	12.51 (5.00)	12.10 (5.06)	13.61 (5.04)	$F_{2,1330}$=12.23	<.001

[a]AUDIT: Alcohol Use Disorders Identification Test.

[b]PHQ-9: Patient Health Questionnaire 9-item.

[c]GAD-7: Generalized Anxiety Disorder 7-item.

Primary Outcome Variables

Clients' mean AUDIT-C scores were 2.33 (SD 2.09) at pretreatment, 2.22 (SD 1.94) at posttreatment, and 2.11 (SD 1.94) at 3-month follow-up. The mixed model analysis predicting clients' AUDIT-C scores revealed a significant decrease in AUDIT-C scores over time (β=−.03, SE 0.01; *P*=.004). Although not significant, there was a negative correlation (*r*=−0.17; *P*=.06) between the intercept and slope, which may indicate that compared with those with lower pretreatment AUDIT-C scores, individuals with higher scores experienced greater reductions in scores over time. Furthermore, although there was a significant *group* (ie, reviewers vs nonreviewers) effect (β=1.10, SE 0.171; *P*<.001), the interaction effect was not significant (*P*=.69). This indicates that although reviewers and nonreviewers varied significantly in pretreatment AUDIT-C scores, reviewing the alcohol resource did not influence changes in their AUDIT-C scores over time. The subsequent mixed model analysis, with clients dichotomized into low-risk and hazardous drinking subgroups, showed a significant decrease in AUDIT-C scores over time for clients in the hazardous drinking subgroup (β=−.21, SE 0.06; *P*=.002). In contrast, *time* was not significant for clients in the low-risk subgroup (*P*=.14). Although there was a significant *group* effect for clients in the low-risk subgroup (β=.74, SE 0.16; *P*<.001), there was no *group* effect for those in the hazardous subgroup (*P*=.91). Furthermore, there was no significant interaction effect for either the low-risk (*P*=.44) or hazardous (*P*=.83) drinking subgroups, indicating that reviewing alcohol resources had no effect on the rate of change in clients' AUDIT-C scores over time, irrespective of drinking problems.

On average, clients consumed 2.89 (SD 6.78) weekly drinks at pretreatment, 2.44 (SD 4.18) weekly drinks at posttreatment, and 2.77 (SD 5.29) weekly drinks at 3-month follow-up. The mixed model analysis predicting clients' previous weekly drinking showed that there was no significant *time* effect (*P*=.81) or interaction effect (*P*=.35). Furthermore, the mixed model analyses dichotomizing clients into low-risk and hazardous drinking subgroups showed no reductions in weekly drinks over time for both clients in the low-risk (*P*=.24) and hazardous (β=−.80, SE 0.40; *P*=.05) drinking subgroups. There was no significant interaction effect for either the low-risk (*P*=.84) or hazardous (*P*=.66) drinking subgroups.

In sum, the results of the primary analyses show that regardless of their pretreatment alcohol use difficulties, there was a significant decrease in clients' AUDIT-C scores over time and no change in clients' previous weekly drinks over time. Furthermore, reviewing the alcohol resources did not influence changes over time in clients' AUDIT-C scores or previous weekly drinks.

Secondary Outcome Variables

Overview

Clients' mean PHQ-9 scores were 13.50 (SD 5.75) at pretreatment, 6.85 (SD 5.36) at posttreatment, and 5.90 (SD 4.97) at 3-month follow-up. Furthermore, their mean GAD-7 scores were 12.64 (SD 5.09) at pretreatment, 6.16 (SD 4.98) at posttreatment, and 5.36 (SD 4.83) at 3-month follow-up. Mixed model analyses revealed significant *time* effects for depression (β=−1.10, SE 0.04; *P*<.001) and anxiety (β=−1.22, SE 0.04; *P*<.001). Yet, there was no significant main effect (*P*=.45) or interaction effect (*P*=.34) of *group* (ie, reviewers vs nonreviewers) predicting depression. Furthermore, although there was no main effect (*P*=.53) of *group* predicting decreases in anxiety, there was a significant interaction effect (β=−.21, SE 0.10; *P*=.04), indicating that the anxiety of alcohol resource reviewers decreased more than that of nonreviewers.

Moreover, the mixed model analysis dichotomizing clients into low-risk and hazardous drinking subgroups showed a statistically significant decrease in depression over time among clients in both the low-risk (β=−1.35, SE 0.05; *P*<.001) and hazardous (β=−1.58, SE 0.19; *P*<.001) drinking subgroups. There were no statistically significant *group* or interaction effects for either the low-risk (group: *P*=.19; interaction: *P*=.18) or hazardous (group: *P*=.17; interaction: *P*=.73) drinking subgroup. These findings suggest that reviewing alcohol resources had no effect on changes in clients' PHQ-9 scores over time, irrespective of their level of drinking.

Similarly, the mixed model analysis dichotomizing clients into low-risk and hazardous drinking subgroups revealed a significant decrease in anxiety over time for clients in both the low-risk (β=−1.20, SE 0.04; *P*<.001) and hazardous (β=−1.41, SE 0.15; *P*<.001) drinking subgroups. There was no significant *group* effect for clients in either the low-risk (*P*=.27) or hazardous (*P*=.13) drinking subgroup. Although the interaction effect was significant in the low-risk subgroup (β=−.26, SE 0.12; *P*=.03), it was not significant in the hazardous subgroup (*P*=.54). These results suggest that for clients in the low-risk drinking group, resource reviewers' anxiety decreased more than that of nonreviewers.

Treatment Engagement and Satisfaction

Treatment engagement was also examined by resource reviewers and nonreviewers (n=900; Table 2). Treatment completion rates were high in both groups, with 88.8% (799/900) of the clients accessing all 5 ICBT lessons. No significant group differences were found for any measure of treatment engagement or satisfaction (all *P*>.01).

Table 2. Treatment engagement and satisfaction.

Variable	Reviewers and nonreviewers (n=900)	Reviewers (n=144)	Nonreviewers (n=756)	Significance	
				Test	*P* value
Engagement					
Accessed lesson 5, n (%)	799 (88.8)	133 (92.4)	666 (88.1)	χ^2_1=2.2	.14
Number of website log-ins, mean (SD)	26.55 (27.78)	26.58 (15.44)	26.54 (29.56)	t_{898}=.01	.99
Satisfaction, n (%)					
Course was worth the time	846 (96)	140 (97.2)	724 (95.8)	χ^2_1=.7	.41
Would recommend course to friend	867 (96.3)	140 (97.2)	727 (96.2)	χ^2_1=.4	.54
Satisfied or very satisfied overall	744 (82.7)	124 (86.1)	620 (82)	χ^2_1=1.4	.23

Evaluation of Alcohol Resource

A total of 144 clients responded to the Alcohol Resource Evaluation Survey. Clients indicated dedicating a moderate amount of effort into the resource (mean 3.79, SD 1.92), moderately agreed that they had learned something new by reviewing the resource (mean 4.31, SD 2.00), found the resource moderately helpful (mean 4.74, SD 1.78), and rated the resource as very understandable (mean 6.15, SD 1.11). Most clients who reviewed the resources indicated that it was worth their time (127/144, 88.2%).

Likes About the Resource

Of the 144 responses, 122 (84.7%) were codable as "likes" (see Table 3 for codes). The most common liked aspect of the alcohol resource was that it was informative (eg, "It was informative and I was able to learn information I wasn't aware of before" [client ID 35854]). This was followed by comments about how the resource gave clients insight into their drinking (eg, "It made me evaluate alcohol use" [client ID 34723]). Some clients also commented on how they liked the way the information was presented (eg, "easy to read and understand, not complicated at all" [client ID 35645]). The remaining comments focused on how clients appreciated learning about the relationship between alcohol consumption and symptoms of depression and anxiety (eg, "It helped me see alcohol in a different way. It helped me see the effects it was implementing on my anxiety." [client ID 35902]), how they were able to use the resource to either better understand or support their loved ones' drinking difficulties (eg, "I was able to understand some things and pass it along to my husband to use" [client ID 35876]), how the resource acted as a review of information they had learned about alcohol use in the past (eg, "It was consistent with and validated other resources I have encountered over the years"), how they liked specific worksheet activities included in the resource (eg, "I like that it gave a worksheet to list the values of why I wanted to reduce drinking" [client ID 36683]), and how they appreciated the information provided about the negative effects of drinking (eg, "It explained a lot about the effects of alcohol on the physical and mental components of the body" [client ID 36029]).

Table 3. Client responses to the Alcohol Resource Evaluation Survey.

Responses	Example	Client ID	Values, n (%)
What did you like about the resource? (n=144)			
Informative	• "Learning things I didn't know."	• 35687	62 (43.1)
Not a like	• "I do not drink. Therefore doesn't pertain to me."	• 35755	23 (16)
Insight into one's drinking	• "Helped me to gain more perspective about my drinking and helped me make the commitment to stop drinking completely"	• 34424	21 (14.6)
Presentation of information	• "Very clear and concise"	• 36081	18 (12.5)
Information on the relationship between alcohol use and symptoms of anxiety or depression	• "It confirmed for me the effects alcohol can have on depression and mood."	• 35688	12 (8.3)
Resource allowed them to help or understand others	• "I used the resource to help me with family members who are alcohol users"	• 35722	10 (6.9)
Resource acted as a refresher	• "it refreshed some knowledge I had"	• 36292	8 (5.6)
Worksheet activity	• "The question prompts were helpful to think about."	• 36027	7 (4.9)
Information on the negative effects of drinking	• "That there are negative consequences beyond just drinking too much."	• 35829	6 (4.2)
What did you not like about the resource? (n=144)			
Nothing	• "There was not anything I didn't like."	• 34368	109 (75.7)
Not relevant to client experience	• "It did not apply to me; I drink a couple beers a year."	• 34054	12 (8.3)
Format or structure issues	• "I wish the PDF was fillable"	• 35890	7 (4.9)
Insight into one's drinking	• "I honestly don't remember anything I didn't like except maybe how it forced me to think more actively about my drinking/alcohol consumption patterns."	• 36643	5 (3.5)
No new information	• "No new information, but not the fault of the resource..."	• 36715	4 (2.8)
Resource did not focus on other substances or addictions	• "You should expand this to include cannabis use too" • "I would like to see other addiction issues discussed–particularly regarding technology"	• 34853 • 34429	3 (2.1)
Generic negative comment	• "I just did not find it helpful."	• 35629	2 (1.4)
Not a dislike	• "I think a weekly sessions with a therapist would be helpful via zoom"	• 34368	2 (1.4)
Did you make any changes to your alcohol use based on the alcohol resource? (n=144)			
Yes			43 (29.9)
Reduced drinking	• "Yes, reduced alcohol consumption"	• 34811	27 (62.8)
Increased awareness of drinking habits	• "It helped me realize sometimes I would grab a drink after a long day when my husband wasn't home because I was lonely."	• 34781	18 (41.9)
Replaced drinking with more helpful coping strategies	• "...made healthier choices when I was having a bad day."	• 34238	7 (16.3)
No	• "No. I read the resource because I was curious (and like information) not because I think I have a problem with alcohol."	• 36099	101 (70.1)

Dislikes About the Resource

Most clients did not report disliking any aspect of the resource. Among clients who shared a dislike, the most common concern was that the resource was not relevant to their personal experience (eg, "I didn't think it really applied, I've never really had issues with alcohol use outside a short stint in my late teens, early twenties." [client ID 36202]). Other clients expressed concerns with the format or structure of the resource (eg, "I wish the PDF was fillable" [client ID 35890]) or felt that the resource did not provide them with any new information (eg, "mostly stuff I already knew" [client ID 34733]). Some clients found it challenging to have increased insight into the frequency of their alcohol consumption or severity of their alcohol concerns (eg, "It reminded me how much I'm binge drinking" [client ID 35312]). A small subgroup of clients thought that the resource should address other addictions or substance use (eg, "You should expand this to include cannabis use too" [client ID 34853]). Two clients responded with generic negative comments that did not fit into the other categories (eg, "I just did not find it helpful" [client ID 35629]). See Table 3 for a summary of clients' dislikes regarding the alcohol resource.

Changes Made to Alcohol Use

In total, 3 types of changes emerged in clients' responses, namely, reduced drinking (eg, "Yes I have cut down and now only will have a drink at a social function" [client ID 34667]), increased awareness of one's drinking habits (eg, "It helped me realize sometimes I would grab a drink after a long day when my husband wasn't home because I was lonely" [client ID 34781]), and replacing drinking with more helpful coping strategies (eg, "made healthier choices when having a bad day" [client ID 34238]).

Discussion

Principal Findings

This observational study investigated whether clients enrolled in an 8-week transdiagnostic ICBT course for depression and anxiety would review, benefit from, and positively evaluate an additional resource for addressing alcohol use. We also aimed to explore the demographic and clinically relevant variables associated with reviewing the alcohol resource. Across all groups, participants showed improvements in alcohol consumption, depression, and anxiety over time. Compared with nonreviewers, clients who accessed the resource were more likely to be older; men; and separated, divorced, or widowed. Furthermore, as expected, reviewers were more likely to consume more weekly drinks, report higher alcohol use difficulties, and have higher levels of hazardous drinking. Ratings and comments from resource reviewers indicated high satisfaction with the resource. The client ratings suggested that the resource was helpful and understandable, and most clients indicated that it was worth their time. Clients' most liked aspects of the resource were that it was informative and that it assisted them in gaining insight into their drinking behavior. Resource reviewers did not differ from nonreviewers in any indices of overall treatment engagement (ie, course completion and website log-ins) or satisfaction with the ICBT course overall.

Interestingly, the prevalence of hazardous or harmful drinking based on AUDIT was slightly higher in this trial (242/1333, 18.15%) than in previous ICBT samples (160/1155, 13.85%) [5]. The elevated levels of hazardous drinking may be explained by the trial occurring during the COVID-19 pandemic, as 20% of Canadians who stayed home during the pandemic reported increased alcohol consumption [29]. Regardless, the high pretreatment hazardous alcohol use rates among clients are intriguing, given that the transdiagnostic ICBT program was intended to primarily address depression and anxiety symptoms. Clients scoring above the cutoff for hazardous drinking would have been eligible for another course offered by the Online Therapy Unit (ie, the Alcohol Change Course) [14], which focused primarily on alcohol use difficulties. It is possible that clients chose to enroll in the transdiagnostic ICBT course because of their desire to focus on depression and anxiety symptoms or because they had limited insight into their drinking concerns. Clients who were highly motivated to change their drinking behaviors may have enrolled in the Alcohol Change Course instead of the transdiagnostic ICBT course; therefore, although hazardous or harmful drinking was common in this trial, the clients who chose the transdiagnostic ICBT course may not have been motivated to change their drinking. As reductions in depression [30], depression, and anxiety [14] have been reported in trials of ICBT for alcohol use, these clients were likely to benefit from either course offered by the Online Therapy Unit.

Findings from this study are consistent with a randomized controlled trial that examined the inclusion of a brief web-based alcohol use intervention in conjunction with a web-based depression intervention [7]. Similar to that trial, we found no additional alcohol-related benefits associated with alcohol consumption, including a brief resource focused on reducing alcohol use. However, the inclusion of resources did not negatively impact clients' engagement with the main intervention targeting depression and anxiety. Of note, while reviewing the resources was not associated with greater reductions in alcohol use, all transdiagnostic ICBT clients—regardless of whether they exhibited low-risk or hazardous drinking—showed decreased alcohol consumption scores, depression, and anxiety over time.

Limitations and Future Directions

This study had several limitations that can help guide future research. The study was observational in nature and had no control group; therefore, causal conclusions about the impact of the resource cannot be made. The Alcohol Resource Evaluation Survey was administered at posttreatment, and responses were missing from approximately one-third of the clients. It is possible that some of the clients who did not respond to the Alcohol Resource Evaluation Survey actually reviewed the resource; however, because of their missing self-reports, we did not have information about their perceptions of the resource and whether the resource was helpful in reducing alcohol consumption by these clients. Future studies could include a resource evaluation survey at midtreatment to ensure higher response rates. Clients could also be asked about their alcohol consumption weekly to allow for tracking of changes in alcohol use over the span of the course. Moreover, it would be beneficial

to use the Alcohol Timeline Followback [31] assessment method (ie, asking about consumption on each of the 7 preceding days) rather than the total weekly drinks to better assess the amount and pattern of drinking, which would allow for the assessment of heavy drinking days.

Only 10.8% (144/1333) of clients in this study reviewed the additional alcohol resource, including only 14.9% (36/242) of those with AUDIT scores indicating hazardous alcohol use. As the ICBT course included 19 additional resources, clients who may have benefited from alcohol may have prioritized other resources. Alternatively, clients who scored above the cutoff for hazardous drinking may not have perceived difficulties with their drinking patterns or they may have been focused on managing their symptoms of depression or anxiety before addressing any concerns related to alcohol. A limitation of the Alcohol Resource Evaluation Survey was that it did not ask clients to provide a reason for why they chose not to review alcohol resources. As such, future studies could include questions to better understand clients' decisions not to review resources. Including a measure of motivation to change drinking behaviors could also be worthwhile, as this study only assessed need based on hazardous or harmful drinking, and it is not possible to determine how motivated these clients were to use resources or change their drinking behaviors. In the future, the main ICBT course content could include information regarding low-risk drinking guidelines and therapists could direct clients' attention to the alcohol resource if they fall within the hazardous risk ranges for their drinking. Such methods could assist in increasing the uptake of alcohol resources, thereby facilitating more substantial opportunities for future research to elucidate the impact of reviewing resources in transdiagnostic ICBT. Therapists could recommend the resource at the beginning of the course based on clients' pretreatment scores on the AUDIT or throughout the course based on clients' concerns related to alcohol consumption. This study did not examine whether therapists recommend resource-predicted clients to review the resource, which could be an area of future study. It would also be worthwhile to ensure consistent follow-up from therapists regarding the resource after it has been recommended to clients.

Nearly one-fourth (326/1333, 24.46%) of clients reported never drinking alcohol on the pretreatment AUDIT; interestingly, 5.8% (19/326) of these clients reviewed the resource. In addition, almost half of the clients (662/1333, 49.66%) did not consume any drinks in the previous week at pretreatment, and 6.8% (45/662) of them reviewed the resource. This may have because of many reasons (eg, having a family member with alcohol use difficulties or having undisclosed or past problems with alcohol). Nonetheless, as only increases in drinking over time are possible for these groups, nondrinkers reviewing alcohol resources may have confounded the results, particularly for the mixed model analysis predicting changes in clients' previous weekly drinks. We can also see echoes of further potential confounds within these nondrinkers' qualitative feedback. For instance, the most common (12/33, 36%) dislike expressed by clients was that the resource was not relevant to their experience. As such, a further direction for research would be to randomly assign clients with alcohol use difficulties to (1) a transdiagnostic ICBT course targeting depression and anxiety or (2) the same transdiagnostic ICBT course plus an additional alcohol resource.

In terms of pretreatment characteristics, it is also important to highlight that most clients in this study identified as a woman and as White. These trends in mental health service use based on gender and ethnicity are not unique to ICBT [32,33]. Although acknowledging that biological sex differences are not inherently representative of differences based on gender identity, findings from epidemiological studies have suggested that alcohol use disorder is more common among male than female individuals [34] and that the relationship between alcohol use disorder and racial and ethnic groups is nuanced and influenced by historical and ongoing discrimination [35]. Therefore, the current sample may not be representative of the diverse population of individuals who experience alcohol-related difficulties. Future studies should aim to recruit a more diverse sample of clients.

Strengths

To our knowledge, no previous studies on transdiagnostic ICBT for depression and anxiety have specifically examined the inclusion of a resource addressing alcohol use. As drinking difficulties are a prevalent concern among ICBT clients with depression and anxiety [5,6], this study makes an important contribution to the literature by examining the uptake and utility of a resource that can address this area of concern. The mixed methods (ie, both quantitative and qualitative data collections) approach is also a strength of this study. Furthermore, the study included a combination of objective and subjective measures of client engagement and perceptions of the resource, which can help inform changes to the content and presentation of the resource for future clinical research and practice.

Conclusions

Consistent with previous research [5,6], nearly one-fifth of clients who enroll in transdiagnostic ICBT are identified as having problems with alcohol. A brief resource focused on alcohol use may be a way for some clients to address their concerns related to alcohol use while concurrently learning strategies to manage their symptoms of depression and anxiety. It appears that the resource is used more often by those who need it than by those who do not, and most reviewers are satisfied with the alcohol resource. Furthermore, some reviewers reported making changes to their alcohol use, such as increased awareness about drinking, reduced drinking, and the use of alternative strategies to cope instead of consuming alcohol. In general, ICBT was associated with improvements in alcohol use for all clients, and there was no added benefit in reviewing the resource. The minimal added benefit from reviewing the resource, especially for those with hazardous drinking, may reflect that more attention needs to be paid to ensuring that the resource is reviewed by those who need it.

Acknowledgments

The authors wish to acknowledge the Patient-Oriented Research Steering Committee, including Matthew Keough and Michael Schaub, for their contributions to developing the alcohol resource, and the Saskatchewan Health Resource Foundation and Saskatchewan Centre for Patient-Oriented Research for funding that supported the development of the resource. The authors would also like to thank the clients, research and clinical staff, students, and web developers associated with the Online Therapy Unit at the University of Regina; the therapists from the Saskatchewan Health Authority; and the Saskatchewan Ministry of Health for funding, which allowed the delivery of internet-delivered cognitive behavioral therapy in Saskatchewan.

Conflicts of Interest

None declared.

References

1. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019 Jul;64(7):465-470. [doi: 10.1177/0706743719839381] [Medline: 31096757]
2. Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, et al. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. JAMA Psychiatry 2017 Sep 01;74(9):875-884 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.2164] [Medline: 28768327]
3. Titov N, Dear BF, Staples LG, Terides MD, Karin E, Sheehan J, et al. Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: a randomized controlled trial. J Anxiety Disord 2015 Oct;35:88-102 [FREE Full text] [doi: 10.1016/j.janxdis.2015.08.002] [Medline: 26422822]
4. Dear BF, Staples LG, Terides MD, Karin E, Zou J, Johnston L, et al. Transdiagnostic versus disorder-specific and clinician-guided versus self-guided internet-delivered treatment for generalized anxiety disorder and comorbid disorders: a randomized controlled trial. J Anxiety Disord 2015 Dec;36:63-77 [FREE Full text] [doi: 10.1016/j.janxdis.2015.09.003] [Medline: 26460536]
5. Sundström C, Edmonds M, Soucy JN, Titov N, Dear BF, Hadjistavropoulos HD. Alcohol and drug use among clients receiving internet-delivered cognitive behavior therapy for anxiety and depression in a routine care clinic - demographics, use patterns, and prediction of treatment completion and outcomes. Internet Interv 2022 Mar;27:100490 [FREE Full text] [doi: 10.1016/j.invent.2021.100490] [Medline: 34987979]
6. Gajecki M, Berman AH, Sinadinovic K, Andersson C, Ljótsson B, Hedman E, et al. Effects of baseline problematic alcohol and drug use on internet-based cognitive behavioral therapy outcomes for depression, panic disorder and social anxiety disorder. PLoS One 2014 Aug 14;9(8):e104615 [FREE Full text] [doi: 10.1371/journal.pone.0104615] [Medline: 25122509]
7. Cunningham JA, Godinho A, Hendershot CS, Kay-Lambkin F, Neighbors C, Griffiths KM, et al. Randomized controlled trial of online interventions for co-occurring depression and hazardous alcohol consumption: primary outcome results. Internet Interv 2021 Nov 02;26:100477 [FREE Full text] [doi: 10.1016/j.invent.2021.100477] [Medline: 34786350]
8. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
9. Berman AH, Bergman H, Palmstierna T, Schlyter F. DUDIT: the drug use disorders identification test manual. Karolinska Institutet. 2003 Mar. URL: https://paihdelinkki.fi/sites/default/files/duditmanual.pdf [accessed 2022-07-21]
10. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. AUDIT: the alcohol use disorders identification test: guidelines for use in primary health care. 2nd edition. World Health Organization. 2001. URL: https://apps.who.int/iris/handle/10665/67205 [accessed 2022-11-21]
11. Dawson DA, Grant BF, Stinson FS. The AUDIT-C: screening for alcohol use disorders and risk drinking in the presence of other psychiatric disorders. Compr Psychiatry 2005 Nov;46(6):405-416. [doi: 10.1016/j.comppsych.2005.01.006] [Medline: 16275207]
12. Bohn MJ, Babor TF, Kranzler HR. The Alcohol Use Disorders Identification Test (AUDIT): validation of a screening instrument for use in medical settings. J Stud Alcohol 1995 Jul;56(4):423-432. [doi: 10.15288/jsa.1995.56.423] [Medline: 7674678]
13. Meneses-Gaya C, Zuardi AW, Loureiro SR, Hallak JE, Trzesniak C, de Azevedo Marques JM, et al. Is the full version of the AUDIT really necessary? Study of the validity and internal construct of its abbreviated versions. Alcohol Clin Exp Res 2010 Aug;34(8):1417-1424. [doi: 10.1111/j.1530-0277.2010.01225.x] [Medline: 20491736]
14. Sundström C, Peynenburg V, Chadwick C, Thiessen D, Wilhems A, Nugent M, et al. Optimizing internet-delivered cognitive behaviour therapy for alcohol misuse-a randomized factorial trial examining effects of a pre-treatment assessment interview and guidance. Addict Sci Clin Pract 2022 Jul 23;17(1):37 [FREE Full text] [doi: 10.1186/s13722-022-00319-0] [Medline: 35871010]

15. Hadjistavropoulos HD, Mehta S, Wilhelms A, Keough MT, Sundström C. A systematic review of internet-delivered cognitive behavior therapy for alcohol misuse: study characteristics, program content and outcomes. Cogn Behav Ther 2020 Jul;49(4):327-346. [doi: 10.1080/16506073.2019.1663258] [Medline: 31599198]
16. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
17. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012 Feb 21;184(3):E191-E196 [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]
18. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
19. Hadjistavropoulos HD, Schneider LH, Mehta S, Karin E, Dear BF, Titov N. Preference trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. J Anxiety Disord 2019 Apr;63:51-60 [FREE Full text] [doi: 10.1016/j.janxdis.2019.02.002] [Medline: 30844613]
20. Larimer ME, Palmer RS, Marlatt GA. Relapse prevention. An overview of Marlatt's cognitive-behavioral model. Alcohol Res Health 1999;23(2):151-160 [FREE Full text] [Medline: 10890810]
21. Butt P, Beirness D, Gliksman L, Paradis C, Stockwell T. Alcohol and health in Canada: a summary of evidence and guidelines for low-risk drinking. Canadian Centre on Substance Abuse. 2011 Nov 25. URL: https://www.ccsa.ca/sites/default/files/2019-04/2011-Summary-of-Evidence-and-Guidelines-for-Low-Risk%20Drinking-en.pdf [accessed 2022-11-15]
22. Hadjistavropoulos HD, Schneider LH, Edmonds M, Karin E, Nugent MN, Dirkse D, et al. Randomized controlled trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. J Anxiety Disord 2017 Dec;52:15-24 [FREE Full text] [doi: 10.1016/j.janxdis.2017.09.006] [Medline: 28964994]
23. IBM SPSS Statistics for Windows. Version 28.0.0.0. IBM Corp. 2021. URL: https://www.ibm.com/products/spss-statistics [accessed 2022-07-21]
24. Little RJ, Rubin DB. Statistical Analysis with Missing Data. 2nd edition. Hoboken, NJ, USA: John Wiley & Sons; Sep 09, 2002.
25. Twisk JW. Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide. 2nd edition. Cambridge, UK: Cambridge University Press; May 2013.
26. Peugh JL. A practical guide to multilevel modeling. J Sch Psychol 2010 Feb;48(1):85-112. [doi: 10.1016/j.jsp.2009.09.002] [Medline: 20006989]
27. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
28. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
29. COVID-19 and increased alcohol consumption: NANOS poll summary report. Canadian Centre on Substance Use and Addiction. 2020. URL: https://www.ccsa.ca/covid-19-and-increased-alcohol-consumption-nanos-poll-summary-report [accessed 2022-11-15]
30. Sundström C, Eék N, Kraepelien M, Fahlke C, Gajecki M, Jakobson M, et al. High- versus low-intensity internet interventions for alcohol use disorders: results of a three-armed randomized controlled superiority trial. Addiction 2020 May;115(5):863-874 [FREE Full text] [doi: 10.1111/add.14871] [Medline: 31691413]
31. Sobell LC, Sobell MB. Timeline follow-back: a technique for assessing self-reported alcohol consumption. In: Litten RZ, Allen JP, editors. Measuring Alcohol Consumption: Psychosocial and Biochemical Methods. Totowa, NJ, USA: Humana Press; 1992:41-72.
32. Pattyn E, Verhaeghe M, Bracke P. The gender gap in mental health service use. Soc Psychiatry Psychiatr Epidemiol 2015 Jul;50(7):1089-1095. [doi: 10.1007/s00127-015-1038-x] [Medline: 25788391]
33. Chiu M, Amartey A, Wang X, Kurdyak P. Ethnic differences in mental health status and service utilization: a population-based study in Ontario, Canada. Can J Psychiatry 2018 Jul;63(7):481-491 [FREE Full text] [doi: 10.1177/0706743717741061] [Medline: 29514512]
34. Agabio R, Pisanu C, Gessa GL, Franconi F. Sex differences in alcohol use disorder. Curr Med Chem 2017;24(24):2661-2670. [doi: 10.2174/0929867323666161202092908] [Medline: 27915987]
35. Vaeth PA, Wang-Schweig M, Caetano R. Drinking, alcohol use disorder, and treatment access and utilization among U.S. racial/ethnic groups. Alcohol Clin Exp Res 2017 Jan;41(1):6-19. [doi: 10.1111/acer.13285]

Abbreviations

AUDIT: Alcohol Use Disorder Identification Test
AUDIT-C: Alcohol Use Disorder Identification Test-Consumption
DUDIT: Drug Use Disorder Identification Test
GAD-7: Generalized Anxiety Disorder 7-item
ICBT: internet-delivered cognitive behavioral therapy

MCAR: missing completely at random
PHQ-9: Patient Health Questionnaire 9-item
QNR: questionnaire nonresponder

Edited by J Torous, G Eysenbach; submitted 01.12.22; peer-reviewed by N Rauseo-Ricupero, V Kaldo; comments to author 10.03.23; revised version received 23.03.23; accepted 28.03.23; published 18.04.23.

Please cite as:
Peynenburg V, Sapkota RP, Lozinski T, Sundström C, Wilhelms A, Titov N, Dear B, Hadjistavropoulos H
The Impacts of a Psychoeducational Alcohol Resource During Internet-Delivered Cognitive Behavioral Therapy for Depression and Anxiety: Observational Study
JMIR Ment Health 2023;10:e44722
URL: https://mental.jmir.org/2023/1/e44722
doi:10.2196/44722
PMID:37071454

Original Paper

Mobile Acceptance and Commitment Therapy With Distressed First-Generation College Students: Microrandomized Trial

Emily Brenny Kroska Thomas[1], PhD; Tijana Sagorac Gruichich[2], BS; Jacob M Maronge[3], PhD; Sydney Hoel[2], BS; Amanda Victory[4], MS; Zachary N Stowe[2], MD; Amy Cochran[5,6], PhD

[1]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, United States
[2]Department of Psychiatry, University of Wisconsin-Madison, Madison, WI, United States
[3]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, United States
[4]Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States
[5]Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI, United States
[6]Department of Mathematics, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
Emily Brenny Kroska Thomas, PhD
Department of Psychological and Brain Sciences
University of Iowa
340 Iowa Avenue, G60 PBSB
Iowa City, IA, 52242
United States
Phone: 1 319 467 1691
Email: emily-kroska@uiowa.edu

Abstract

Background: Extant gaps in mental health services are intensified among first-generation college students. Improving access to empirically based interventions is critical, and mobile health (mHealth) interventions are growing in support. Acceptance and commitment therapy (ACT) is an empirically supported intervention that has been applied to college students, via mobile app, and in brief intervals.

Objective: This study evaluated the safety, feasibility, and effectiveness of an ACT-based mHealth intervention using a microrandomized trial (MRT) design.

Methods: Participants (N=34) were 18- to 19-year-old first-generation college students reporting distress, who participated in a 6-week intervention period of twice-daily assessments and randomization to intervention. Participants logged symptoms, moods, and behaviors on the mobile app Lorevimo. After the assessment, participants were randomized to an ACT-based intervention or no intervention. Analyses examined proximal change after randomization using a weighted and centered least squares approach. Outcomes included values-based and avoidance behavior, as well as depressive symptoms and perceived stress.

Results: The findings indicated the intervention was safe and feasible. The intervention increased values-based behavior but did not decrease avoidance behavior. The intervention reduced depressive symptoms but not perceived stress.

Conclusions: An MRT of an mHealth ACT-based intervention among distressed first-generation college students suggests that a larger MRT is warranted. Future investigations may tailor interventions to contexts where intervention is most impactful.

Trial Registration: ClinicalTrials.gov NCT04081662; https://clinicaltrials.gov/show/NCT04081662

International Registered Report Identifier (IRRID): RR2-10.2196/17086

(JMIR Ment Health 2023;10:e43065) doi:10.2196/43065

KEYWORDS

acceptance and commitment therapy; randomized controlled trials; mobile health; mHealth; first-generation college students; psychological flexibility; distress; depression

Introduction

Background

Despite previous research regarding its necessity, there is a gap in access to mental health services [1-5], and the gap has been further widened by the COVID-19 pandemic [6,7]. According to Mental Health America, 24.7% of adults with mental illness report unmet needs relating to treatment [8]. Gaps in access to mental health services are further exacerbated on college campuses and among students [9,10]. The prevalence of mental health concerns and barriers to treatment on campus have been termed a crisis [11-13]. One approach to closing treatment gaps is to provide more accessible treatments via technology in order to ameliorate the current mental health crisis college students face [14-17].

One way to improve accessibility is through mobile health (mHealth) interventions. mHealth interventions have demonstrated effectiveness in a variety of conditions, including depression, anxiety, bipolar disorder, borderline personality disorder, and post-traumatic stress disorder [18-23]. Specifically, just-in-time adaptive interventions are successful in providing individuals with the support they need based on their situation and context, and technological advances provide the framework to apply these concepts to mental health [24]. Just-in-time adaptive interventions highlight the effectiveness of interventions, but also when, which, and for whom interventions are most effective, providing critical contextual sensitivity that is often missed in traditional 2-armed randomized-controlled trials. Beyond the development and evaluation of mHealth interventions, when addressing psychiatric conditions, it is important to expand beyond diagnoses and target transdiagnostic psychotherapeutic processes that extend beyond syndromes into functioning.

Psychological Flexibility

One such process is psychological flexibility, defined as the ability to engage in behavior that is consistent with one's values even when challenged or distressed [25,26]. Psychological flexibility is targeted in acceptance and commitment therapy (ACT), a transdiagnostic intervention with demonstrated efficacy [27].

Psychological flexibility is comprised of interrelated processes, often divided into three core pillars: (1) *openness to experiences*—willingness to make contact with emotions, thoughts, physical sensations, urges, and memories without judgment; (2) *awareness*—purposefully paying attention to the experiences of the present moment and the ability to notice the function of one's behaviors; and (3) *engagement with values*—behavioral pursuit of personally chosen values in a consistent and flexible manner. These processes are theoretically interwoven, and attention to these processes allows for targeted intervention driven by precise case conceptualization.

Several mHealth interventions have examined the efficacy of ACT in a variety of different samples. For example, SmartQuit, a mobile app designed for smoking cessation, demonstrated higher engagement and quit rates than a non–ACT-based app [28]. In a more recent randomized controlled trial, an ACT-based mobile intervention increased the likelihood of quitting smoking [29]. Others have used both mobile apps and in-person or phone interventions [30,31]. Several studies have tested a web-based ACT intervention though not via mobile app [32-38]. Importantly, ACT has also demonstrated effectiveness in brief interventions, indicating that change in psychological flexibility can be observed even with a small "dose" [36,38-42].

First-Generation College Students

College students are known to be at risk for mental health conditions during this developmental window, especially for depression, anxiety, and eating disorders [43]. At particular risk are first-generation college students (FGCSs), defined here as students whose parents or legal guardians have attained less education than a bachelor's degree. Risk is demonstrated in additional stress compared to non-FGCSs, greater need for service while also seeking fewer services, and working while being a student, among other factors [44-46]. mHealth interventions are a plausible solution to address the need for both convenience and efficacy.

Objectives of This Study

This study sought to examine the safety, feasibility, and preliminary effectiveness of an mHealth ACT intervention for FGCSs reporting distress. The goal of the intervention was to determine whether mobile ACT was safe for delivery (ie, did not worsen depression), feasibility (ie, adherence to in-app assessments), and preliminary effectiveness (the *proximal* impact of intervention). Primary outcomes included values-based and avoidance behaviors, and secondary outcomes included depressive symptoms and stress. The authors hypothesized that the intervention would be safe; participants would respond to over 60% of the in-app assessments; and the intervention would reduce avoidance behavior, depressive symptoms, and stress while increasing values-based behavior.

Methods

Overview

The protocol for this study and a parallel trial with patients with bipolar disorder was published [47]. To summarize, this study examined a brief 6-week mHealth intervention with FGCSs with primary effectiveness outcomes of values-based and avoidance-based behavior and secondary outcomes of depressive symptoms and stress. The microrandomized trial (MRT) included randomization to ACT intervention or no intervention twice per day after assessments were completed.

Electronic informed consent was obtained. Given the focus on depressive symptoms, suicidality was closely monitored throughout assessments. Outcomes included safety, feasibility, and effectiveness.

Ethics Approval

The study was approved by the institutional review board at the University of Wisconsin-Madison (#2019-0819) and was registered at ClinicalTrials.gov (NCT04081662).

Participants

FGCSs were recruited from the University of Wisconsin-Madison through advertisements and mass electronic mailings. Potential participants contacted the research team via completion of the screening survey or by phone, wherein a screening phone call was conducted. Inclusion criteria were being an adult, a full-time first- or second-year student, an FGCS, fluent in English, and having access to a smartphone. First-generation status was defined as a student whose parents had not completed a 4-year college degree. FGCSs had to report distress on 4 or more of the last 7 days that interfered with functioning when asked "Over the past seven days, on how many days did you experience distress that interfered with your ability to fulfill your responsibilities in one or more domains of life (eg, school, home, social, work, and intimate relationship)?"; this criterion was used as college students may be more likely to endorse stress than anxiety or depressive symptoms [48]. Notably, *stress* and *distress* refer to different constructs, and the authors' aim was to recruit participants who were experiencing interference with functioning secondary to distress. Potential participants had to be willing to receive a consent form for consideration via email, as Research Electronic Data Capture (REDCap, Vanderbilt University) software was used to obtain electronic informed consent. After screening for eligibility, the research team reviewed the informed consent and study procedures with the potential participant, and if interested, the document was sent via REDCap. Recruitment for the study began in the fall semester of 2019, and a second recruitment effort occurred in the fall of 2020. The research team aimed to recruit 50 participants, but due to the COVID-19 pandemic, the number of consented participants was less than the goal (N=34). Details on how a target sample size was determined are found in the published protocol [47].

Procedures

Assessment

Participants completed a baseline assessment after consenting to participate. The assessment included measures of current symptoms, stress, functioning, and psychological flexibility [47]. Daily in-app assessments were completed throughout the 6-week intervention period. At the conclusion of the intervention (day 42), participants completed a web-based assessment, including current symptoms, stress, functioning, and psychological flexibility. In addition, a mobile app engagement survey was administered. Furthermore, a 3- and 6-month follow-up survey assessed symptoms, stress, functioning, and psychological flexibility. Participants were compensated for completion of the baseline and follow-up assessments and for each week of the 6-week intervention period in which they completed at least 50% of daily in-app assessments.

Mobile App

After consenting to participate and completing the baseline questionnaire, participants downloaded a mobile app called Lorevimo (Log, Review, and Visualize your Mood). The app was designed by the senior author, and participants could download the app without charge from the Apple or Google Play stores. Upon opening Lorevimo, the participants set typical wake and bedtimes for weekdays and weekends. These times determined the morning and evening assessment intervals, with morning occurring 2-7 hours after waking and evening 3 hours before or 2 hours after bedtime. Participants consented to receive notifications from the Lorevimo app. Notifications were sent at the start of the interval and every 2 hours following the initial notification until symptoms were logged or it was within 30 minutes of the participant's reported bedtime. Finally, participants watched an introductory video that introduced the ACT Matrix [49], a therapeutic tool used to help participants notice and sort emotions, thoughts, and behaviors. The video was 20 minutes in length and depicted a role play of a therapist (ET) and student (SH) walking through the ACT Matrix. At the end of the video, the therapist noted that the terminology would be used each day throughout the intervention, defining *toward* behaviors as those that move a person toward personally held values (ie, values-based behaviors), and *away* behaviors as those that function to avoid unwanted internal experiences (eg, emotions and thoughts).

Daily Assessments

Participants responded to twice-daily prompts in Lorevimo. Prompts were time-sensitive, and morning prompts assessed symptoms since waking, and evening prompts assessed symptoms since about lunchtime. Participants completed the Patient Health Questionnaire–2 [50], a measure of the 2 key symptoms of depression identified in the Diagnostic and Statistical Manual-5 [51] to assess depressive symptoms. Items assess dysphoria and anhedonia on an ordinal scale from 0 to 3, ranging from *absent* to *severe*. In addition, participants completed the Perceived Stress Scale- 4 [52] to assess perceived stress. Items are rated on a 0-4 ordinal scale ranging from *never* to *very often*. Finally, participants completed a 4-item ACT activity scale that was developed for the study and parallel trial. Questions are stated with the intended construct following:

1. "In a few words, what behavior are you engaged in now?" (behavioral form)
2. "Does this behavior move you toward who/what matters or away from internal experiences? [Reminder: if the behavior is both, choose which best fits.]" (behavioral function)
3. "Since [lunchtime or waking up], how much energy was consumed by trying to get rid of unwanted feelings, thoughts, and other internal experiences (eg, suppressing, distracting, avoiding)?" (avoidance behavior)
4. "Since [lunchtime or waking up], how much energy was consumed by pursuing your values? (eg, making choices that align with who you want to be or who/what matters)" (values-based behavior).

The behavioral form question was a free response, and the behavioral function was categorical (toward or away). The avoidance and values-based behavior questions were rated on a 0 to 6 ordinal scale, ranging from *none* to *all of my energy*.

Intervention

After completing the in-app assessment, participants were randomized to receive an intervention or no intervention (1:1). This repeated randomization occurred every time an assessment

was completed. Throughout the 6-week intervention period, a participant who completed all assessments could receive up to 84 interventions. If the participant was randomized to no intervention, the app returned them to the Lorevimo homepage. If randomized to the intervention, the intervention was then randomly chosen from 84 possible prompts such that each intervention prompt was equally likely to be chosen, regardless of whether that intervention had been previously received.

The intervention prompts were developed by the first author with feedback from the research team. The prompts spanned the three core pillars of ACT [53]: awareness of internal experiences, openness to experiences, and engagement with values, discussed from here forward as awareness, openness, and engagement. The prompts were brief and in congruence with the microintervention design. Awareness intervention prompts focused on the promotion of mindful attention to context, internal experiences, and behaviors. Perspective taking and noticing the many parts of the human experience were emphasized. Sample awareness intervention prompts include: "Throughout your day, what can you notice with your five senses (sight, sound, taste, touch, smell)?" and "In what situations do you notice yourself acting on autopilot?" Openness prompts were intended to facilitate acceptance of internal experiences (eg, thoughts, emotions, physical sensations, memories) and detachment from thoughts. Sample openness prompts include: "When you experience a difficult emotion, what do you notice yourself doing to avoid or suppress the emotion?" and "When you tell yourself to not think or feel a certain way, when does it work, and when does it not work?" Engagement prompts centered on identification and clarification of personal values, consistency of actions with one's values, and behavioral pursuit of one's values. Sample engagement prompts include: "What is the smallest step you could take toward something that matters to you?" and "Over the past week, did you notice yourself doing something in service of what matters most, even when difficult thoughts or emotions were present?" Openness, awareness, and engagement intervention prompts were represented in equal proportion of the total 84 prompts (ie, 28 prompts in each domain).

Statistical Analyses

Key outcomes included: (1) safety, operationalized as nonworsening in depressive symptoms from baseline to follow-up; (2) feasibility, operationalized as completion of in-app assessments; and (3) preliminary effectiveness, operationalized as the *proximal* impact of the intervention on values-based and avoidance behavior. We included sociodemographic characteristics, intervention characteristics, and in-app ratings prior to randomization as covariates in the analyses.

Two-tailed hypothesis tests were used for all analyses, and significance was defined as $P<.05$. To examine feasibility, a one-sample *z*-test assessed whether participants completed 50% of the assessments per day (1 of 2) for at least 60% of the intervention days (42 total). Safety was examined whether the change in depressive symptoms and stress was significantly different from 0 using a *z*-test, as well as using a sign test to determine whether equal proportions of individuals reported decreased depressive symptoms as those who reported increases in depressive symptoms. Effectiveness was examined using a weighted and centered least squares method [54,55]. When intervention assignment is randomized with a constant probability, as is done in this study, the estimation procedure is effectively equivalent to fitting a generalized estimating equation (GEE) with an independent working correlation structure. As in GEE, a working model is specified for the population-level mean, which in the case of an MRT is the mean of a given proximal outcome conditional on a participant completing the assessment, thus being available for randomization. We specified a linear working model for this conditional mean that consisted of an intercept, intervention delivery, time, intervention delivery × time interaction, and additional control variables predictive of missingness (these variables are described below). A model was built separately for each primary outcome (values-based behavior and avoidance behavior) and each secondary outcome (depressive symptoms and stress). The inferential target of estimation was the coefficient associated with intervention delivery and its interaction with time, representing the population's average effect of delivering an intervention on a given outcome as a function of time and conditional on being available for randomization. A sandwich estimator was used to calculate robust standard errors [56]. Additional analyses were performed by repeating the exact estimation procedure except for changing the proximal outcome and adding interactions terms with intervention delivery.

For effectiveness, statistical analyses were powered to analyze the primary outcomes (values-based behavior and avoidance behavior). However, additional analyses were performed to analyze secondary and exploratory outcomes and interactions, and these analyses involve hypothesis tests. Multiple comparisons were not controlled for in these additional analyses, and as such, the reported *P* values should be considered nominal.

The percentage missing for primary outcomes was greater than 10%, so as prespecified in the protocol, additional variables that predicted missingness were added to the linear working model. Variables that predicted missingness were age, time of day, day in the study, count of prior completed in-app assessments, count of prior missing data points, and depressive symptoms reported immediately prior to randomization. Selection of these variables was guided by an increase of 2 in the quasi-information criterion (QIC) [57], which is similar to the commonly used BIC, but for GEEs. This was implemented using the *geepack* package in R (R Foundation) [56,58]. Due to the observed missingness, the accuracy of results relies on the validity of the missingness model.

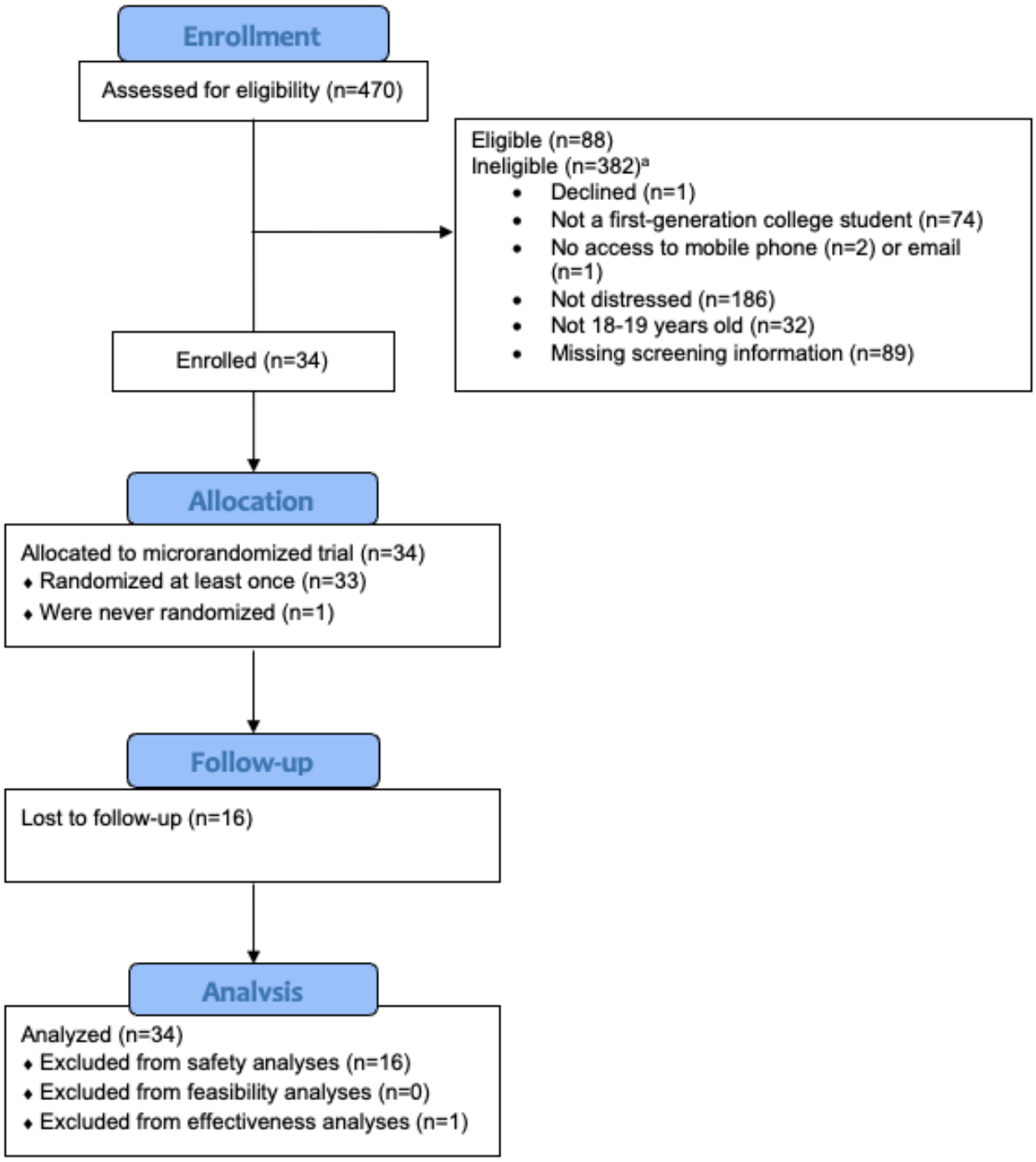
Results

Study Flow

A total of 470 individuals completed the web-based eligibility screening, of which 88 were eligible. Of these, 42 ultimately provided signed informed consent to participate. After consent, participants were asked to complete a baseline assessment, and 34 of the consented participants completed the baseline assessment. Following the baseline, participants were asked to

download the mobile app, complete information about sleep and wake times, enable push notifications, and watch the introductory video. Of the 34 participants who completed the baseline, 33 downloaded the app and logged symptoms in the app at least once. The intervention lasted for 6 weeks, and at 3- and 6 months following the baseline, 18 participants completed the 3-month follow-up, and 15 participants completed the 6-month follow-up. For analyses, the 34 participants who completed the baseline assessment will be used to determine feasibility outcomes; the 18 participants who completed the 3-month follow-up assessments will be used to determine safety outcomes; and the 33 participants who logged symptoms at least once will be used to determine preliminary effectiveness (Figure 1).

Figure 1. CONSORT flow diagram. [a]Some individuals were ineligible for more than 1 reason.



Sample Characteristics

Table 1 summarizes the characteristics of the sample (N=34). They had an average age of 18.53 (SD 0.53) years and were 85% (n=28) female. The majority were White (n=21, 62%) and non-Hispanic (n=31, 91%).

Table 1. Characteristics of the sample population (N=34).[a]

Variable	Values
Age (years), mean (SD)	18.53 (0.53)
Female, n (%)[a]	28 (85)
Race, n (%)	
African American or Black	3 (9)
Asian	7 (21)
Caucasian	21 (62)
More than one	2 (6)
Other	1 (3)
Hispanic	3 (9)
Baseline PHQ-9[b] score, mean (SD)	10.35 (4.85)

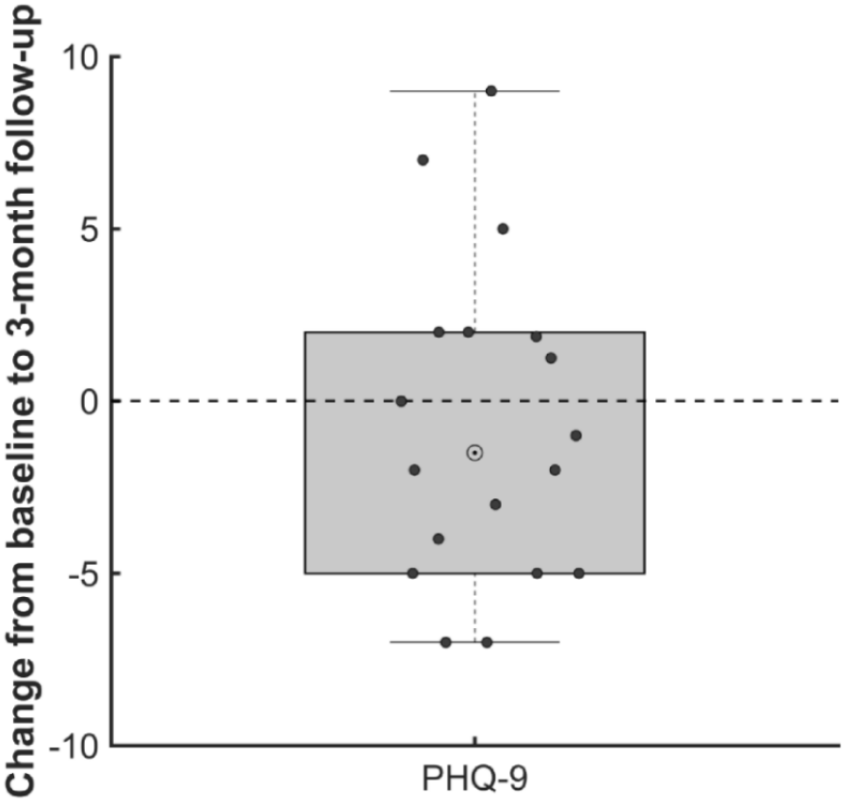
[a]One individual reported nonbinary gender identity.

[b]PHQ-9: Patient Health Questionnaire–9.

Safety

Analyses examined whether Patient Health Questionnaire–9 (PHQ-9) scores worsened from baseline assessment to 3-month follow-up. Figure 2 illustrates the change from baseline to 3-month follow-up in PHQ-9 scores for the 18 participants who completed the baseline and follow-up measures. Depressive symptom severity decreased slightly with an average decrease in PHQ-9 score of 0.72 points (t_{17}=.65; P=.52). Using the cutoff of a score of 10 or higher on the PHQ-9 as an indication of moderate or higher depressive symptoms, 4 participants who were depressed at baseline were no longer depressed at follow-up, compared to 2 participants who were not depressed at baseline but were depressed at follow-up (67%; z=0.82; P=.41). Furthermore, 7 participants were not depressed at baseline or follow-up, leaving 5 participants who were depressed both at baseline and follow-up. Importantly, on average, it appears participants who completed the baseline and 3-month follow-up assessments did not report worsened depressive symptoms following the intervention period.

Figure 2. Change in Patient Health Questionnaire–9 (PHQ-9) scores from baseline to 3-month follow-up.



Feasibility

Because app engagement is a primary concern with mobile interventions, our second analysis investigated whether participants were available for in-app randomization, that is, whether a participant logged their symptoms in app at 1 of the 84 time points (=2 per day×42 days). After completing consent, participants were available for randomization for an average of 57% of the time points. Excluding the 1 participant who never logged symptoms, availability increased to an average of 58% of the time points. In addition, participants were available at least once per day for an average of 71% of the days. This average was significantly larger than our prespecified target of 60% (t_{33}=2.13; *P*=.04). Again, excluding the participant who never logged symptoms, participants were available at least once per day for an average of 73% of the days.

Effectiveness

Models assessed whether being randomized to intervention impacted the outcome, while controlling for relevant covariates, including the number of prior assessments completed, age (18 or 19 years old), morning (0) or evening (1) assessment, and day of intervention (centered). The most proximal measurement of depression was included as a covariate to estimate mood at the time point prior to randomization. Finally, the day × intervention interaction examined whether interventions were more effective at any point in the 42-day intervention. These analyses exclude the participant who never logged symptoms, thereby making them unavailable for randomization.

Primary outcomes included energy focused on values and energy focused on avoidance. For values-focused energy, intervention predicted increased values-focused energy at the next assessment (χ^2_1=5.58; *P*=.02). Higher depressive symptoms predicted lower values-focused energy at the next assessment (χ^2_1=10.06; *P*=.002). The full model is reported in Table 2. For avoidance-focused energy, the intervention did not predict avoidance (*P*=.24). Higher depressive symptoms predicted higher avoidance-focused energy at the next assessment (χ^2_1=68.24; *P*<.001). The full model is reported in Table 3.

Secondary outcomes included depressive symptoms and perceived stress. For depression, intervention predicted lower depressive symptoms at the next assessment (χ^2_1=8.56; *P*=.003). Higher depressive symptoms predicted higher depressive symptoms at the next assessment (χ^2_1=43.97, *P*<.001). The full model is reported in Table 4. For perceived stress, intervention trended, but was not statistically significant, toward predicting decreased perceived stress (χ^2_1=3.72; *P*=.05). Moreover, higher depressive symptoms predicted higher perceived stress at the next assessment (χ^2_1=32.20; *P*<.001). The full model is reported in Table 5.

Exploratory analyses examined the specific process being targeted in the intervention (ie, engagement, awareness, and openness) as predictive of the outcomes for which the intervention had a significant effect. With values-focused energy, interventions targeting awareness (χ^2_1=11.99; *P*=.001) predicted greater values-focused energy at the next assessment. Table 2 presents the exploratory model. With depression, interventions targeting awareness (χ^2_1=9.58; *P*=.002) and openness (χ^2_1=5.08; *P*=.02) predicted lower depressive symptoms at the next assessment (Table 4). An additional exploratory model examined the interaction between intervention and depressive symptoms (at assessment just before intervention) in predicting depressive symptoms (at next assessment), finding no significant interaction (*P*=.72). Further, exploratory models investigated *for whom* the intervention may be most effective, finding no significant interaction between intervention and sex in predicting depressive symptoms and values-focused energy. Similarly, an age×intervention interaction was nonsignificant in predicting depressive symptoms and values-focused intervention. Finally, due to the small number in non-White racial groups, a categorical comparison of White and non-White participants was conducted, wherein the race×intervention interaction was investigated as a nonsignificant predictor of depressive symptoms or values-focused energy.

Table 2. Predictors of values-focused energy, including primary analysis and exploratory analysis of intervention type.

Predictor	Estimate	SE	χ^2_1	*P* value	95% CI
Primary model					
Evening[a]	−0.11	0.11	0.95	.33	−0.33 to 0.11
Prior assessment[b]	0.03	0.04	0.46	.50	−0.05 to 0.10
Day[c]	−0.06	0.08	0.57	.45	−0.20 to 0.09
Prior missing[d]	0.04	0.07	0.32	.58	−0.10 to 0.18
Age[e]	0.01	0.36	0.001	.98	−0.69 to 0.71
Depression[f]	−0.15	0.05	10.06	.002	−0.23 to −0.06
Intervention[g]	0.22	0.09	5.58	.02	0.04 to 0.40
Day×intervention	0.01	0.01	0.51	.47	−0.01 to 0.02
Exploratory model					
Evening[a]	−0.11	0.11	1.05	.31	−0.33 to 0.10
Prior assessment[b]	0.02	0.04	0.46	.50	−0.05 to 0.10
Day[c]	−0.05	0.07	0.52	.47	−0.20 to 0.09
Prior missing[d]	0.04	0.07	0.33	.57	−0.10 to 0.17
Age[e]	0.01	0.36	0.00	.98	−0.69 to 0.71
Depression[f]	−0.15	0.05	9.46	.002	−0.24 to −0.05
Intervention[g]					
Engagement	0.16	0.14	1.29	.26	−0.11 to 0.42
Awareness	0.39	0.11	11.99	.001	0.17 to 0.62
Openness	0.10	0.14	0.52	.47	−0.17 to 0.38

[a]Evening (1) indicates prior assessment was in the evening.

[b]Prior assessment is a count of assessments completed.

[c]Day was centered at 21.5.

[d]Prior missing is the count of prior missing data points.

[e]Age is coded 1 for 19 years old and 0 for 18 years old.

[f]Depression is the depressive symptoms reported at the prior assessment.

[g]Intervention (1) is randomization to intervention.

Table 3. Predictors of avoidance-focused energy.

Predictor	Estimate	SE	χ^2_1	*P* value	95% CI
Evening[a]	−0.47	0.14	11.17	.001	−0.74 to −0.19
Prior assessment[b]	0.01	0.03	0.23	.63	−0.04 to 0.07
Day[c]	−0.03	0.06	0.22	.64	−0.14 to 0.09
Prior missing[d]	0.05	0.05	1.19	.28	−0.04 to 0.15
Age[e]	0.18	0.22	0.67	.41	−0.25 to 0.06
Depression[f]	0.32	0.04	68.24	<.001	0.24 to 0.40
Intervention[g]	−0.07	0.06	1.40	.24	−0.19 to 0.05
Day×intervention	−0.003	0.01	0.19	.67	−0.01 to 0.01

[a]Evening (1) indicates prior assessment was in the evening.

[b]Prior assessment is a count of assessments completed.

[c]Day was centered at 21.5.

[d]Prior missing is the count of prior missing data points.

[e]Age is coded 1 for 19-years old and 0 for 18 years old.

[f]Depression is the depressive symptoms reported at the prior assessment.

[g]Intervention (1) is randomization to intervention.

Table 4. Predictors of depressive symptoms, including primary analysis and exploratory analysis of intervention type.

Predictor	Estimate	SE	χ^2_1	*P* value	95% CI
Primary model					
Evening[a]	−0.34	0.16	4.33	.04	−0.66 to −0.02
Prior assessment[b]	0.01	0.02	0.62	.43	−0.02 to 0.05
Day[c]	−0.02	0.04	0.22	.64	−0.09 to 0.05
Prior missing[d]	0.01	0.03	0.20	.66	−0.05 to 0.07
Age[e]	−0.03	0.16	0.03	.87	−0.34 to 0.28
Depression[f]	0.47	0.07	43.97	<.001	0.33 to 0.61
Intervention[g]	−0.22	0.08	8.56	.003	−0.36 to −0.07
Day×intervention	−0.01	0.01	1.45	.23	−0.02 to 0.004
Exploratory model					
Evening[a]	−0.34	0.16	4.32	.04	−0.65 to −0.02
Prior assessment[b]	0.01	0.02	0.63	.43	−0.02 to 0.05
Day[c]	−0.02	0.03	0.35	.55	−0.09 to 0.05
Prior missing[d]	0.01	0.03	0.19	.67	−0.05 to 0.07
Age[e]	−0.02	0.16	0.02	.88	−0.33 to 0.28
Depression[f]	0.47	0.07	44.11	<.001	0.33 to 0.61
Intervention[g]					
Engagement	−0.04	0.08	0.22	.64	−0.20 to 0.12
Awareness	−0.38	0.12	9.58	.002	−0.62 to −0.14
Openness	−0.23	0.10	5.08	.02	−0.43 to −0.03

[a]Evening (1) indicates prior assessment was in the evening.

[b]Prior assessment is a count of assessments completed.

[c]Day was centered at 21.5.

[d]Prior missing is the count of prior missing data points.

[e]Age is coded 1 for 19 years old and 0 for 18 years old.

[f]Depression is the depressive symptoms reported at the prior assessment.

[g]Intervention (1) is randomization to intervention.

Table 5. Predictors of perceived stress symptoms.

Predictor	Estimate	SE	χ^2_1	*P* value	95% CI
Evening[a]	−0.46	0.15	8.82	.003	−0.76 to −0.16
Prior assessment[b]	0.02	0.04	0.19	.66	−0.05 to 0.09
Day[c]	0.01	0.07	0.01	.92	−0.13 to 0.14
Prior missing[d]	−0.02	0.07	0.07	.80	−0.15 to 0.12
Age[e]	0.21	0.37	0.32	.57	−0.51 to 0.93
Depression[f]	0.65	0.12	32.20	<.001	0.43 to 0.88
Intervention[g]	−0.38	0.20	3.72	.05	−0.77 to 0.01
Day×intervention	0.01	0.01	0.18	.67	−0.02 to 0.03

[a]Evening (1) indicates prior assessment was in the evening.

[b]Prior assessment is a count of assessments completed.

[c]Day was centered at 21.5.

[d]Prior missing is the count of prior missing data points.

[e]Age is coded 1 for 19 years old and 0 for 18 years old.

[f]Depression is the depressive symptoms reported at the prior assessment.

[g]Intervention (1) is randomization to intervention.

Discussion

Principal Findings

Given the treatment gaps in mental health, particularly among college students in the United States, there is a substantial need for accessible, empirically based treatments. mHealth interventions are one method for translating traditional psychotherapies into a modality that may improve accessibility and reach. The present MRT examined the safety, feasibility, and preliminary effectiveness of an ACT-based mHealth intervention. Twice-daily assessments were administered in-app, and if completed, participants were randomized to intervention or no intervention. Safety analyses indicated that on average, depressive symptoms decreased from baseline to 3-month follow-up. Feasibility findings indicated that participants responded at least once per day to an in-app assessment at a greater rate (71%) than the authors' prespecified target of 60%. The findings indicated that intervention was associated with increased values-focused energy and decreased depressive symptoms at the next assessment. Moreover, exploratory analyses indicated that awareness interventions predicted increased values-focused energy, and awareness and openness interventions predicted decreased depressive symptoms.

This study provides preliminary support for an mHealth ACT-based microintervention. The findings indicate that an intervention of this magnitude may be appropriate for FGCS reporting distress more days than not. The most important to establish in this study was that participants were, on average, responsive to in-app assessments, a prerequisite to receiving an intervention. As such, the findings herein provide strong data to support adherence to the intervention among college students. Although the intervention was deemed safe in terms of a general decrease in depressive symptoms, an intervention of this magnitude may not be sufficient. Future research should investigate *for whom* an ACT-based mHealth intervention is most helpful and perhaps examine a stepped care model, wherein the mHealth intervention is followed by an increased level of care for those who remain in need. The findings align with prior work, indicating that ACT is impactful in brief intervals [38,39,41,42], via internet-delivered interventions [32-38], via mobile app [28,29,40], and with college students [36,38,59].

Importantly, traditional randomized trials have limitations in terms of conclusions that can be drawn, wherein individuals are randomized once (eg, ACT vs treatment-as-usual). The results would, in this case, offer information about the cumulative effectiveness of the ACT sessions as compared to the treatment-as-usual condition. Microrandomized interventions, however, offer more granular and proximal information about individual interventions. Twice-daily assessments, followed by randomization, allow for examination of the impact of intervention on the assessment immediately following (eg, intervention in the morning and assessment in the evening). Beyond the effectiveness of the intervention, one can also investigate *when*, *for whom*, and *what type* of interventions are most effective. The findings indicated that interventions were not more effective as the days of intervention increased (day×intervention interaction). However, it appeared that interventions delivered after the evening assessment were more predictive of decreased avoidance-focused energy, depressive symptoms, and perceived stress, but not with increased values-focused energy. Moreover, intervention effectiveness did not vary by depressive symptoms at the assessment immediately preceding intervention (interaction), although depressive symptoms preceding intervention were significantly predictive of all outcomes in the expected direction. In terms of *for whom* the intervention is most effective, the findings indicated that the impact of intervention did not vary by age (18 and 19 years old), sex (female and male), or race (White and non-White).

Finally, exploratory analyses suggested that awareness interventions were most impactful for increasing values-focused energy, and awareness and openness interventions were most impactful for decreasing depressive symptoms. Such findings provide granular information on the tailoring of mHealth interventions that will provide a strong foundation for future studies investigating the effectiveness in a larger sample. Furthermore, increased information on for whom the intervention may be most effective will be greatly helped by diversifying the sample in terms of age, gender identity, race, and ethnicity. Moreover, future investigations may address to what extent findings are similar or different across students of different years, and the role of psychiatric history might be investigated.

It is important to note the timing of the study in relation to the COVID-19 pandemic. The first wave of participants was recruited during the fall of 2019, and as such, it is possible that some portion of the 6-week intervention, if not the 3- and 6-month follow-up assessments occurred during the rise of COVID-19 in the United States (March 2020). Another recruitment wave occurred in the fall of 2021. The COVID-19 pandemic has impacted university students' mental health [60,61]. Although the participants may have experienced increased stress over the course of the study, ACT was deemed impactful for university students during the pandemic [38]. Given the ongoing mental, physical, and social impacts of the COVID-19 pandemic, the importance of mHealth interventions for college students cannot be overstated.

Limitations

This study has several limitations that should be considered. The study did not include a diagnostic interview to determine whether symptoms were clinically significant or warranting diagnosis, and instead, measurements relied on self-report of symptoms. Moreover, the twice-daily in-app assessments were based on short forms of validated assessments but were thus limited in scope to the specific symptoms assessed. Future research should consider assessing a wider breadth of symptoms. Assessment of values-focused energy and avoidance energy was conducted with questions designed for this study, and as such, the questions are not part of a validated scale. Future research should identify valid measures of the amount of energy one devotes to values and avoidance. In addition, twice-daily assessments may not be feasible outside of a research context, and future explorations of feasibility in real-world settings may be helpful following a larger efficacy trial. The sample was also limited to 18- and 19-year-old FGCSs reporting distress more days than not at screening, and thus the generalizability of the findings is limited. Further, many eligible students did not respond to email communications after filling the eligibility survey; future studies may consider alternative follow-up communications methods after a web-based screening survey. Finally, significant attrition occurred between the end of the intervention and follow-up assessments, and as such, safety data were based on a subset of the sample. It is unclear how these responders may differ from nonresponders in the follow-up assessment.

Conclusions

The MRT findings presented herein provide preliminary support for the safety, feasibility, and effectiveness of a microrandomized ACT-based mHealth intervention for distressed FGCSs. Amidst large treatment gaps in the United States, particularly among college students, the identification of accessible interventions is an important step toward reducing treatment gaps and increasing the reach of empirically based interventions. Moreover, among at-risk groups, including college students generally and FGCSs specifically, interventions of this sort may be examined for preventive properties. Such an approach may improve students' skills in navigating adversity before it arises, thereby improving resilience, awareness of one's skills, and competencies in the management of distress.

Acknowledgments

We would like to extend our gratitude to the research participants who made this research possible with their continued generosity and dedication. We also acknowledge Rena Doyle, Melvin McInnis, and Susan Murphy for their vital contributions to this work. The funding for the FGCS cohort study was provided by the Baldwin Wisconsin Idea Endowment, Seed Project Grant, the National Institute of Mental Health (K01 MH112876), and the Clinical and Translational Science Award (CSTA) program through the NIH National Center for Advancing Translational Sciences (NCATS) (grant ULITR002373).

Conflicts of Interest

ZNS has received research and salary support from the National Institute of Health and the Center for Disease Control and has consulted for Sage Therapeutics. EBKT has consulted with Macmillan Learning on work unrelated to this research.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 372 KB - mental_v10i1e43065_app1.pdf]

References

1. Kohn R, Ali AA, Puac-Polanco V, Figueroa C, López-Soto V, Morgan K, et al. Mental health in the Americas: an overview of the treatment gap. Rev Panam Salud Publica 2018;42:e165 [FREE Full text] [doi: 10.26633/RPSP.2018.165] [Medline: 31093193]

2. Leong FT, Kalibatseva Z. Cross-cultural barriers to mental health services in the United States. Cerebrum 2011;2011:5 [FREE Full text] [Medline: 23447774]
3. Alonso J, Liu Z, Evans-Lacko S, Sadikova E, Sampson N, Chatterji S, WHO World Mental Health Survey Collaborators. Treatment gap for anxiety disorders is global: results of the world mental health surveys in 21 countries. Depress Anxiety 2018;35(3):195-208 [FREE Full text] [doi: 10.1002/da.22711] [Medline: 29356216]
4. Evans-Lacko S, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, Benjet C, Bruffaerts R, et al. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO world mental health (WMH) surveys. Psychol Med 2018;48(9):1560-1571 [FREE Full text] [doi: 10.1017/S0033291717003336] [Medline: 29173244]
5. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004;82(11):858-866 [FREE Full text] [Medline: 15640922]
6. Lee H, Singh GK. Monthly trends in access to care and mental health services by household income level during the COVID-19 pandemic, United States, April: December 2020. Health Equity 2021;5(1):770-779 [FREE Full text] [doi: 10.1089/heq.2021.0036] [Medline: 34909547]
7. Adams-Prassl A, Boneva T, Golin M, Rauh C. The impact of the coronavirus lockdown on mental health: evidence from the United States. Econ Policy 2022;37(109):139-155. [doi: 10.1093/epolic/eiac002]
8. Reinert M, Fritze D, Nguyen T. The state of mental health in America 2022. Mental Health America. 2021. URL: https://mhanational.org/sites/default/files/2022%20State%20of%20Mental%20Health%20in%20America.pdf [accessed 2023-04-14]
9. Kadison R, DiGeronimo TF. College of the Overwhelmed: The Campus Mental Health Crisis and What To Do About It. San Francisco, CA: Jossey-Bass; 2004.
10. Kruisselbrink Flatt A. A suffering generation: six factors contributing to the mental health crisis in North American higher education. Coll Q 2013;16(1):n1 [FREE Full text]
11. Castillo LG, Schwartz SJ. Introduction to the special issue on college student mental health. J Clin Psychol 2013;69(4):291-297. [doi: 10.1002/jclp.21972] [Medline: 23381839]
12. Ebert DD, Mortier P, Kaehlke F, Bruffaerts R, Baumeister H, Auerbach RP, WHO World Mental Health-International College Student Initiative collaborators. Barriers of mental health treatment utilization among first-year college students: first cross-national results from the WHO world mental health international college student initiative. Int J Methods Psychiatr Res 2019;28(2):e1782 [FREE Full text] [doi: 10.1002/mpr.1782] [Medline: 31069905]
13. Xiao H, Carney DM, Youn SJ, Janis RA, Castonguay LG, Hayes JA, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
14. Cortelyou-Ward K, Rotarius T, Honrado JC. Using technology to improve access to mental health services. Health Care Manag 2018;37(2):101-108. [doi: 10.1097/HCM.0000000000000211] [Medline: 29557818]
15. Hogan MF, Goldman ML. New opportunities to improve mental health crisis systems. Psychiatr Serv 2021;72(2):169-173. [doi: 10.1176/appi.ps.202000114] [Medline: 32988327]
16. Lattie EG, Lipson SK, Eisenberg D. Technology and college student mental health: challenges and opportunities. Front Psychiatry 2019;10:246 [FREE Full text] [doi: 10.3389/fpsyt.2019.00246] [Medline: 31037061]
17. Rudd BN, Beidas RS. Digital mental health: the answer to the global mental health crisis? JMIR Ment Health 2020;7(6):e18472 [FREE Full text] [doi: 10.2196/18472] [Medline: 32484445]
18. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: a pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
19. Rajagopalan A, Shah P, Zhang MW, Ho RC. Digital platforms in the assessment and monitoring of patients with bipolar disorder. Brain Sci 2017;7(11):150 [FREE Full text] [doi: 10.3390/brainsci7110150] [Medline: 29137156]
20. Rizvi SL, Hughes CD, Thomas MC. The DBT Coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: a preliminary evaluation and challenges to client utilization. Psychol Serv 2016;13(4):380-388. [doi: 10.1037/ser0000100] [Medline: 27797571]
21. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: there is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
22. Wright JH, Mishkind M. Computer-assisted CBT and mobile apps for depression: assessment and integration into clinical care. Focus 2020;18(2):162-168 [FREE Full text] [doi: 10.1176/appi.focus.20190044] [Medline: 33162853]
23. Economides M, Ranta K, Nazander A, Hilgert O, Goldin PR, Raevuori A, et al. Long-term outcomes of a therapist-supported, smartphone-based intervention for elevated symptoms of depression and Anxiety: quasiexperimental, pre-postintervention study. JMIR Mhealth Uhealth 2019;7(8):e14284 [FREE Full text] [doi: 10.2196/14284] [Medline: 31452521]
24. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

25. Levin ME, MacLane C, Daflos S, Seeley JR, Hayes SC, Biglan A, et al. Examining psychological inflexibility as a transdiagnostic process across psychological disorders. J Contextual Behav Sci 2014;3(3):155-163 [FREE Full text] [doi: 10.1016/j.jcbs.2014.06.003] [Medline: 29057212]
26. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: The Process and Practice of Mindful Change. New York: Guilford Press; 2011.
27. Gloster AT, Walder N, Levin ME, Twohig MP, Karekla M. The empirical status of acceptance and commitment therapy: a review of meta-analyses. J Contextual Behav Sci 2020;18:181-192. [doi: 10.1016/j.jcbs.2020.09.009]
28. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
29. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
30. Järvelä-Reijonen E, Karhunen L, Sairanen E, Muotka J, Lindroos S, Laitinen J, et al. The effects of acceptance and commitment therapy on eating behavior and diet delivered through face-to-face contact and a mobile app: a randomized controlled trial. Int J Behav Nutr Phys Act 2018;15(1):22 [FREE Full text] [doi: 10.1186/s12966-018-0654-8] [Medline: 29482636]
31. Merwin RM, Moskovich AA, Babyak M, Feinglos M, Honeycutt LK, Mooney J, et al. An open trial of app-assisted acceptance and commitment therapy (iACT) for eating disorders in type 1 diabetes. J Eat Disord 2021;9(1):6 [FREE Full text] [doi: 10.1186/s40337-020-00357-6] [Medline: 33407910]
32. Bendig E, Bauereiss N, Schmitt A, Albus P, Baumeister H. ACTonDiabetes-a guided psychological internet intervention based on acceptance and commitment therapy (ACT) for adults living with type 1 or 2 diabetes: results of a randomised controlled feasibility trial. BMJ Open 2021;11(7):e049238 [FREE Full text] [doi: 10.1136/bmjopen-2021-049238] [Medline: 34244277]
33. Fauth EB, Novak JR, Levin ME. Outcomes from a pilot online acceptance and commitment therapy program for dementia family caregivers. Aging Ment Health 2022;26(8):1620-1629. [doi: 10.1080/13607863.2021.1942432] [Medline: 34233133]
34. Han A, Yuen HK, Jenkins J, Yun Lee H. Acceptance and commitment therapy (ACT) guided online for distressed caregivers of persons living with dementia. Clin Gerontol 2022;45(4):927-938. [doi: 10.1080/07317115.2021.1908475] [Medline: 33794127]
35. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-based acceptance and commitment therapy for depressive symptoms with minimal support: a randomized controlled trial. Behav Modif 2015;39(6):805-834. [doi: 10.1177/0145445515598142] [Medline: 26253644]
36. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: a randomized controlled clinical trial. Behav Res Ther 2016;78:30-42. [doi: 10.1016/j.brat.2016.01.001] [Medline: 26848517]
37. Räsänen P, Muotka J, Lappalainen R. Examining mediators of change in wellbeing, stress, and depression in a blended, internet-based, ACT intervention for university students. Internet Interv 2020;22:100343. [doi: 10.1016/j.invent.2020.100343]
38. Pang NTP, Tio VCS, Singh ASB, Tseu MWL, Shoesmith WD, Abd Rahim MA, et al. Efficacy of a single-session online ACT-based mindfulness intervention among undergraduates in lockdown amid COVID-19 pandemic. Trends Psychiatry Psychother 2021;44:e20200172 [FREE Full text] [doi: 10.47626/2237-6089-2020-0172] [Medline: 34392668]
39. Fiedorowicz JG, Dindo L, Ajibewa T, Persons J, Marchman J, Holwerda SW, et al. One-day acceptance and commitment therapy (ACT) workshop improves anxiety but not vascular function or inflammation in adults with moderate to high anxiety levels in a randomized controlled trial. Gen Hosp Psychiatry 2021;73:64-70 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.09.009] [Medline: 34619441]
40. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. Behav Modif 2017;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]
41. Kroska EB, Roche AI, O'Hara MW. How much is enough in brief acceptance and commitment therapy? A randomized trial. J Contextual Behav Sci 2020;15:235-244. [doi: 10.1016/j.jcbs.2020.01.009]
42. Dindo L, Zimmerman MB, Hadlandsmyth K, StMarie B, Embree J, Marchman J, et al. Acceptance and commitment therapy for prevention of chronic postsurgical pain and opioid use in at-risk veterans: a pilot randomized controlled study. J Pain 2018;19(10):1211-1221 [FREE Full text] [doi: 10.1016/j.jpain.2018.04.016] [Medline: 29777950]
43. Pedrelli P, Nyer M, Yeung A, Zulauf C, Wilens T. College students: mental health problems and treatment considerations. Acad Psychiatry 2015;39(5):503-511 [FREE Full text] [doi: 10.1007/s40596-014-0205-9] [Medline: 25142250]
44. Stebleton MJ, Soria KM, Huesman RL. First-generation students' sense of belonging, mental health, and use of counseling services at public research universities. J Coll Couns 2014;17(1):6-20. [doi: 10.1002/j.2161-1882.2014.00044.x]
45. Walpole M. Socioeconomic status and college: how SES affects college experiences and outcomes. Rev High Educ 2003;27(1):45-73. [doi: 10.1353/rhe.2003.0044]

46. Engle J. Postsecondary access and success for first-generation college students. Am Acad 2007;3(1):25-48 [FREE Full text]
47. Kroska EB, Hoel S, Victory A, Murphy SA, McInnis MG, Stowe ZN, et al. Optimizing an acceptance and commitment therapy microintervention via a mobile app with two cohorts: protocol for micro-randomized trials. JMIR Res Protoc 2020;9(9):e17086 [FREE Full text] [doi: 10.2196/17086] [Medline: 32965227]
48. Ramón-Arbués E, Gea-Caballero V, Granada-López JM, Juárez-Vela R, Pellicer-García B, Antón-Solanas I. The prevalence of depression, anxiety and stress and their associated factors in college students. Int J Environ Res Public Health 2020;17(19):7001 [FREE Full text] [doi: 10.3390/ijerph17197001] [Medline: 32987932]
49. Polk KL, Schoendorff B, Webster M, Olaz FO. The Essential Guide to the ACT Matrix : A Step-by-Step Approach to Using the ACT Matrix Model in Clinical Practice. Oakland, CA: New Harbinger Publications; 2016.
50. Kroenke K, Spitzer RL, Williams JBW. The patient health questionnaire-2: validity of a two-item depression screener. Med Care 2003;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
51. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th ed. Washington, DC: American Psychiatric Publishing; 2013.
52. Cohen S, Williamson G. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health: The Claremont Symposium on Applied Social Psychology. Newbury Park, CA: SAGE Publications; 1988:31-67.
53. Strosahl KD, Robinson PA, Gustavsson T. Brief Interventions for Radical Change: Principles and Practice of Focused Acceptance and Commitment Therapy. Oakland, CA: New Harbinger Publications; 2012.
54. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. J Am Stat Assoc 2018;113(523):1112-1121 [FREE Full text] [doi: 10.1080/01621459.2017.1305274] [Medline: 30467446]
55. Klasnja P, Smith S, Seewald N, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of heartSteps. Ann Behav Med 2019;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
56. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Softw 2005;15(2):1-11. [doi: 10.18637/jss.v015.i02]
57. Pan W. Akaike's information criterion in generalized estimating equations. Biometrics 2001;57(1):120-125. [doi: 10.1111/j.0006-341x.2001.00120.x] [Medline: 11252586]
58. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2013. URL: https://www.r-project.org/ [accessed 2023-04-14]
59. Grégoire S, Lachance L, Bouffard T, Dionne F. The use of acceptance and commitment therapy to promote mental health and school engagement in university students: a multisite randomized controlled trial. Behav Ther 2018;49(3):360-372. [doi: 10.1016/j.beth.2017.10.003] [Medline: 29704966]
60. Copeland WE, McGinnis E, Bai Y, Adams Z, Nardone H, Devadanam V, et al. Impact of COVID-19 pandemic on college student mental health and wellness. J Am Acad Child Adolesc Psychiatry 2021;60(1):134-141.e2 [FREE Full text] [doi: 10.1016/j.jaac.2020.08.466] [Medline: 33091568]
61. Wang X, Hegde S, Son C, Keller B, Smith A, Sasangohar F. Investigating mental health of US college students during the COVID-19 pandemic: cross-sectional survey study. J Med Internet Res 2020;22(9):e22817 [FREE Full text] [doi: 10.2196/22817] [Medline: 32897868]

Abbreviations

ACT: acceptance and commitment therapy
FGCS: first-generation college student
GEE: generalized estimating equation
Lorevimo: Log, Review, and Visualize your Mood
mHealth: mobile health
MRT: microrandomized trial
PHQ-9: Patient Health Questionnaire–9

Edited by J Torous; submitted 01.10.22; peer-reviewed by D Eisenberg, T Ranjan; comments to author 06.02.23; revised version received 21.03.23; accepted 28.03.23; published 15.05.23.

Please cite as:
Thomas EBK, Sagorac Gruichich T, Maronge JM, Hoel S, Victory A, Stowe ZN, Cochran A
Mobile Acceptance and Commitment Therapy With Distressed First-Generation College Students: Microrandomized Trial
JMIR Ment Health 2023;10:e43065
URL: https://mental.jmir.org/2023/1/e43065
doi:10.2196/43065
PMID:37184896

Original Paper

Feasibility, Perceived Impact, and Acceptability of a Socially Assistive Robot to Support Emotion Regulation With Highly Anxious University Students: Mixed Methods Open Trial

A Jess Williams[1], PhD; Maureen Freed[2], MPhil; Nikki Theofanopoulou[1], MSc; Claudia Daudén Roquet[1], PhD; Predrag Klasnja[3], PhD; James Gross[4], PhD; Jessica Schleider[5], PhD; Petr Slovak[1], PhD

[1]Department of Informatics, King's College London, London, United Kingdom
[2]Psychodynamic Studies, University of Oxford, Oxford, United Kingdom
[3]School of Information, University of Michigan, Ann Arbor, MI, United States
[4]Psychophysiology Laboratory, University of Stanford, Stanford, CA, United States
[5]Department of Psychology, Stony Brook University, New York, NY, United States

Corresponding Author:
A Jess Williams, PhD
Department of Informatics
King's College London
Strand Campus
London, WC2R 2LS
United Kingdom
Phone: 44 7969432592
Email: amy_jess.williams@kcl.ac.uk

Abstract

Background: Mental health difficulties among university students have been rising rapidly over the last decade, and the demand for university mental health services commonly far exceeds available resources. Digital interventions are seen as one potential solution to these challenges. However, as in other mental health contexts, digital programs often face low engagement and uptake, and the field lacks usable, engaging, evidence-supported mental health interventions that may be used flexibly when students need them most.

Objective: The aim of this study is to investigate the feasibility and acceptability of a new, in situ intervention tool (Purrble) among university students experiencing anxiety. As an intervention, Purrble was designed to provide in situ support for emotion regulation (ER)—a well-known transdiagnostic construct—directly in the moments when individuals are facing emotionally challenging situations. A secondary aim is to consider the perceived impact of Purrble on youth mental health, as reported by students over a 7-week deployment.

Methods: A mixed methods open trial was conducted with 78 under- and postgraduate students at Oxford University. Participants were recruited based on moderate to high levels of anxiety measured by Generalized Anxiety Disorder-7 at baseline (mean 16.09, SD 3.03). All participants had access to Purrble for 7 weeks during the spring term with data on their perceived anxiety, emotion dysregulation, ER self-efficacy, and engagement with the intervention collected at baseline (pre), week 4 (mid), and week 8 (postintervention). Qualitative responses were also collected at the mid- and postintervention points.

Results: The findings demonstrated a sustained engagement with Purrble over the 7-week period, with the acceptability further supported by the qualitative data indicating that students accepted Purrble and that Purrble was well-integrated into their daily routines. Exploratory quantitative data analysis indicated that Purrble was associated with reductions in student anxiety (dz=0.96, 95% CI 0.62-1.29) and emotion dysregulation (dz=0.69, 95% CI 0.38-0.99), and with an increase in ER self-efficacy (dz=–0.56, 95% CI –0.86 to –0.26).

Conclusions: This is the first trial of a simple physical intervention that aims to provide ongoing ER support to university students. Both quantitative and qualitative data suggest that Purrble is an acceptable and feasible intervention among students, the engagement with which can be sustained at a stable level across a 7-week period while retaining a perceived benefit for those who use it (n=32, 61% of our sample). The consistency of use is particularly promising given that there was no clinician engagement

or further support provided beyond Purrble being delivered to the students. These results show promise for an innovative intervention model, which could be complementary to the existing interventions.

(JMIR Ment Health 2023;10:e46826) doi:10.2196/46826

KEYWORDS

emotion regulation; students; anxiety; digital intervention; mixed-methods

Introduction

Background

Globally, the prevalence of mental health difficulties among students is a major concern [1,2]. Recent epidemiological studies estimate that about one-fifth (20.3%) of all university students are affected by mental health disorders every year [1,3,4] with the incidence rates rising substantially over the last 10 years [5]. In particular, anxiety disorders have been highlighted as a concern among student populations; for example, a recent review across 21 countries shows 11.7%-14.7% of the cross-national population being impacted [1]. In recent years, this concern has grown as generalized anxiety disorder alone is now thought to be affecting 16.7%-18.6% of students over 12 months to lifetime prevalence [3]. Alongside anxiety disorders, students are known to struggle with mood and substance use disorders [1-3] that can be associated with the onset of university education [1].

Traditional university services have struggled to meet the increasing needs of this population [6-8], with student demand for mental health services far exceeding changes in enrollment numbers [9,10]. Combined with difficulties such as underresourced counseling centers and lack of staff capacity, students often face extended waiting times when seeking help [5,11]. This may further exacerbate symptoms of mental ill-health and discourage future help-seeking [12].

Digital Mental Health Interventions for University Students

One potential solution to these challenges involves the use of digital technologies that are generally regarded positively by university students [13]. Prior work has shown that web-based service provision can be efficacious with a recent systematic review [14] showing small intervention effects on anxiety (Hedges g=0.27), depression (g=0.18), and stress (g=0.20) in randomized controlled trials, amid substantial heterogeneity and inclusion of high risk of bias studies in the sample. The 48 intervention programs reviewed were typically delivered through websites (89.6%), focused on cognitive behavioral therapy (or third-wave cognitive behavioral therapy) skills training (66%), and relied on a passive control group (eg, waitlist or psychoeducation materials for behavior change) [14]. We note, however, that these effects are smaller than those seen in similar systematic reviews of interventions for general populations (eg, for digital anxiety treatments, g=0.62 [15]).

Moreover, the implementation of digital programs within routine services remains challenging in student populations [16-18]. Pragmatic trials on web-based interventions show low engagement and uptake rates [19-22] similar to web-based mental health interventions in other domains [23]. To overcome such a lack of participant engagement, there is a need to develop usable, engaging, evidence-supported mental health interventions that may be used flexibly based on when students need them most (eg, when stress levels are particularly high and coping skills most warrant deployment).

Emotion Regulation and Emerging Digital Interventions

Emotion regulation (ER) is a well-known transdiagnostic factor for a range of mental health difficulties [24-27]. Across young people, difficulties with ER have been associated with a range of mental health difficulties, such as eating disorders [28], anxiety and depression [29], as well as self-harm and suicidal thoughts [30]. In student populations particularly, difficulties with ER have been found to be moderately severe [31,32], with dysregulation related to practices such as maladaptive perfectionism [33] and coping [34].

While the research on digitally mediated interventions targeting ER is limited to date (cf, recent review [35]), there is one emerging intervention tool—"Purrble"—which was designed to provide in-the-moment ER support in daily life [24,36,37] (see Intervention section for more details). Purrble was originally developed for use with child populations, and there is emerging evidence that it could deliver significant benefits through improved ER [24,36,37]. For example, Purrble has been shown to facilitate positive parental-child interactions [24,36], support children to self-soothe in the moment [24,37], and enable conversations around children's emotions [37]. Within interviews, parents and children recognized that Purrble was incorporated into perceived ER practices (such as disengagement or distraction), which helped to calm or settle the child in emotion-eliciting situations [38]; the children also remained engaged with Purrble across the studies [24,37,38]. From a theory-of-change perspective, Purrble is based on the Gross Process model of ER [39]: it has been designed to impact two separate stages of the ER process, (1) the attention deployment stage, whereby an individual's attention is shifted from the emotion-eliciting situations toward interacting with the toy; and (2) the response modulation stage, by facilitating downregulation through pleasant tactile interactions—see prior work [36] for further details.

In summary, while the emerging evidence—together with the theory of change—suggests that Purrble could support ER across a broader age range, no work has so far investigated if and how older populations such as university students might engage with Purrble during stressful periods (eg, during the college term) and how this may impact prevalent mental health concerns, such as anxiety.

This Study

The aim of this feasibility and acceptability study is to evaluate a novel digital intervention device, Purrble, designed to provide in-the-moment ER support in daily life [24] with high-anxiety students in naturalistic contexts. Therefore, the specific objectives are to determine the feasibility of Purrble as an emotion regulatory intervention by considering retention rate, engagement measures, and perceived impact on mental health; to assess the acceptability of Purrble by exploring students' perceptions and experiences of engaging and appropriating the device using the open-text responses to mid- and postdeployment surveys; and to determine the association between Purrble deployment and students' anxiety symptoms, ER difficulties, and ER beliefs.

Methods

This study is an open trial deployment of an in situ intervention to enhance in-the-moment ER in high-anxiety university students.

Ethics Approval

Ethics approval was obtained from the University of Oxford ethics board (CUREC: 310221) prior to the start of the investigation.

Recruitment

Participants were recruited using advertisements in a university-wide newsletter that included a link to register interest in the study. To take part, interested individuals completed a short screening tool made of a battery of validated measures. To be eligible, one needed to be (1) an Oxford University student, (2) 18 years or older, (3) living in Oxford at the start of the study, and (4) score 10 or above on the Generalized Anxiety Disorder-7 (GAD-7), indicating moderate to severe anxiety symptoms. Due to the limited number of intervention devices available, additional criteria were needed for participant selection, offering priority to nonbinary or male participants due to the high volume of interest from cisgender female students and those with severe anxiety. All selected eligible students were sent consent forms to complete.

Intervention

Purrble is a small, inexpensive, web-based plush animal (Figure 1) that aims to guide the user to downregulate their unwanted emotions through a combination of sensors and haptic vibrations, with the theory-of-change targeting the attentional deployment and response modulation components of the ER process [24,25]. The smart toy's internal state is communicated to users through vibration patterns that mimic a heartbeat, with faster rates corresponding to higher "stress" levels. The fundamental game loop is the following: whenever the smart toy wakes up from sleep, it is startled and has a rapid heartbeat. Shakes, sudden movements, or pressing its ears also wake and "startle" the toy, while calm stroking movements and hugs gradually slow the heartbeat, which eventually changes to a purring vibration, indicating a calm, happy state. The toy also makes varying gentle sounds (ie, sighs, coos, giggles, and grunts) that correspond to its internal state and complement the vibration-based feedback. The full description of the intervention including the detailed theory-of-change model is available [36].

What makes Purrble a potentially unique intervention mechanism is the combination of (1) physical form, (2) associated ongoing availability to provide support as and when needed in everyday settings, and (3) the targeting of the process of in-the-moment ER. Such ongoing in-the-moment ER support—if effective—has the potential to create a positive feedback loop and drive further engagement: for example, if a student is able to reduce their anxiety by interacting with Purrble (and thus feels better), it is more likely that Purrble will become embedded into their daily ER practice. In addition, the in-the-moment physical intervention delivery aims to remove the need for any a priori training on the part of the student and makes Purrble an interestingly complementary approach to existing cognitive interventions (which are reliant on training modules while lacking in-the-moment support for skills application).

Figure 1. Intervention tool—Purrble.

Measures

Demographics

Information, such as university, college, student status, location, and gender, was collected as part of the screening process. These were used to ensure inclusion criteria were met.

Feasibility Measures

Digital engagement of Purrble was assessed using Twente Engagement With eHealth Technologies Scale (TWEETS) [40], which consists of 9 items measured on a 5-point Likert scale ranging from strongly disagree (0) to strongly agree (4), with higher scores indicating greater engagement. This measure has been shown to have good reliability, in this sample, Cronbach α was .88.

Within the postdeployment survey, participants were asked whether they felt Purrble had helped or hurt their mental health using a 5-point Likert scale of strongly disagree (1) to strongly agree (5). This provided a bespoke measure of Purrble's perceived impact on their own mental health.

Acceptability Measures

Exploratory open-ended questions were included in the mid- and postintervention surveys. These queried students about situations where Purrble was used, why or why not they used Purrble, perceived emotional impact, usefulness, and personal ER strategies.

Mental Health Measures

Generalized Anxiety Disorder-7

The GAD-7 [41] is a 7-item screening tool designed to assess the presence and severity of generalized anxiety disorder. Individuals are required to rate the frequency of anxiety symptoms on a 4-point Likert scale from not at all (0) to nearly every day (3). Higher scores indicate greater severity of anxiety symptoms, with a threshold of 10 suggesting moderate and 15 implying severe anxiety. In this sample, Cronbach α was .88.

Difficulties in Emotion Regulation Scale-18

The Difficulties in Emotion Regulation Scale-18 (DERS-18) [42] is an 18-item questionnaire that assesses how well participants regulate their emotions. Participants are asked to rate how often each statement applies to them on a 5-point Likert scale ranging from almost never (1) to almost always (5). Higher total scores imply greater difficulties with ER. In this sample, Cronbach α was .91.

Beliefs About Emotion (ER Beliefs)

Following the adaptation of the Implicit Beliefs About Emotions scale [43,44], ER beliefs was used as a 4-item measure to assess personal beliefs about the malleability of emotions. Participants are asked to score these statements using a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). Those with higher scores are thought to hold fixed beliefs about ER, whereas lower scores suggest more malleable beliefs that may be changed. In this sample, Cronbach α was .83.

Procedure

Assessments (overview in Table 1) were conducted on the web using King's College London Qualtrics, with the baseline assessment (anxiety and ER items) taking place the week prior to Purrble deployment and feasibility, acceptability, and mental health measures asked in mid- and postdeployment surveys. Across a 7-week deployment period, participants had access to their individual Purrble with the instructions to use it "as much or as little as they liked." This was to ensure that any engagement was unbiased and offered an understanding of how students appropriated Purrble over time. Optional daily surveys were sent to participants each evening (~5 minutes), asking about their day and Purrble engagement (Multimedia Appendix 1). Other than the bespoke feasibility measure of impact, all feasibility, acceptability, and mental health measures were asked in both mid- and postdeployment surveys.

Following the intervention period, participants were invited to a semistructured interview, which took place via Zoom (Zoom Video Communications). This was to explore their perceptions and experiences of having had Purrble, whether they had engaged with the intervention, and their understanding as to whether Purrble had influenced their ER skills at all.

Table 1. Data collection procedure across the study.

Survey	Predeployment	Deployment							Follow-up
	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Baseline assessment (GAD-7[a], DERS-18[b], ER[c] beliefs)	✓								
Daily surveys		__[d]	__[d]	__[d]	__[d]	__[d]	__[d]	__[d]	
Middeployment survey (GAD-7, DERS-18, ER beliefs, TWEETS[e], exploratory open-ended questions)					✓				
Postdeployment survey (GAD-7, DERS-18, ER beliefs, TWEETS, exploratory open-ended questions)									✓
Interviews with a subset of participants									✓

[a]GAD-7: Generalized Anxiety Disorder-7.

[b]DERS-18: Difficulties in Emotion Regulation Scale-18.

[c]ER: emotion regulation.

[d]Collected daily.

[e]TWEETS: Twente Engagement With eHealth Technologies Scale.

Mixed Methods Analytic Plan

Statistical data were analyzed using RStudio (Posit) and the *lmer* package. For baseline measures, averages and SDs were calculated to give an insight into the characteristics of the participant sample.

Intervention feasibility was assessed in three ways by calculating descriptive statistics: (1) calculating retention rate across the study period as percentages of students who submitted mid- or postquestionnaires, (2) self-reported engagement with the intervention through a total score of the TWEETS measure for those retained in the study, and (3) perceived impact of intervention as a percentage of student responses at postdeployment survey. The daily items asking about Purrble engagement were not analyzed due to their optional status and thus high attrition (>70% by week 8 of the study).

To assess the acceptability of Purrble, iterative deductive thematic analysis was used to explore the perceptions and experiences of engaging and appropriating Purrble, as per their own needs. This was achieved by considering participant responses to open-ended questions collected during the exploratory surveys. The qualitative data were analyzed following the 6 phases outlined by Braun and Clarke [45,46]. Following survey data cleaning, all data were imported into NVivo (version 12; Lumivero), and AJW began deductively coding the data. Codes were clustered to develop a meaningful preliminary framework, which was discussed with PS. Additional considerations and reflections were discussed, leading to a revised thematic framework. This was reviewed and finalized by all authors. The semistructured interviews of this study are analyzed elsewhere [47].

To explore the association between Purrble deployment and students' anxiety symptoms (GAD-7), ER difficulties (DERS-18), and ER beliefs, each of the outcome variables was fitted to progressively more complex linear mixed models to identify the best fit, following the analysis steps from prior analogous work [48].

The starting model predicted the outcome variable as a function of time and allowed for both random slopes and intercept for individual participants (unless this reduced the model fit): <outcome> ~ timepoint + (timepoint | participant). We then included predetermined covariates, that is, engagement (TWEETS) and gender. For DERS-18 and ER beliefs, we also included baseline GAD-7 as another covariate. Restricted maximum likelihood methods were used for estimation (the default approach within lme4), which yields unbiased estimates under the assumption that the missing data mechanism is ignorable (ie, data are missing at random [49]). Given the promising trajectory for all variables, within-group effect sizes (Cohen *d*; 95% CI) were calculated to reflect this change from pre- (week 0) to postdeployment (week 7) on each of the outcome variables.

Results

Sample Characteristics

In total, 80 (42 undergraduates, 36 postgraduates) students were enrolled in the study. Two participants did not return signed consent forms and were excluded. Of the full sample (n=78), most participants identified as female (n=48). The remaining participants identified as male (n=18) and nonbinary (n=10). Two participants did not list their gender identity. The average age was 23.5 (range 18-40) years.

At the start of the trial, participants self-reported having severe anxiety (mean 16.09, SD 3.03) alongside high scores of emotion dysregulation (mean 55.95, SD 12.76) and lower levels of beliefs

that one can control their ER abilities (mean 11.24, SD 3.21). Between genders, symptoms of anxiety were higher among females (mean 18.0, SD 1.75) when compared to male (mean 14.3, SD 2.56) and nonbinary (mean 14.2, SD 2.90) counterparts.

Feasibility

In total, 300 students responded to the study call. As there was a limited number of Purrble units, 80 of the 142 students who met all inclusion criteria were invited to take part based on the additional criteria. In total, 78 participants provided valid consent forms and completed the baseline assessment; therefore, the enrollment rate was 97.5%. Of the 78 students enrolled, 57 responded to the midpoint assessment (week 4), which decreased to 52 respondents at postdeployment (week 8). Therefore, the overall retention rate across the 7-week deployment including follow-up was 66.7%. We do not see differences in baseline GAD-7 scores between those who dropped out (mean 15.9, SD 3.24; n=25) and those who remained in the study (mean 16.2, SD 2.94; n=51). However, there is some indication that female participants were less likely to drop out (6 out of 28) than male participants (13 out of 32) and nonbinary participants (3 out of 10); it is possible that these differences could be linked to the lower baseline levels of anxiety for male and nonbinary participants.

Across the 7-week deployment, self-reported engagement was measured at the midpoint (week 4) and postdeployment (week 8) through TWEETS. At both time points, students reported a moderate level of engagement with the intervention, with a limited difference from week 4 (mean 23.75, SD 5.09) to week 8 (mean 21.79, SD 6.79). Students who responded to these surveys maintained a good degree of engagement with Purrble throughout the deployment.

Purrble was indicated as a feasible intervention, as generally students reported it to have had a positive impact on their mental health. Of those students who remained in the study, 61% (n=32) reported that Purrble had helped their mental health, while only 9% (n=5) disagreed with this statement. Similarly, 93% (n=48) stated that they “strongly disagreed” or “disagreed” with the statement that Purrble had hurt their mental health, and the remaining 7% (n=4) answered neutrally.

Acceptability

Overview

To determine Purrble acceptability as a tool in students’ daily lives, the qualitative responses from the midpoint and posttrial surveys were explored considering the engagement and appropriation of the intervention. Four themes were developed: (1) primary uses for Purrble, (2) appropriation mechanisms, (3) Purrble is “just” a toy, and (4) empathetic responses (Table 2). The themes identified are described later with example quotes. Additional illustrative quotes by subtheme can be found in Multimedia Appendix 2.

Table 2. Thematic framework including the prevalence of instance of subtheme.

Theme and subtheme	Values, n (%)
Primary uses for Purrble	
Tool for anxiety and stress reduction	66 (60.6)
Calming during night routine	27 (24.8)
Prevention of mental health spiral	6 (5.5)
Appropriation mechanisms	
Grounding	25 (22.9)
Mindfulness	13 (11.9)
Self-stimulation	30 (27.5)
Purrble is “just” a toy	
Feelings of embarrassment	11 (10.1)
“If levels are too high”	15 (13.8)
Purrble-related anxiety	21 (19.3)
Empathetic response	
Caring for something else helps one’s ability to self-soothe	32 (29.4)
Guilt relating to “upsetting” Purrble	8 (7.3)

Primary Uses for Purrble

The most frequently reported use for Purrble by students was as a tool for anxiety and stress reduction. The students described using Purrble as a means to cope with anxiety or anxiety-inducing situations, with some indicating specific types of anxiety responses that Purrble was particularly useful for; for example, “Purrble is useful when I have a quality of anxiety that is panicky—by which I mean, heart palpitations and shakes, where my physiological response is in itself problematic” (R_339Y). Purrble was also often incorporated into a calming night routine, either acting as a device to replace negative sleep hygiene practices (“something to focus on that is not a screen, before I go to bed” [R_OveS]) or as a comforting object while sleeping. For a few people, Purrble was used to prevent a mental

health spiral, supporting students in stopping their experiences of poor mental health from becoming more severe and difficult to cope with. Purrble was described as being particularly useful to act as a tool to "snap out of the spiraling nature of invasive thoughts" (R_2EGJ).

Together, these descriptions show students' perceptions of Purrble as an acceptable and effective intervention for anxiety and stress as well as highlight it as a potential mechanism to break the ruminative aspect of mental health difficulties.

Appropriation Mechanism

The students also described a number of mechanisms that led to the positive effects stated earlier. Most commonly, students reported using Purrble as a tangible grounding tool—something that allowed them to center themselves when experiencing specific stressors and thus reduce their anxiety: "when I'm dissociating, I use the Purrble to ground myself" (R_1jfX). Purrble was also used alongside mindfulness processes and acted as a physical reminder and a practical tool to bring them to that mindfulness headspace. Finally, Purrble also seemed to support self-stimulation when anxious, allowing students to relax or calm down through touching or stroking the device: "The action of stroking him rhythmically is very calming, and takes the focus away from the source of panic/stress" (R_25vd).

All of these strategies can be linked to specific ER processes (attention deployment, cognitive change, and response modulation), suggesting that Purrble could have the intended impact of supporting in situ ER.

Purrble Is "Just" a Toy

While most students felt Purrble was helpful to them, some also had reservations about situations where Purrble could be used. This was commonly associated with feelings of embarrassment due to Purrble being perceived as designed for children and therefore was not always thought to be "appropriate" for adult use in public due to assumed judgment or stigma from others.

A small subset of students also felt that Purrble was not an acceptable option if their levels of distress "are too high," either as it would not help ("knowing that it is merely a substitute and not an actual living being" [R_3Rz4]) or because it would be too difficult to engage ("it is hard to make myself reach out for him" [R_10ZQ]). Finally, for a few students, there were moments when Purrble interaction could heighten the student's sensory perceptions, leading to Purrble-related anxiety ("the rare/occasional times where Purrble's heartbeat/sounds trigger my overstimulation/anxiety" [R_28Z2]).

Empathetic Response

Across students, Purrble appeared to induce empathetic responses: the participants sought to comfort Purrble when it was "distressed." The "need" and ability to care for something else helped self-soothing: it allowed the students to detach themselves from their distress and promoted a focus on Purrble, which encouraged students to relax. "It made me feel quite parental and allowed me to get out of my own head for a minute" (R_2Cba). However, the same empathic response sometimes induced feelings of guilt relating to "upsetting" Purrble by waking it to start an interaction.

Exploring Associations of Purrble Deployment and Mental Health Measures

Findings indicate consistent and statistically significant reductions in all 3 outcome measures when comparing baseline with midterm (week 3) and end-of-term (week 7) observations (Tables 3 and 4). Additionally, these improvements over time are consistent across all models regardless of whether correlates such as TWEETS or GAD-7 baseline are factored in.

First, there were large reductions of anxiety (dz=0.96, 95% CI 0.62-1.29 and dav=1.22, 95% CI 0.85-1.57) with post–week 7 anxiety showing an 83% chance of being lower than baseline. Second, the results show medium changes to difficulties with ER (dz=0.69, 95% CI 0.38-0.99 and dav=0.56, 95% CI 0.26-0.85) with postdeployment scores showing a 75% chance of being lower than baseline. Finally, the outcomes showed medium changes to ER beliefs (dz=–0.56, 95% CI –0.86 to –0.26 and dav=–0.51, 95% CI –0.80 to –0.21) with postdeployment showing a 71% chance of being lower than baseline difficulties with ER.

Midpoint engagement score (TWEETS) was not significantly associated with a reduction in anxiety or difficulties with ER scores but was associated with ER beliefs outcomes (with higher TWEETs at baseline leading to a stronger increase in ER beliefs). In contrast, higher baseline anxiety was associated with a lower reduction in difficulties in ER but not with changes in ER beliefs.

Table 3. Results of mixed-linear models considering the impact of Purrble on anxiety symptoms.

	Dependent variable: generalized anxiety disorder
Mid–time point	–4.18[a] (0.63)
Post–time point	–5.38[a] (0.67)
TWEETS[b]: mid–time point	–0.10 (0.09)
Gender	
Nonbinary	–0.45 (1.48)
Not listed	3.44 (1.76)
Female	2.22[c] (1.06)
Constant	17.64[a] (2.44)
Observations	159
Log likelihood	–441.71
Akaike information criterion	901.41
Bayesian information criterion	929.03

[a]*P*<.001.
[b]TWEETS: Twente Engagement With eHealth Technologies Scale.
[c]*P*<.05.

Table 4. Results of mixed-linear model considering the impact of Purrble on emotion regulation.

	Dependent variable	
	Difficulties in Emotion Regulation	Beliefs About Emotion
Mid–time point	–3.42[a] (1.32)	1.31[b] (0.36)
Post–time point	–7.56[b] (1.41)	1.59[b] (0.38)
TWEETS[c]: mid–time point	–0.51 (0.30)	0.235[b] (0.07)
GAD-7[d]: baseline	1.69[a] (0.52)	–0.05 (0.12)
Constant	41.39[b] (10.95)	6.03[e] (2.59)
Observations	159	159
Log likelihood	–584.41	–375.42
Akaike information criterion	1182.81	764.83
Bayesian information criterion	1204.29	786.32

[a]*P*<.01.
[b]*P*<.001.
[c]TWEETS: Twente Engagement With eHealth Technologies Scale.
[d]GAD-7: Generalized Anxiety Disorder-7.
[e]*P*<.05.

Discussion

Principal Findings

This study aimed to determine whether a socially assistive robot, Purrble, would be a feasible and acceptable intervention for highly anxious university students. The overall findings support the feasibility of Purrble and indicate that Purrble was acceptable as an intervention within this sample, offering insights into how students engaged and appropriated Purrble to suit their individual needs. Furthermore, the explorative analysis showed positive within-subject associations of Purrble deployment with reductions in students' self-reported anxiety symptoms and with the perceived improvement of ER.

Around one-third of the student sample disengaged by week 4 of this study, but those who were retained used Purrble consistently across the trial period, a finding consistent across questionnaires, open-text responses, and qualitative interview data [47]. The overall number of retained participants was comparable to other studies [50-52]. Interestingly, Purrble maintained a stable level of engagement among those

retained—as measured by TWEETS score—throughout the deployment with only a minor (and statistically nonsignificant) decrease in use over the 7-week deployment. This is surprising as prior research shows that engagement with digital interventions tends to drop off markedly over time [18,53,54], and ongoing engagement is a key challenge in traditional therapeutic approaches [55]. For example, prior work has reported decreasing rates of app use each week [50], decreasing rates of program adherence over time [51], and only 10.1% of youths completing all full 10 cognitive behavioral therapy sessions [52]. The consistency of Purrble engagement is particularly striking given the lack of any therapist engagement (eg, clinician's calls) at any point during the intervention that would often be deployed to encourage sustained engagement [56]. This supports emerging arguments about the benefits of interventions that draw on extensive user-centered design approaches and focus on addressing the day-to-day needs of target populations [18,57,58]. In summary, both quantitative and qualitative data suggest that Purrble is an acceptable and feasible intervention among students and can sustain a stable-level engagement across a 7-week period with acceptable dropouts while retaining a perceived benefit for those who use it.

Our exploratory analysis also shows large within-subject effects on student anxiety in week 7 (d=0.96), which are akin to results from other recent web-based open trials [52,59-61]. Moreover, the results also show medium effect sizes on the 2 ER constructs: beliefs about own ER self-efficacy (dz=–0.56) and difficulties with ER (dz=–0.69). This provides support for the theory-of-change underpinning Purrble, assuming that repeated in-the-moment downregulation experiences can both improve perceived self-efficacy and mindsets about ER as well as help scaffold more constructive ER practices [47]. We note that what is particularly interesting in this work is that we have observed both changes in the proximal ER constructs and in anxiety even though Purrble intervention does not explicitly contain any cognitive components (eg, cognitive reappraisal or exposure hierarchies), which underpin traditional ER interventions for mood disorders (eg, Unified Protocol [62,63] or ER therapy [64,65]). This suggests that the effects emerging from in-the-moment experiences of Purrble use could be complementary to the traditionally more cognitively oriented interventions with the potential for mutual amplification (eg, where experiential effects of Purrble further support cognitive training and vice versa).

Implications

Given the promising results found from this trial, implications for wider research are considered. First, given the large effect sizes seen in this open trial, it would be prudent to examine Purrble alongside a control treatment within a randomized controlled trial to determine whether Purrble is indeed an effective intervention.

Second, if effective, Purrble deployments could support professional clinical services at universities. One example of this would be Purrble as a complementary resource for those on waitlists. Given that waitlists are long in many countries: in the United Kingdom, for example, 54% of students reported waiting over 3 months to start treatment, and a further 12% reported waiting over a year [66]; additional support for those on waitlists is needed to aid well-being, help-seeking, and perceptions of care [67].

Third, Purrble could support emergent ER skills taught in traditional therapies, such as dialectical behavior therapy [68], serving as an additional support tool provided while on a waitlist, during or between therapeutic sessions, and potentially being of use to maintain skills once therapy is concluded. In particular, dialectical behavior therapy focuses on teaching skills, such as mindfulness, distress tolerance (distraction, reframing the moment, self-soothing, and problem-solving), interpersonal effectiveness, and ER, and seems aligned with the grounding and mindful interaction promoting effects of Purrble as reported by the students in this sample [47].

Limitations

The following limitations of this study need to be considered. First, as an exploratory open trial, the size of the effects could have been affected by a range of time effects. For example, we would expect that the alignment of the study with term time could also lead to a reduction in anxiety scores, for example, due to a reduction of examination-induced anxiety at the end of the term. Additionally, we might see potential reduction-to-mean effects, as the sample was selected on the elevated need of anxiety symptoms at the start of term. Finally, it is possible that the data were also affected by COVID-19-driven cohort effects: at the time the study was conducted, the United Kingdom had recently ended a national lockdown due to COVID-19, which had detrimental effects on young people's—and particularly students'—mental well-being [8]. It is likely that the students were at risk of heightened stress during the study, which might have further amplified the observed effects and perceived need.

Second, the study has not fully explored any gender effects in the acceptability or feasibility of Purrble. Initially, a large number of female students responded to the intervention advert, which prompted the selection of nonfemale responders and clinical need inclusion criteria. The initial disproportionate response from females may reflect a greater receptivity to the intervention concept, or potentially, this reflects a higher incidence of emotional distress, which is often presented as anxiety. For example, females (38%) present more commonly with clinically significant anxiety than males (20.3%) [10]. Alternatively, female participants are often overrepresented in psychological research [69]. As this trial was to assess the feasibility and acceptability of Purrble, a generalization sample across gender identities was selected. Alternatively, in future trials, there may be a benefit to maintaining the clinical threshold but removing gender limits.

Finally, additional mental health difficulties (eg, depression) or life factors (eg, grief) were not measured. Given that mental health difficulties are often comorbid, for example, 27.8% of a recent student sample reported both anxiety and depression [2], capturing such experiences is vital to understanding the impact and effectiveness of Purrble. Thus, it is not currently possible to examine in greater detail whether Purrble was useful for students dealing with these difficulties, which is discussed within

the thematic analysis. It is also unclear whether the ER strategies induced by Purrble translate to these difficulties.

Conclusions

Purrble appears to represent a feasible in situ intervention to aid ER within a university student population. In particular, Purrble was seen to be an acceptable tangible device that was useful in stressful and anxious moments, where students could downregulate their emotions. Moreover, its use was associated with a decrease in anxiety symptoms and an increase in ER competence among anxious college students. From this study, there are several research implications, such as exploring the impact of Purrble compared to a control group to better determine how Purrble is effective as an ER intervention and to what extent within highly anxious student populations. By using a longitudinal study design for this follow-up study, better evidence would be provided regarding the true impact of the intervention compared to this study, which aligned with the end of term time.

Conflicts of Interest

PS has been involved in the development of what is now Purrble as part of his postdoctoral fellowship and serves as a paid research adviser to the Committee for Children (CfC) but has no financial stake in either CfC (Purrble brand owners) or Sproutel (company manufacturing Purrble). CfC has donated the Purrble units used in the study, and Sproutel has not been involved. Neither Committee for Children nor Sproutel had access to the data or had been part of the data collection in any way nor did they approve or see the publication before it was submitted.

Multimedia Appendix 1
Optional daily survey questions.
[DOCX File , 15 KB - mental_v10i1e46826_app1.docx]

Multimedia Appendix 2
Illustrative quotes.
[DOCX File , 16 KB - mental_v10i1e46826_app2.docx]

References

1. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization world mental health surveys. Psychol Med 2016;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
2. Jenkins PE, Ducker I, Gooding R, James M, Rutter-Eley E. Anxiety and depression in a sample of UK college students: a study of prevalence, comorbidity, and quality of life. J Am Coll Health 2021;69(8):813-819. [doi: 10.1080/07448481.2019.1709474] [Medline: 31995452]
3. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, et al. WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018;127(7):623-638 [FREE Full text] [doi: 10.1037/abn0000362] [Medline: 30211576]
4. Blanco C, Okuda M, Wright C, Hasin DS, Grant BF, Liu SM, et al. Mental health of college students and their non-college-attending peers: results from the National Epidemiologic Study on alcohol and related conditions. Arch Gen Psychiatry 2008;65(12):1429-1437 [FREE Full text] [doi: 10.1001/archpsyc.65.12.1429] [Medline: 19047530]
5. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007-2017). Psychiatr Serv 2019;70(1):60-63 [FREE Full text] [doi: 10.1176/appi.ps.201800332] [Medline: 30394183]
6. Huang HCH, Ougrin D. Impact of the COVID-19 pandemic on child and adolescent mental health services. BJPsych Open 2021;7(5):e145 [FREE Full text] [doi: 10.1192/bjo.2021.976] [Medline: 34348823]
7. Shaw H, Robertson S, Ranceva N. What was the impact of a global pandemic (COVID-19) lockdown period on experiences within an eating disorder service? A service evaluation of the views of patients, parents/carers and staff. J Eat Disord 2021;9(1):14 [FREE Full text] [doi: 10.1186/s40337-021-00368-x] [Medline: 33468242]
8. Coronavirus: impact on young people with mental health needs. Young Minds. 2020. URL: https://healtheducationresources.unesco.org/library/documents/coronavirus-impact-young-people-mental-health-needs [accessed 2023-09-26]
9. Duffy ME, Twenge JM, Joiner TE. Trends in mood and anxiety symptoms and suicide-related outcomes among U.S. undergraduates, 2007-2018: evidence from two national surveys. J Adolesc Health 2019;65(5):590-598. [doi: 10.1016/j.jadohealth.2019.04.033] [Medline: 31279724]
10. King N, Pickett W, McNevin SH, Bowie CR, Rivera D, Keown-Stoneman C, et al. Mental health need of students at entry to university: baseline findings from the U-Flourish Student Well-Being and Academic Success Study. Early Interv Psychiatry 2021;15(2):286-295. [doi: 10.1111/eip.12939] [Medline: 32048460]

11. Xiao H, Carney DM, Youn SJ, Janis RA, Castonguay LG, Hayes JA, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
12. Vidourek RA, King KA, Nabors LA, Merianos AL. Students' benefits and barriers to mental health help-seeking. Health Psychol Behav Med 2014;2(1):1009-1022 [FREE Full text] [doi: 10.1080/21642850.2014.963586] [Medline: 25750831]
13. Dunbar MS, Sontag-Padilla L, Kase CA, Seelam R, Stein BD. Unmet mental health treatment need and attitudes toward online mental health services among community college students. Psychiatr Serv 2018;69(5):597-600 [FREE Full text] [doi: 10.1176/appi.ps.201700402] [Medline: 29540117]
14. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: a systematic review and meta-analysis. Int J Methods Psychiatr Res 2019;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
15. Pauley D, Cuijpers P, Papola D, Miguel C, Karyotaki E. Two decades of digital interventions for anxiety disorders: a systematic review and meta-analysis of treatment effectiveness. Psychol Med 2023;53(2):567-579 [FREE Full text] [doi: 10.1017/S0033291721001999] [Medline: 34047264]
16. Harrer M, Apolinário-Hagen J, Fritsche L, Salewski C, Zarski AC, Lehr D, et al. Effect of an internet- and app-based stress intervention compared to online psychoeducation in university students with depressive symptoms: results of a randomized controlled trial. Internet Interv 2021;24:100374 [FREE Full text] [doi: 10.1016/j.invent.2021.100374] [Medline: 33718001]
17. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
18. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
19. Kaltenthaler E, Parry G, Beverley C, Ferriter M. Computerised cognitive-behavioural therapy for depression: systematic review. Br J Psychiatry 2008;193(3):181-184 [FREE Full text] [doi: 10.1192/bjp.bp.106.025981] [Medline: 18757972]
20. Lillevoll KR, Vangberg HCB, Griffiths KM, Waterloo K, Eisemann MR. Uptake and adherence of a self-directed internet-based mental health intervention with tailored e-mail reminders in senior high schools in Norway. BMC Psychiatry 2014;14:14 [FREE Full text] [doi: 10.1186/1471-244X-14-14] [Medline: 24443820]
21. Whiteside U, Lungu A, Richards J, Simon GE, Clingan S, Siler J, et al. Designing messaging to engage patients in an online suicide prevention intervention: survey results from patients with current suicidal ideation. J Med Internet Res 2014;16(2):e42 [FREE Full text] [doi: 10.2196/jmir.3173] [Medline: 24509475]
22. Woodford J, Farrand P, Bessant M, Williams C. Recruitment into a guided Internet based CBT (iCBT) intervention for depression: lesson learnt from the failure of a prevalence recruitment strategy. Contemp Clin Trials 2011;32(5):641-648. [doi: 10.1016/j.cct.2011.04.013] [Medline: 21570485]
23. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
24. Slovák P, Theofanopoulou N, Cecchet A, Cottrell P, Bertran FA, Dagan E, et al. I just let him cry...: designing socio-technical interventions in families to prevent mental health disorders. Proc ACM Hum-Comput Interact 2018;2(CSCW):1-34. [doi: 10.1145/3274429]
25. Gross JJ. Emotion regulation: current status and future prospects. Psychol Inq 2015;26(1):1-26. [doi: 10.1080/1047840X.2014.940781]
26. Sloan E, Hall K, Moulding R, Bryce S, Mildred H, Staiger PK. Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: a systematic review. Clin Psychol Rev 2017;57:141-163 [FREE Full text] [doi: 10.1016/j.cpr.2017.09.002] [Medline: 28941927]
27. Fernandez KC, Jazaieri H, Gross JJ. Emotion regulation: a transdiagnostic perspective on a new RDoC domain. Cognit Ther Res 2016;40(3):426-440 [FREE Full text] [doi: 10.1007/s10608-016-9772-2] [Medline: 27524846]
28. Trompeter N, Bussey K, Forbes MK, Mitchison D. Emotion dysregulation within the CBT-E model of eating disorders: a narrative review. Cogn Ther Res 2021;45(6):1021-1036 [FREE Full text] [doi: 10.1007/s10608-021-10225-5]
29. Young KS, Sandman CF, Craske MG. Positive and negative emotion regulation in adolescence: links to anxiety and depression. Brain Sci 2019;9(4):76 [FREE Full text] [doi: 10.3390/brainsci9040076] [Medline: 30934877]
30. Brausch AM, Clapham RB, Littlefield AK. Identifying specific emotion regulation deficits that associate with nonsuicidal self-injury and suicide ideation in adolescents. J Youth Adolesc 2022;51(3):556-569 [FREE Full text] [doi: 10.1007/s10964-021-01525-w] [Medline: 34686951]
31. Squires LR, Hollett KB, Hesson J, Harris N. Psychological distress, emotion dysregulation, and coping behaviour: a theoretical perspective of problematic smartphone use. Int J Ment Health Addiction 2020;19(4):1284-1299. [doi: 10.1007/s11469-020-00224-0]

32. Kapatais A, Williams AJ, Townsend E. The mediating role of emotion regulation on self-harm among gender identity and sexual orientation minority (LGBTQ+) individuals. Arch Suicide Res 2023;27(2):165-178. [doi: 10.1080/13811118.2022.2064254] [Medline: 35446245]
33. Dobos B, Piko BF, Mellor D. What makes university students perfectionists? The role of childhood trauma, emotional dysregulation, academic anxiety, and social support. Scand J Psychol 2021;62(3):443-447 [FREE Full text] [doi: 10.1111/sjop.12718] [Medline: 33742444]
34. Mülder LM, Deci N, Werner AM, Reichel JL, Tibubos AN, Heller S, et al. Antecedents and moderation effects of maladaptive coping behaviors among German university students. Front Psychol 2021;12:645087 [FREE Full text] [doi: 10.3389/fpsyg.2021.645087] [Medline: 34025514]
35. Slovak P, Antle A, Theofanopoulou N, Roquet CD, Gross J, Isbister K. Designing for emotion regulation interventions: an agenda for HCI theory and research. ACM Trans Comput-Hum Interact 2023;30(1):1-51. [doi: 10.1145/3569898]
36. Slovak P, Ford BQ, Widen S, Roquet CD, Theofanopoulou N, Gross JJ, et al. An in situ, child-led intervention to promote emotion regulation competence in middle childhood: protocol for an exploratory randomized controlled trial. JMIR Res Protoc 2021;10(11):e28914 [FREE Full text] [doi: 10.2196/28914] [Medline: 34751666]
37. Theofanopoulou N, Slovak P. Exploring technology-mediated parental socialisation of emotion: leveraging an embodied, in-situ intervention for child emotion regulation. 2022 Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; April 29-May 5, 2022; New Orleans, LA, USA p. 1-16 URL: https://dl.acm.org/doi/proceedings/10.1145/3491102 [doi: 10.1145/3491102.3502130]
38. Theofanopoulou N, Isbister K, Edbrooke-Childs J, Slovák P. A smart toy intervention to promote emotion regulation in middle childhood: feasibility study. JMIR Ment Health 2019;6(8):e14029 [FREE Full text] [doi: 10.2196/14029] [Medline: 31381502]
39. Gross JJ. The extended process model of emotion regulation: elaborations, applications, and future directions. Psychol Inq 2015;26(1):130-137. [doi: 10.1080/1047840x.2015.989751]
40. Kelders SM, Kip H, Greeff J. Psychometric evaluation of the Twente Engagement with Ehealth Technologies Scale (TWEETS): evaluation study. J Med Internet Res 2020;22(10):e17757 [FREE Full text] [doi: 10.2196/17757] [Medline: 33021487]
41. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097 [FREE Full text] [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
42. Victor SE, Klonsky ED. Validation of a brief version of the Difficulties in Emotion Regulation Scale (DERS-18) in five samples. J Psychopathol Behav Assess 2016;38(4):582-589. [doi: 10.1007/s10862-016-9547-9]
43. Tamir M, John OP, Srivastava S, Gross JJ. Implicit theories of emotion: affective and social outcomes across a major life transition. J Pers Soc Psychol 2007;92(4):731-744. [doi: 10.1037/0022-3514.92.4.731] [Medline: 17469955]
44. De Castella K, Goldin P, Jazaieri H, Ziv M, Dweck CS, Gross JJ. Beliefs about emotion: links to emotion regulation, well-being, and psychological distress. Basic Appl Soc Psychol 2013;35(6):497-505. [doi: 10.1080/01973533.2013.840632]
45. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa] [Medline: 32100154]
46. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
47. Roquet CD, Theofanopoulou N, Freeman JL, Schleider J, Gross JJ, Davis K, et al. Exploring situated & embodied support for youth's mental health: design opportunities for interactive tangible device. 2022 Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; April 29-May 5, 2022; New Orleans, LA, USA p. 1-16 URL: https://dl.acm.org/doi/proceedings/10.1145/3491102 [doi: 10.1145/3491102.3502135]
48. Schleider JL, Dobias ML, Sung JY, Mullarkey MC. Future directions in single-session youth mental health interventions. J Clin Child Adolesc Psychol 2020;49(2):264-278 [FREE Full text] [doi: 10.1080/15374416.2019.1683852] [Medline: 31799863]
49. Bates D, Kliegl R, Vasishth S, Baayen H. Parsimonious mixed models. ArXiv Preprint posted online on May 26, 2018. [FREE Full text] [doi: 10.5860/choice.189890]
50. McCloud T, Jones R, Lewis G, Bell V, Tsakanikos E. Effectiveness of a mobile app intervention for anxiety and depression symptoms in university students: randomized controlled trial. JMIR Mhealth Uhealth 2020;8(7):e15418 [FREE Full text] [doi: 10.2196/15418] [Medline: 32735221]
51. Rith-Najarian LR, Chorpita BF, Gong-Guy E, Hammons HR, Chavira DA. Feasibility of a web-based program for universal prevention of anxiety and depression in university students: an open trial. J Am Coll Health 2022;70(8):2519-2526. [doi: 10.1080/07448481.2020.1869749] [Medline: 33577431]
52. Bantjes J, Kazdin AE, Cuijpers P, Breet E, Dunn-Coetzee M, Davids C, et al. A web-based group cognitive behavioral therapy intervention for symptoms of anxiety and depression among university students: open-label, pragmatic trial. JMIR Ment Health 2021;8(5):e27400 [FREE Full text] [doi: 10.2196/27400] [Medline: 34042598]
53. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]

54. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
55. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
56. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. J Med Internet Res 2017;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
57. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]
58. Graham AK, Wildes JE, Reddy M, Munson SA, Taylor CB, Mohr DC. User-centered design for technology-enabled services for eating disorders. Int J Eat Disord 2019;52(10):1095-1107 [FREE Full text] [doi: 10.1002/eat.23130] [Medline: 31313370]
59. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
60. March S, Spence SH, Donovan CL, Kenardy JA. Large-scale dissemination of internet-based cognitive behavioral therapy for youth anxiety: feasibility and acceptability study. J Med Internet Res 2018;20(7):e234 [FREE Full text] [doi: 10.2196/jmir.9211] [Medline: 29973338]
61. Palacios JE, Richards D, Palmer R, Coudray C, Hofmann SG, Palmieri PA, et al. Supported internet-delivered cognitive behavioral therapy programs for depression, anxiety, and stress in university students: open, non-randomised trial of acceptability, effectiveness, and satisfaction. JMIR Ment Health 2018;5(4):e11467 [FREE Full text] [doi: 10.2196/11467] [Medline: 30552076]
62. Ellard KK, Fairholme CP, Boisseau CL, Farchione TJ, Barlow DH. Unified protocol for the transdiagnostic treatment of emotional disorders: protocol development and initial outcome data. Cogn Behav Pract 2010;17(1):88-101. [doi: 10.1016/j.cbpra.2009.06.002] [Medline: 33762811]
63. Ehrenreich-May J, Kennedy SM, Sherman JA, Bilek EL, Buzzella BA, Bennett SM, et al. Unified Protocols for Transdiagnostic Treatment of Emotional Disorders in Children and Adolescents: Therapist Guide. Oxford: Oxford University Press; 2018.
64. Mennin DS, Fresco DM. Emotion regulation therapy. In: Handbook of Emotion Regulation. New York: The Guilford Press; 2014:469-490.
65. Renna ME, Quintero JM, Soffer A, Pino M, Ader L, Fresco DM, et al. A pilot study of emotion regulation therapy for generalized anxiety and depression: findings from a diverse sample of young adults. Behav Ther 2018;49(3):403-418 [FREE Full text] [doi: 10.1016/j.beth.2017.09.001] [Medline: 29704969]
66. Docherty M, Thornicroft G. Specialist mental health services in England in 2014: overview of funding, access and levels of care. Int J Ment Health Syst 2015;9:34 [FREE Full text] [doi: 10.1186/s13033-015-0023-9] [Medline: 26322123]
67. Grünzig SD, Baumeister H, Bengel J, Ebert D, Krämer L. Correction to: effectiveness and acceptance of a web-based depression intervention during waiting time for outpatient psychotherapy: study protocol for a randomized controlled trial. Trials 2018;19(1):393 [FREE Full text] [doi: 10.1186/s13063-018-2806-1] [Medline: 30025535]
68. Harvey LJ, Hunt C, White FA. Dialectical behaviour therapy for emotion regulation difficulties: a systematic review. Behav change 2019;36(3):143-164 [FREE Full text] [doi: 10.1017/bec.2019.9]
69. Dickinson ER, Adelson JL, Owen J. Gender balance, representativeness, and statistical power in sexuality research using undergraduate student samples. Arch Sex Behav 2012;41(2):325-327 [FREE Full text] [doi: 10.1007/s10508-011-9887-1] [Medline: 22228196]

Abbreviations

DERS-18: Difficulties in Emotion Regulation Scale-18
ER: emotion regulation
GAD-7: Generalized Anxiety Disorder-7
TWEETS: Twente Engagement With eHealth Technologies Scale

Edited by J Torous; submitted 27.02.23; peer-reviewed by D Joshi, A Kazdin, C Laranjeira, N Hung; comments to author 01.04.23; revised version received 14.07.23; accepted 18.07.23; published 31.10.23.

Please cite as:
Williams AJ, Freed M, Theofanopoulou N, Daudén Roquet C, Klasnja P, Gross J, Schleider J, Slovak P
Feasibility, Perceived Impact, and Acceptability of a Socially Assistive Robot to Support Emotion Regulation With Highly Anxious University Students: Mixed Methods Open Trial
JMIR Ment Health 2023;10:e46826
URL: https://mental.jmir.org/2023/1/e46826
doi:10.2196/46826
PMID:37906230

Original Paper

Effectiveness of an App-Based Short Intervention to Improve Sleep: Randomized Controlled Trial

Bianka Vollert[1*], Dipl-Psych; Luise Müller[1], MSc; Corinna Jacobi[1], Prof Dr; Mickey Trockel[2], MD; Ina Beintner[1*], PhD

[1]Department of Clinical Psychology and Psychotherapy, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany

[2]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, United States

[*]these authors contributed equally

Corresponding Author:
Bianka Vollert, Dipl-Psych
Department of Clinical Psychology and Psychotherapy
Faculty of Psychology
Technische Universität Dresden
Chemnitzer Strasse 46
Dresden, 01187
Germany
Phone: 49 351 463 38576
Email: bianka.vollert@tu-dresden.de

Abstract

Background: A growing body of evidence for digital interventions to improve sleep shows promising effects. The interventions investigated so far have been primarily web-based; however, app-based interventions may reach a wider audience and be more suitable for daily use.

Objective: This study aims to evaluate the intervention effects, adherence, and acceptance of an unguided app-based intervention for individuals who wish to improve their sleep.

Methods: In a randomized controlled trial, we evaluated the effects of an app-based short intervention (Refresh) to improve sleep compared with a waitlist condition. Refresh is an 8-week unguided intervention covering the principles of cognitive behavioral therapy for insomnia (CBT-I) and including a sleep diary. The primary outcome was sleep quality (insomnia symptoms) as self-assessed by the Regensburg Insomnia Scale (RIS). The secondary outcomes were depression (9-item Patient Health Questionnaire [PHQ-9] score) and perceived insomnia-related impairment.

Results: We included 371 participants, of which 245 reported poor sleep at baseline. About 1 in 3 participants who were allocated to the intervention group never accessed the intervention. Active participants completed on average 4 out of 8 chapters. Retention rates were 67.4% (n=250) at postassessment and 57.7% (n=214) at the 6-month follow-up. At postintervention, insomnia symptoms in the intervention group had improved more than those in the waitlist group, with a small effect (d=0.26) in the whole sample and a medium effect (d=0.45) in the subgroup with poor sleep. Effects in the intervention group were maintained at follow-up. Perceived insomnia-related impairment also improved from pre- to postassessment. No significant intervention effect on depression was detected. Working alliance and acceptance were moderate to good.

Conclusions: An app-based, unguided intervention is a feasible and effective option to scale-up CBT-I-based treatment, but intervention uptake and adherence need to be carefully addressed.

Trial Registration: ISRCTN Registry ISRCTN53553517; https://www.isrctn.com/ISRCTN53553517

(JMIR Ment Health 2023;10:e39052) doi:10.2196/39052

KEYWORDS

sleep; insomnia; cognitive behavioral treatment for insomnia; eHealth; mobile app

Introduction

Sleep is of vital importance for our everyday functioning, health, and well-being [1]. However, difficulties falling or staying asleep are common among the general population worldwide [2-4]. Approximately 1 in 3-4 adults experience symptoms of insomnia, while up to 10%-15% cross the threshold to full-syndrome insomnia [5-7]. Poor sleep is associated with disabling daytime consequences including fatigue, reduced performance, irritability, mood disturbances, and impaired health-related quality of life [8-11].

Individuals with insomnia often report symptoms that persist over several months, indicating a chronic course of insomnia. In a longitudinal study, more than 70% of the participants experienced insomnia for at least one year and 1 in 2 reported that symptoms lasted even longer [12]. Insomnia symptoms can be perpetuated by dysfunctional cognitions and maladaptive sleep-related behaviors and habits that contribute to a psycho-physiological hyperarousal that is incompatible with sleep [13,14].

Insomnia has been linked to various health risks including cardiovascular disease, high blood pressure, and diabetes [15-17]. Poor sleep also has effects on the brain and cognition; impairments in performance, concentration, working memory, and response time are associated with chronic sleep disorders [18-20]. Insomnia also increases the risk for other mental disorders, especially depression [21-23].

Because of substantial direct (eg, medication and increased health care utilization) and indirect costs (eg, absenteeism, reduced work productivity, and accidents), insomnia is associated with a high socioeconomic burden [24-26].

Given the high prevalence, associated health risks, and the global burden, it is crucial that insomnia is recognized early and treated adequately [14]. Cognitive behavioral therapy for insomnia (CBT-I) is consistently deemed the first choice of treatment for insomnia [27,28]. It is considered more suitable than a (pure) pharmacological treatment, especially given the limited evidence for long-term effects of the latter [29]. CBT-I consists of several components including psychoeducational, cognitive, and behavioral elements [30].

CBT-I has shown to be effective in reducing temporary as well as persisting insomnia symptoms in several studies and meta-analyses (eg, [31-36]). In addition, studies suggest that CBT-I improves (comorbid) depression and may even prevent the onset of depression [37-40].

However, reach and access to face-to-face, therapist-delivered CBT-I is limited [41-43], due to both structural and attitudinal barriers. Structural barriers include limited treatment availability, inadequate screening and referrals, and long waiting times [41]. Within-person attitudinal barriers include reservations about help-seeking in general or nonpharmacological treatment options in particular [44,45].

Low-threshold and scalable CBT-based self-help interventions may be a means to overcome these barriers and facilitate access to adequate treatment [46], especially if they are provided online [47]. Internet-based interventions have become increasingly popular in the prevention and treatment of mental disorders (eg, [48]), and offer easy accessibility, flexibility, and anonymity. They allow users to integrate an intervention into daily life and in many cases also to automatically monitor progress [49].

There is a growing body of evidence for internet-based CBI-I interventions [50,51]. Meta-analyses revealed large improvements in insomnia severity (g=1.09) and other sleep outcomes (eg, sleep efficiency and total sleep time) that are comparable to those found in face-to-to interventions [52,53]. Besides, quality of life, executive functions, and work-related health and productivity have shown to be improved through internet-based CBT-I [54-57]. Findings indicate that internet-based CBT-I also prevents or reduces symptoms of depression [58-61].

Fully automated internet-based self-help interventions without human guidance may have the greatest potential when it comes to scalability and cost-effectiveness. While some studies suggest that guided interventions are more effective than unguided ones [53,62], the latter may still yield large and sustainable effects compared with a waitlist or active control condition [63-68].

So far, most research on online interventions has focused on web-based interventions [65]. In the past decade, mobile internet use has been growing and interventions that can be accessed on a smartphone promise a higher potential to reach users than those that require a large screen [69]. By contrast, mobile use may be associated with different user requirements and expectations, including the length and duration of treatment modules or the presentation of information, which may affect adherence and treatment effects.

While there are already a large number of consumer-targeting apps addressing and tracking sleep (behavior) available in mobile app stores, most of them lack evidence [70]. We only identified 2 randomized controlled trials evaluating app-based interventions for individuals with insomnia. One compared a fully automated Dutch intervention focusing on sleep restriction and relaxation with a waitlist condition [65]. The other compared a Persian self-help intervention based on a combination of Theory of Planned Behavior, Health Action Process Approach, and CBT-I with an active control condition (patient education) [71]. Although both interventions did not require any human input, the app by Horsch et al [65] included a conversation tool (chatbot) between the app and the participants as well as a number of persuasive strategies. Both randomized controlled trials applied strict inclusion criteria for study participation (ie, insomnia symptoms in accordance with the criteria for a DSM-5 [Diagnostic and Statistical Manual of Mental Disorders, 5th edition] diagnosis of insomnia) and found that the app-based intervention was superior to the control condition in the improvement of insomnia symptoms.

Given the relative lack of evidence for mobile CBT-I interventions, the aim of this study is to evaluate an unguided app-based training for individuals who wish to improve their sleep. The primary aim is to evaluate the effect of the intervention on sleep quality (insomnia symptoms). Secondary, exploratory aims include the intervention effects on insomnia-related impairment and depression symptoms as well

as participant adherence to the intervention, working alliance, and intervention acceptance.

Methods

Study Design

Following an uncontrolled feasibility pilot study with 189 participants (S. Buntzel, unpublished data, 2018), we conducted a randomized controlled trial comparing an intervention group with access to a mobile CBT-I intervention with a waiting list control group (Multimedia Appendix 1). Outcomes were assessed through online questionnaires at baseline (pretreatment), postintervention (8 weeks after randomization), and 6-month follow-up. The primary outcome was the change in sleep quality (insomnia symptoms and Regensburg Insomnia Scale [RIS] total score) at postassessment compared with the baseline assessment.

Participants were randomized in a 1:1 ratio. The randomization was block stratified by gender (female vs male); severity of insomnia symptoms, poor (RIS total score ≥13 [72]) versus good sleep (RIS total score <13); and the consumption of sleep-inducing drugs (yes vs no). The random allocation sequence was implemented on the data collection platform.

Participants received information about their group assignment by email. Participants in the intervention group had immediate access to the intervention. By contrast, participants in the waiting list group received access to Refresh after the 6-month follow-up had been completed if they were still interested in using the intervention. Participants in both groups were invited to the post- and follow-up assessments via emails. They received up to 5 reminders at intervals of 3 days.

Ethics Approval

The study was approved by the Ethics Committee (Institutional Review Board) at Technische Universität Dresden (reference number EK 111032919) and registered prospectively in the ISRCTN database (registration number ISRCTN53553517).

Inclusion and Exclusion Criteria

We included participants over 18 years of age who were fluent in German and had access to the internet during the intervention period. Exclusion criteria were (1) current treatment for depression, (2) a history of psychotic or bipolar disorders, and (3) a suicidal ideation according to the answer in the last item of the 9-item Patient Health Questionnaire (PHQ-9). Poor sleep was not necessary to participate in the study. All adults showing interest in improving their sleep were welcome if none of the exclusion criteria were present.

Recruitment/Procedure

Participants were recruited through flyers, postcards, and posters in several sleep laboratories in Germany as well as in practices of general practitioners, pharmacies, and medical supply stores in Dresden. In addition, recruitment activities took place at Technische Universität Dresden and included a press release, student email newsletters as well as postcards and posters distributed on the campus. A newspaper article was published in a regional newspaper. Social media channels (eg, Facebook groups dealing with sleep issues) were used to address further potential participants. The recruitment material included a link and a QR code leading to the study website. Interested participants received written information about the study and informed consent was obtained online. Study data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) electronic data capture tools hosted at the Centre for Clinical Studies at Technische Universität Dresden. REDCap [73] is a secure, web-based application designed to support data capture for research studies, which provides (1) an intuitive interface for validated data entry; (2) audit trails for tracking data manipulation and export procedures; (3) automated export procedures for seamless data downloads to common statistical packages; and (4) procedures for importing data from external sources.

Intervention

Refresh is an unguided, fully automated app-based intervention adapted and translated from an e-mail-delivered CBT for sleep health program for college students, which was originally developed and evaluated in the United States [74], and subsequently translated and evaluated in a Japanese college student population [75]. The intervention was adapted for the general adult population and designed as an app-based self-help intervention. Refresh was implemented using a commercially available eHealth platform [76]. Some screenshots of Refresh are provided in Multimedia Appendix 2.

The intervention consists of 8 consecutive chapters that can be completed in about 10 minutes each and cover all CBT-I components recommended by the German Sleep Society (DGSM): psychoeducation, sleep hygiene, stimulus control, sleep restriction, cognitive restructuring, and relaxation [30]. Table 1 summarizes the content of the individual chapters.

Table 1. Content of the Refresh intervention.

Chapter	Content
1	• Introduction • Psychoeducation: (1) rapid eye movement/non–rapid eye movement sleep and stages of sleep; (2) consequences of lack of sleep • Introduction to the "30-second sleep diary"
2	• Psychoeducation: (1) sleep-wake cycle and the circadian clock; (2) the 2-process model of sleep [77] • Establishing a regular sleep schedule
3	• Introduction to sleep restriction (only for people with poor sleep) • Relaxation: (1) simple breathing relaxation exercise
4	• Mindfulness: (1) mindful breathing; (2) body scan
5	• Stimulus control • Finding compromise: (1) neighbors, spouses, babies • Sleep restriction • Review of the sleep schedule
6	• Rumination and worries: (1) strategies to reduce nighttime rumination; (2) take a mindfulness-based approach to worry
7	• Cognitive reappraisal • Sleep restriction: (1) review of the sleep schedule
8	• Sleep restriction: (1) review of the sleep schedule • Wrap up

The chapters are multimedia based with short text passages, audio and video content, vignettes, and questions (multiple choice or free text) to foster an active engagement with the intervention content as well as to tailor subsequent content based on the participants' preference (eg, whether they preferred information to be presented as text or video). The number of pages per chapter and the amount of plain text per page were kept to a minimum as factors such as extensive text content and text content complexity are likely to increase the risk of nonadherence [78]. The intervention is supplemented by a "30-second sleep diary" to be filled in every morning to monitor insomnia symptoms. For a duration of 8 weeks, an automated reminder to fill in the sleep diary was sent every morning at 7 AM. There were no reminders for progressing through the intervention. At the beginning, participants received a short automated feedback about their sleep based on the baseline RIS score. Participants reporting poor sleep were especially encouraged to use parts of the intervention that were marked as for "people with poor sleep" (eg, sleep restriction).

The intervention is self-paced, but the recommended duration was 8 weeks with 1 chapter per week.

Measures/Outcomes

Regensburg Insomnia Scale

Sleep quality (insomnia symptoms) was assessed using the German RIS [72]. The scale covers psychophysiological (cognitive, emotional, and behavioral) aspects of insomnia during the previous 4 weeks. The instrument consists of 10 items to be answered on a 5-point self-rating scale ranging from 0 to 4. A total score of ≥13 indicates poor sleep [72]. The RIS has shown to be a well-accepted, valid instrument that has discriminative power and is sensitive to detect improvements in insomnia parameters after CBT-I [72].

An additional item asked the participants about their perceived insomnia-related impairment in the past 7 days on a scale ranging from 0 to 100.

The 9-Item Patient Health Questionnaire

Depressive symptoms and depression severity were measured using the German version of the PHQ-9 [79]. This widely used brief self-report instrument consists of 9 items covering the DSM-5 diagnostic criteria of depression that are scored on a 4-point Likert scale from 0 (not at all) to 3 (nearly every day). The PHQ-9 has been shown to have good psychometric properties [80] and to be sensitive to changes in depression symptoms over time [81].

Working Alliance Inventory-Short Revised

Acceptance of the intervention was assessed using the subscales "task" and "goal" of the Working Alliance Inventory-Short Revised (WAI-SR) [82,83], adapted for online interventions. The inventory is based on Bordin's Alliance Theory [84]. The "task" subscale (4 items) measures the agreement on the tasks of the intervention, whereas the "goal" subscale (4 items) measures the agreement on the goals of the intervention. Because of the unguided nature of the study, the "bond" subscale (measuring the quality of an affective bond between patient [participant] and therapist [coach]) was removed. Items are answered on a scale from 1 (seldom) to 5 (always), but participants could also choose the additional answer category "I don't know." In addition to the WAI-SR, each chapter of the intervention could be rated on a 5-point scale.

Adherence Markers

Adherence to the intervention was defined as the number of chapters completed and the proportional progress through the intervention (in percentage). In addition, the number of entries in the sleep diary was used to describe the usage of the intervention.

Sample Size Calculation

Based on results from the pilot study (S. Buntzel, unpublished data, 2018), we assumed a pre-post between-group effect size (Cohen *d*) of 0.30 and a dropout rate of 60% at postintervention. To detect the anticipated effect with an 80% probability at a significance level of 5%, a sample size of 586 participants was required.

Statistical Analyses

All statistical analyses were performed with SPSS Statistics version 27 (IBM Corp). Statistical significance was set at α=.05. Differences in baseline sociodemographic and clinical scores between the intervention group and the control group as well as between participants with poor and good sleep were analyzed using Pearson chi-square tests for dichotomous variables and independent sample *t* tests (2-sided, unpaired) for metric variables. Differences between completers and noncompleters were analyzed in the same way.

Adherence and acceptance data were analyzed descriptively and compared between participants with poor and good sleep by applying independent sample *t* tests. To identify potential predictors of adherence, Pearson correlations were calculated between adherence markers (number of chapters completed and number of sleep diary entries) and baseline variables.

The primary (RIS) and secondary (perceived insomnia-related impairment and PHQ-9 score) outcomes were analyzed using linear repeated-measures mixed-effect models with restricted maximum likelihood estimation and an unstructured covariance matrix. This method follows the intention-to-treat (ITT) approach and is recommended for randomized controlled trials with missing data [85]. Group, time, and interaction of group × time were entered as fixed variables with group as a between-group variable and time as a within-group variable. The outcomes RIS, perceived insomnia-related impairment, and PHQ-9 scores were entered as dependent variables in separate analyses. Within-group effect sizes (Cohen *d*) were calculated by dividing the estimated mean change from baseline to postassessment (or follow-up) by the pooled SD [86]. Between-group effect sizes (Cohen *d*) were computed based on the difference of the estimated mean change from baseline to postassessment (or follow-up) in the intervention group compared with the control group divided by the pooled SD at baseline as recommended by Morris [87].

The analyses for the whole sample were repeated including only participants with poor sleep at baseline as indicated by an RIS score of ≥13 at baseline. Improvements in insomnia symptoms were tested for clinical relevance by comparing the frequency of participants with poor sleep (RIS ≥13) at baseline in both study arms who had improved at postintervention and at follow-up using Pearson chi-square tests for assessment completers. In addition, the odds of suffering from poor sleep (RIS ≥13) at postintervention were compared between the 2 groups using a more conservative ITT approach including all randomized participants with poor sleep. We used a logistic regression model and adjusted for sex, age, and RIS score at baseline. Data on participants who did not complete postassessment questionnaires were imputed by assuming poor sleep at postintervention.

Results

Participants

The recruitment period and baseline assessments ran from April 2019 to May 2020. Follow-up data collection was completed in September 2020. Given that the pre-post dropout rate was much smaller than expected, we included fewer participants than originally planned. A total of 393 completed the baseline assessment, of which 371 were randomized (Figure 1). About two-thirds of participants (n=245, 66.0%) reported poor sleep according to an RIS total score of ≥13.

A total of 250 participants (67.4%) provided the primary outcome (RIS total score) at postassessment and 214 (57.7%) completed the RIS at the follow-up assessment. Dropout was higher in the intervention group than in the control group at both assessment points (χ^2_{1post}=24.477; *P*<.001 and χ^2_{1FU6}=28.247; *P*<.001). Participants with poor initial sleep did not differ from those with good sleep in dropout rates. There also were no differences in any of the baseline measures between study dropouts and completers at postassessment. At follow-up, women (χ^2_1=5.813; *P*=.02) and participants with a higher level of education (χ^2_1=7.832; *P*=.005) were more likely to complete the assessment. In the subgroup of participants with poor sleep, education also predicted completion at postassessment (χ^2_1=5.163; *P*=.02).

The majority of participants in the full sample were women (223/371, 62.8%), in a relationship (271/371, 73.0%), and reported a higher level of education (324/371, 87.3%; Table 2). About 1 in 4 had children and about 1 in 3 were university students. The mean age of participants was 37.3 (SD 14.24) years. Sleep was poor on average (indicated by an RIS score ≥13) and the average level of depression severity was mild.

The subgroups of participants with poor or good sleep differed significantly in most of the sociodemographic variables and in the clinical scores (Table 2). For example, participants with poor sleep were significantly older (*P*<.001), were more likely to have children (*P*<.001), and more likely to be self-employed (*P*=.04). They also reported more severe depressive symptoms (*P*<.001) than those with good sleep and were more likely to receive treatment (*P*<.001) or use medication for their sleep problems (*P*<.001).

There were no significant differences between the intervention group and the control group in sociodemographic characteristics, clinical scores (see Multimedia Appendix 3), or the proportion of participants with poor sleep (117/186, 62.9% vs 128/185, 69.2%, respectively; χ^2_1=1.634; *P*=.23.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow of participants. FU6: 6-month follow-up; PHQ-9: 9-item Patient Health Questionnaire.

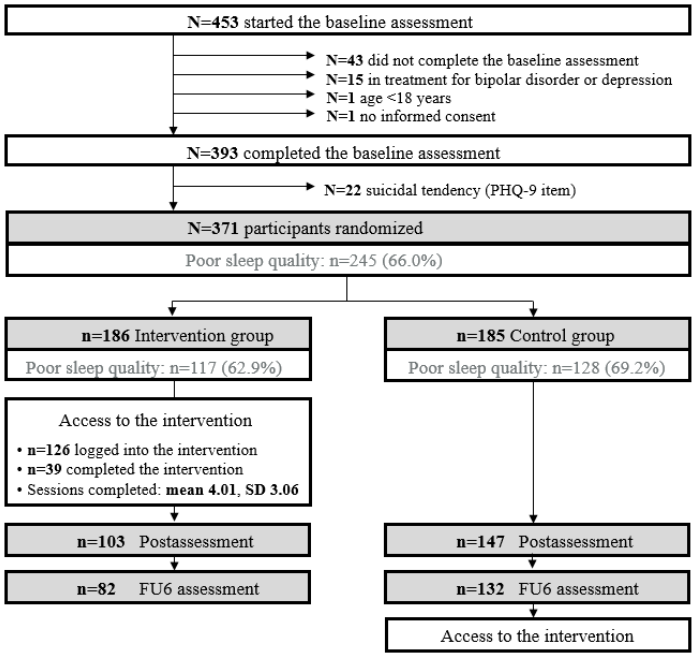


Table 2. Baseline characteristics of participants.

Characteristics	Full sample (N=371)	Poor sleep (n=245)	Good sleep (n=126)	*P* value
Female gender, n (%)	233 (62.8)	156 (63.7)	77 (61.1)	.65
In a relationship, n (%)	271 (73.0)	184 (75.1)	87 (69.0)	.23
Children (age <18 years), n (%)	89 (24.0)	69 (28.2)	20 (15.9)	.006
Higher education[a], n (%)	324 (87.3)	205 (83.7)	119 (94.4)	.003
In education, n (%)	122 (32.9)	60 (24.5)	62 (49.2)	<.001
Self-employed, n (%)	26 (7.0)	22 (9.0)	4 (3.2)	.04
Shift work, n (%)	34 (9.2)	25 (10.2)	9 (7.1)	.33
In treatment, n (%)	41 (11.1)	39 (15.9)	2 (1.6)	<.001
Medication, n (%)	58 (15.6)	58 (23.7)	0 (0.0)	<.001
Sick leave, n (%)	14 (3.8)	13 (5.3)	1 (0.8)	.04
Age, mean (SD)	37.30 (14.24)	41.16 (14.61)	29.82 (9.92)	<.001
RIS[b] at baseline, mean (SD)	15.33 (6.24)	18.72 (4.63)	8.75 (2.67)	<.001
Perceived insomnia-related impairment, mean (SD)	47.23 (27.37)	58.97 (21.64)	24.40 (22.53)	<.001
PHQ-9[c] score at baseline, mean (SD)	7.44 (3.56)	8.64 (3.01)	5.11 (3.02)	<.001

[a]At least level 4 according to the European Qualifications Framework.

[b]RIS: Regensburg Insomnia Scale.

[c]PHQ-9: 9-item Patient Health Questionnaire (Depression).

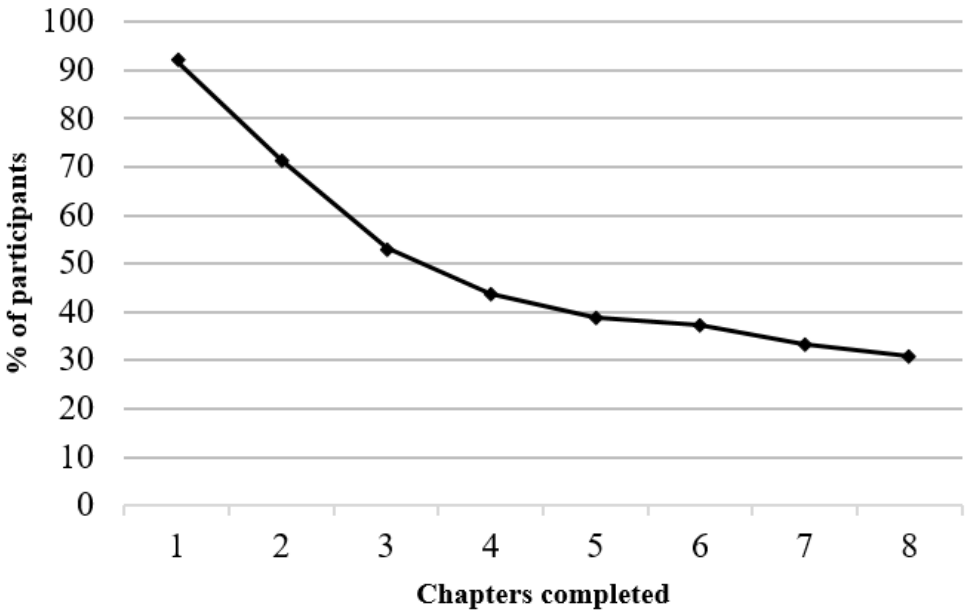
Adherence to the Intervention

A total of 60/186 (32.3%) participants never logged into the platform or opened the first chapter. Participants who logged into the intervention at least once (n=126) completed on average 4 chapters (SD 3.06; range 0-8) and opened 42/83 (51%) intervention pages (SD 38.14). A total of 39/126 (31.0%) participants completed all chapters of the intervention and 55/126 (43.7%) completed at least half of the intervention; 116/126 (92.1%) completed at least chapter 1 (Figure 2). A total of 106/126 participants (84.1%) used the sleep diary at least once and made entries for 25.23 days (SD 22.3; range 0-83).

Male participants were more likely to not log into the intervention at all (30/65, 46.2% vs 30/121, 24.8%; χ^2_1=8.829; *P*=.003). Participants with children made fewer entries in the

sleep diary than those without (*P*=.02). None of the other sociodemographic variables were predictive for the number of chapters completed or the number of sleep diary entries, and also both RIS total score and PHQ-9 total score at baseline did not predict adherence to the intervention.

Figure 2. Chapter completion rates in participants who accessed the intervention.



Working Alliance and Acceptance

As many as 87 participants from the intervention group provided WAI-SR data that allowed calculating the score of the Task and Goal subscales. The participants agreed between "sometimes" and "fairly often" with the tasks of the intervention and between "fairly often" and "often" with the goals of the intervention. Acceptance did not differ between participants with poor and good sleep at baseline (Table 3).

Participants who answered the session rating of the intervention chapters rated chapters 1-6 as "good" on average, whereas chapters 7 and 8 were rated as "moderate." There was no significant difference in session ratings between participants with poor and good sleep, except for session 2 that was rated better and nearly "excellent" by participants with good sleep (*P*<.001).

Table 3. Perceived working alliance (WAI-SR[a] subscales).

Subscale	IG[b] (all; n=87), mean (SD)	IG poor sleep (n=58), mean (SD)	IG good sleep (n=29), mean (SD)	*P* value
WAI-SR Task	2.78 (0.99)	2.82 (0.96)	2.70 (1.05)	.60
WAI-SR Goal	3.17 (0.98)	3.18 (0.95)	3.16 (1.07)	.95

[a]WAI-SR: Working Alliance Inventory-Short Revised.

[b]IG: intervention group.

Intervention Effects

Results of the mixed model analyses are presented in Table 4. For the full sample, a significant group × time interaction was found for the RIS at postassessment ($F_{1,269.31}$=11.93; *P*=.001) with a small between-group effect (*d*=0.26). Participants in the intervention group showed a stronger reduction (mean difference = –2.66; d_{within}=0.42) in the RIS scores than those in the control group (mean difference = –1.04; d_{within}=.17). At follow-up, the group × time interaction failed to reach significance ($F_{1,227.63}$=3.54; *P*=.06). Within-group improvements in the perceived insomnia-related impairment due to sleep problems as well as in the PHQ-9 did not differ significantly between the study arms (Table 3).

In the subgroup analyses including only participants with poor sleep at baseline, a significant group × time interaction was found for the RIS at postassessment ($F_{1,180.66}$=12.59; *P*<.001; *d*=0.47) and for the perceived insomnia-related impairment at postassessment ($F_{1,184.45}$=6.69; *P*=.01; *d*=0.45) as well as for the RIS at the follow-up assessment ($F_{1,147.56}$=4.60; *P*=.03; *d*=0.33). Although the intervention group showed somewhat larger reductions in the PHQ-9 than the control group, the difference was not significant (P_{post}=.13; P_{FU6}=.15; also see Table 4).

Table 4. Results of the intention-to-treat analyses including all participants (N=371) and participants with poor baseline sleep only (n=245).

Results	Baseline, mean (SD)	Post (estimated), mean (SE)	Effect pre-post (within)		Effect pre-post (between)		Follow-up (estimated), mean (SE)	Effect pre-FU6[a] (within)		Effect pre-FU6 (between)	
			Cohen *d*	95% CI	Cohen *d*	*P* value		Cohen *d*	95% CI	Cohen *d*	*P* value
All participants (N=371)											
RIS[b] total											
IG[c]	15.11 (6.34)	12.45 (0.47)	0.42	0.27 to 0.57	0.26 (0.05 to 0.46)[d]	.001	12.23 (0.52)	0.45	0.30 to 0.60	0.16 (–0.05 to 0.36)	.06
CG[e]	15.55 (6.14)	14.51 (0.44)	0.17	0.02 to 0.31	N/A[f]		13.66 (0.46)	0.31	0.16 to 0.45	N/A	
Perceived insomnia-related impairment[g]											
IG	45.66 (27.52)	34.76 (2.54)	0.39	0.25 to 0.54	0.17 (–0.04 to 0.37)	.16	35.85 (2.78)	0.36	0.21 to 0.50	0.05 (–0.16 to 0.25)	.67
CG	48.81 (27.20)	42.45 (2.22)	0.23	0.09 to 0.37	N/A		40.31 (2.32)	0.31	0.16 to 0.46	N/A	
PHQ-9[h]											
IG	7.41 (3.41)	6.13 (0.32)	0.37	0.23 to 0.60	0.13 (–0.08 to 0.33)	.25	6.24 (0.45)	0.34	0.19 to 0.49	0.17 (–0.03 to 0.37)	.25
CG	7.47 (3.72)	6.65 (0.30)	0.22	0.07 to 0.37	N/A		6.91 (0.36)	0.15	0.01 to 0.29	N/A	
Subgroup with poor sleep at baseline (N=245)											
RIS total											
IG	18.73 (5.02)	15.00 (0.51)	0.74	0.53 to 0.94	0.47 (0.22 to 0.73)[d]	<.001	14.81 (0.62)	0.78	0.57 to 0.98	0.33 (0.08 to 0.59)[d]	.03
CG	18.71 (4.27)	17.18 (0.44)	0.36	0.18 to 0.53	N/A		16.34 (0.51)	0.55	0.37 to 0.74	N/A	
Perceived insomnia-related impairment											
IG	57.97 (22.17)	41.37 (2.89)	0.74	0.54 to 0.95	0.45 (0.20 to 0.71)[d]	.01	43.27 (3.36)	0.66	0.46 to 0.86	0.26 (0.01 to 0.51)	.18
CG	59.88 (21.19)	53.13 (2.43)	0.32	0.14 to 0.49	N/A		50.88 (2.64)	0.42	0.24 to 0.60	N/A	
PHQ-9											
IG	8.61 (3.07)	6.94 (0.43)	0.54	0.35 to 0.73	0.25 (0.00 to 0.51)	.13	7.27 (0.60)	0.43	0.24 to 0.62	0.33 (0.08 to 0.59)	.15
CG	8.66 (3.34)	7.81 (0.36)	0.25	0.08 to 0.43	N/A		8.40 (0.45)	0.08	–0.10 to 0.25	N/A	

[a]FU6: 6-month follow-up

[b]RIS: Regensburg Insomnia Scale.

[c]IG: intervention group.

[d]Significant group × time interaction.

[e]CG: control group.

[f]N/A: not applicable.

[g]Refers to subjective impairment in the last 7 days (scale ranging from 0 to 100).

[h]PHQ-9: 9-item Patient Health Questionnaire-9 (Depression).

Evaluation of Clinical Relevance

To assess clinically relevant improvements, we calculated how many participants with poor sleep at baseline had reduced their RIS score below 13 points at postintervention and follow-up assessments. At the postintervention assessment, this was true for 25/69 (36%) assessment completers in the intervention group and 16/102 (15.7%) assessment completers in the control group.

This difference was statistically significant (χ^2_1=9.531; *P*=.002). At follow-up, there was no significant difference between the groups (21/53 IG completers vs 25/90 CG completers; χ^2_1=2.145; *P*=.14). In the ITT analysis of clinically relevant improvements, the odds of suffering from poor sleep at postintervention were also significantly larger in the control group than in the intervention group (odds ratio 0.462, 95% CI 0.218-0.976; *P*=.02).

To assess clinically relevant deterioration, we assessed how many participants with good sleep at baseline had increased their RIS score to 13 points or more at postintervention and follow-up assessments. At postassessment, this was true for 1/34 (3%) assessment completers in the intervention group and 10/45 (22%) assessment completers in the control group. This difference was significant (χ^2_1=6.007; *P*=.01). At follow-up, there was no significant difference between the groups (2/29 IG completers vs 2/42 CG completers; χ^2_1=0.147; *P*=.71).

Discussion

Summary of the Trial Objectives

The aim of this study was to evaluate an unguided app-based training to improve sleep quality (insomnia symptoms) in individuals who wish to improve their sleep compared with a waitlist control group. Specifically, we examined participant adherence to the intervention; working alliance and intervention acceptance; and intervention effects on sleep quality (insomnia symptoms), insomnia-related impairment, and depression symptoms.

Adherence, Working Alliance, and Acceptance

Nonusage attrition was comparable to other digital interventions [88], with a larger loss of participants during the early intervention phase and about 1 in 3 participants completing the intervention. About 1 in 3 participants who were allocated to the intervention group never accessed the intervention. Failure to commence the intervention was 2 times as likely in men, with 1 in 2 never accessing the intervention. The proportion of participants who never started the intervention was markedly lower in a previous study investigating an unguided mobile CBT-I–based intervention [65]. However, we applied broader inclusion criteria compared with [65], that is, we also included individuals reporting slight insomnia symptoms or symptoms below the cutoff (RIS <13). Because our sample was healthier, it can be assumed that the insomnia-related distress of these participants was lower and may have resulted in a lower motivation to start the intervention after completing the baseline assessment. Furthermore, we did not exclude participants with depression. Evidence suggests that adherence to an internet-based CBT-I program is reduced if participants experience other psychiatric or medical problems next to insomnia [89].

Because of the heterogeneity in adherence markers reported (if any were reported), adherence can be compared in detail only with 2 other recent clinical trials [66,90] that investigated unguided digital CBT-I interventions. Participants, on average, completed 4 out of 8 chapters in our study, which is comparable with adherence in another trial, in which participants completed 3.4 out of 6 modules [90]. However, in Lorenz et al [66], adherence was higher, with participants completing 5.6 out of 6 treatment sessions. In our study participants completed the sleep diary, on average, on 25 out of 56 days (45%), whereas they did so for 86% of days in one of the other trials [66].

Participants in the intervention group were able to form a moderate working alliance with the unguided intervention, with a slightly larger concordance regarding goals than regarding tasks. Working alliance in psychotherapy tends to be stronger, but a similar pattern in concordance regarding goals versus tasks is typically also seen in patients undergoing psychotherapy [91], which could be explained by a lesser degree of ambivalence toward treatment goals compared with the actual behavior changes (tasks) that are necessary to achieve them. Working alliance has not been investigated in recent clinical trials investigating unguided digital CBT-I interventions.

The first 6 chapters of the intervention, which provided psychoeducation and various strategies to improve sleep-related habits and behaviors, were on average rated “good,” whereas the final 2 chapters, which mainly focused on reviewing and reinforcing the sleep restriction method and provided fewer new strategies, were rated “moderate.”

Intervention Effects

Sleep Quality (Insomnia Symptoms)

Participants in both groups were able to improve insomnia symptoms during the 6-month assessment period, but participants in the intervention group achieved this improvement faster. In the full sample, the intervention was associated with a statistically significant improvement of insomnia symptoms at postintervention that was maintained at follow-up, although the latter just failed to reach statistical significance. Effect sizes were, however, small. In the subsample of participants with poor sleep quality at baseline, the intervention yielded a medium effect on insomnia symptoms at postintervention and a small effect at follow-up, both being statistically significant. The subgroup effect on insomnia symptoms at postintervention is comparable to that achieved in a previous study evaluating an app-based CBT-I program that did not cover all CBT-I components, with stricter inclusion criteria regarding the presence of insomnia symptoms [65]. In another study that assessed the effects of an app-based CBT-I program, insufficient data have been reported to determine ITT effect sizes, and therefore, a direct comparison is not feasible [71]. Overall, unguided web-based interventions yielded medium to large effects in previous studies [66,68,90,92,93]. Assessment completers with poor sleep at baseline were more likely to achieve clinically relevant improvement at postintervention in the intervention group than in the waitlist group. Moreover, considering all included participants, the odds of suffering from poor sleep at postintervention were higher in the waitlist group than in the intervention group.

In the intervention group, insomnia symptoms improved during the intervention period and then remained unchanged between the postintervention and follow-up periods, whereas in the waitlist group a slight improvement was noted between each of

the 3 assessment points. Similar to the intervention group, the improvement in the waitlist group was more pronounced in individuals with poor sleep at baseline. Gradual improvement in insomnia symptoms and sleep quality or efficiency in the untreated control group was also present in some of the other studies on digital CBT-I [65,68,90,92]. These findings indicate that while spontaneous improvements may play a role in insomnia, interventions have the potential of bringing on improvements much faster, thus reducing time with impaired quality of life by promptly supporting individuals at the nadir of their sleep problems.

Our findings also indicate that the intervention may have a protective effect on sleep in individuals with good sleep at baseline. At postintervention, sleep had deteriorated from good to poor only for 1 individual in the intervention group, but over 1 in 5 in the waitlist group. However, that protective effect was not maintained at follow-up.

In comparison with other digital CBT-I studies, we did we did not exclude individuals with insomnia symptoms below the threshold as these individuals still may experience psychological distress or the subjective need to improve their sleep. For these individuals the intervention may have the potential to reduce subthreshold symptoms or prevent symptom progression and the onset of a full syndrome insomnia.

Insomnia-Related Impairment

We observed an intervention effect on insomnia-related impairment only in individuals with poor sleep at baseline, which was statistically significant at postintervention, but just failed statistical significance at follow-up. The pattern of improvement in the 2 groups was parallel to that for insomnia symptoms. In the intervention group, the average impairment was reduced during the intervention period and then remained largely unchanged between the postintervention and follow-up periods, whereas in the waitlist group we observed a slight reduction between each of the 3 assessment points. Again, spontaneous improvements in sleep may play a role.

Depression

Contrary to other studies on digital CBI-I [65,66,71,90,93], we observed no significant intervention effects on depression. On average, participants in the full sample as well as those in the subsample with poor sleep at baseline on average only reported mild symptoms of depression, so there was limited room for improvement.

Strengths and Limitations

Compared with previous studies on digital CBT-I–based interventions, we were able to recruit a sample of participants that was large enough to detect even the small intervention effects that are to be expected in an unguided intervention. The study had a waitlist control condition and a longer follow-up period than 1 of the 2 previous studies investigating a mobile intervention [65]. Other than free access to the intervention, study participants received no incentives for taking part in the study or engaging with the intervention, thus avoiding potential incentive-caused bias that may impact intervention adherence. Thus, we can assume that people who used the intervention intensively had an intrinsic motivation to do so. Despite the lack of incentives, dropout at postintervention (121/371, 32.6%) was smaller than expected from our pilot study and within the range of dropout rates (0%-44%) reported in self-help CBT-I interventions [94]. However, not completing trial assessments (ie, dropout attrition [88]) does not necessarily mean that participants do not use an intervention (ie, nonusage attrition). Some people may be interested in using the intervention, but may not wish to complete assessments. Adherence to the intervention was comparable to other trials on digital interventions, and 1 in 3 participants completed the treatment. However, a substantial number of participants never started the intervention. Prerequisites and time requirements to participate in our study were very low. Although many clinical studies require participants to undergo a longer clinical interview during the screening process, participants in our study only had to fill in a relatively short anonymous online survey. This may have resulted in a larger proportion of participants who were ambivalent about the intervention. Although people could participate in the study regardless of whether they had insomnia symptoms, the proportion of individuals with poor sleep at baseline was high, indicating that the intervention was mostly reaching the users it was designed for.

Limitations of our study were the lack of more specific measures for insomnia-related impairment and socioeconomic consequences, including absenteeism or loss of productivity, as well as an objective measure of sleep quality. However, subjective measures of sleep quality and insomnia symptoms are widely used in CBT-I research and the added value of objective measures is a matter of debate [95]. The participants in this trial might have also suffered from sleep disorders other than insomnia (eg, obstructive sleep apnea, circadian rhythm sleep-wake disorders), in which an intervention covering the principles of CBT-I is likely to have limited or no effect. Future studies may thus benefit from including insomnia diagnosis confirmed by a clinician. Adherence to the intervention was approximated by examining usage data. We did not collect information on whether and how participants in the intervention group implemented the suggested behavior changes. Besides, the study design was not suitable to determine the extent to which the components of the interventions contributed to the effects. Secondary analyses are therefore needed to investigate possible moderators or mediators, which are not the subject of this publication. In accordance with other trials evaluating fully automated internet-based CBT-I, the majority of our sample were highly educated participants who are not representative of the general population. However, even if low education is associated with insomnia-related symptoms [96,97], sleep problems are also frequent in people with a higher level of education. Nevertheless, future research should include a more heterogeneous sample with regard to the socioeconomic level and increase efforts to also reach a lower educated population that might be especially underserved.

Conclusions

In our study, the unguided app-based intervention was associated with a short-term improvement of insomnia symptoms that remained stable over 6 months. Participants in the waitlist condition also reported an improvement in their insomnia

symptoms, but this improvement took longer. Effects were more pronounced in individuals with low sleep at baseline. We detected no effect on depression symptoms in our sample, which may largely be due to the low average symptom load at baseline. Although the inclusion criteria for our study were broad, our recruitment strategy mostly reached the individuals the intervention was designed for.

Overall, an anonymous unguided intervention is a feasible option to deliver CBT-I–based treatment to a large number of people, either through self-referral or by primary care providers. Even if the individual effects are smaller than in guided self-help or face-to-face CBT-I, the public health impact could still be profound given the scalability of unguided approaches. Given that very few people have access to face-to-face CBT-I, making it available digitally can increase its availability [98] and public health impact [99]. As in face-to-face CBT-I, some people will not respond to the intervention and precautions must be taken to prevent negative expectations regarding future psychological interventions in those individuals. Symptoms can be continuously measured within the intervention and if participants do not respond, a more intensive treatment can be recommended. Thus, a digital self-help program can serve as a first step of treatment within a stepped care approach [100]. Moderator analyses can contribute to identifying those who will likely benefit from unguided app-based CBT-I and those who need more guidance or face-to-face treatment.

It is noteworthy, however, that 1 in 4 women and 1 in 2 men who were assigned to the intervention group did not start the intervention. In clinical practice, the motivation and ability to adhere to a self-guided intervention should thus be carefully assessed.

In participants who did start the intervention, adherence was comparable to other digital interventions. In future adaptations of the interventions, adherence may be improved by involving a multidisciplinary team of psychologists, user experience specialists, and user experience designers in the design of the user interface, which unfortunately was not possible in our study due to budget constraints. In this context, adherence to the intervention may benefit from persuasive elements such as reminders, automated positive reinforcement, graphical reports, or tailored recommendations. However, such elements need to be implemented in a way that allows users some level of control (eg, options to determine the frequency of reminders [101]). However, adding several persuasive technology features does not necessarily result in better outcomes [102]. Adding some form of human interaction to Refresh could also improve adherence, but this in turn would result in a reduced scalability [103]. Automated feedback that is given when users show low engagement with the intervention may be an alternative [101]. Increased adherence by implementing design features to encourage engagement with the intervention might contribute to more pronounced improvements in insomnia symptoms.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1). CONSORT: Consolidated Standards of Reporting Trials.
[PDF File (Adobe PDF File), 1176 KB - mental_v10i1e39052_app1.pdf]

Multimedia Appendix 2
Screenshots of the Refresh intervention.
[PDF File (Adobe PDF File), 537 KB - mental_v10i1e39052_app2.pdf]

Multimedia Appendix 3
Baseline characteristics of participants in the intervention group and the control group.
[PDF File (Adobe PDF File), 21 KB - mental_v10i1e39052_app3.pdf]

References

1. Weeß H, Landwehr R. Phänomenologie, Funktion und Physiologie des Schlafes. PiD - Psychotherapie im Dialog 2009 Jun 22;10(02):101-106. [doi: 10.1055/s-0029-1223312]
2. Ohayon MM. Epidemiological Overview of sleep Disorders in the General Population. Sleep Med Res 2011 Apr 30;2(1):1-9. [doi: 10.17241/smr.2011.2.1.1]
3. van de Straat V, Bracke P. How well does Europe sleep? A cross-national study of sleep problems in European older adults. Int J Public Health 2015 Sep;60(6):643-650. [doi: 10.1007/s00038-015-0682-y] [Medline: 25975374]
4. Schlack R, Hapke U, Maske U, Busch M, Cohrs S. [Frequency and distribution of sleep problems and insomnia in the adult population in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2013 May 25;56(5-6):740-748. [doi: 10.1007/s00103-013-1689-2] [Medline: 23703493]
5. Baglioni C, Altena E, Bjorvatn B, Blom K, Bothelius K, Devoto A, et al. The European Academy for Cognitive Behavioural Therapy for Insomnia: An initiative of the European Insomnia Network to promote implementation and dissemination of treatment. J Sleep Res 2020 Apr;29(2):e12967 [FREE Full text] [doi: 10.1111/jsr.12967] [Medline: 31856367]

6. Dopheide JA. Insomnia overview: epidemiology, pathophysiology, diagnosis and monitoring, and nonpharmacologic therapy. Am J Manag Care 2020 Mar;26(4 Suppl):S76-S84 [FREE Full text] [doi: 10.37765/ajmc.2020.42769] [Medline: 32282177]
7. Morin CM, Jarrin DC, Ivers H, Mérette C, LeBlanc M, Savard J. Incidence, Persistence, and Remission Rates of Insomnia Over 5 Years. JAMA Netw Open 2020 Nov 02;3(11):e2018782 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.18782] [Medline: 33156345]
8. Buysse DJ, Thompson W, Scott J, Franzen PL, Germain A, Hall M, et al. Daytime symptoms in primary insomnia: a prospective analysis using ecological momentary assessment. Sleep Med 2007 Apr;8(3):198-208 [FREE Full text] [doi: 10.1016/j.sleep.2006.10.006] [Medline: 17368098]
9. Kyle SD, Morgan K, Espie CA. Insomnia and health-related quality of life. Sleep Med Rev 2010 Feb;14(1):69-82. [doi: 10.1016/j.smrv.2009.07.004] [Medline: 19962922]
10. Roth T, Ancoli-Israel S. Daytime consequences and correlates of insomnia in the United States: results of the 1991 National Sleep Foundation Survey. II. Sleep 1999 May 01;22 Suppl 2:S354-S358. [Medline: 10394607]
11. Espie CA, Kyle SD, Hames P, Cyhlarova E, Benzeval M. The daytime impact of DSM-5 insomnia disorder: comparative analysis of insomnia subtypes from the Great British Sleep Survey. J Clin Psychiatry 2012 Dec;73(12):e1478-e1484. [doi: 10.4088/JCP.12m07954] [Medline: 23290331]
12. Morin CM, Bélanger L, LeBlanc M, Ivers H, Savard J, Espie CA, et al. The natural history of insomnia: a population-based 3-year longitudinal study. Arch Intern Med 2009 Mar 09;169(5):447-453. [doi: 10.1001/archinternmed.2008.610] [Medline: 19273774]
13. Jansson M, Linton S. Psychological mechanisms in the maintenance of insomnia: arousal, distress, and sleep-related beliefs. Behav Res Ther 2007 Mar;45(3):511-521 [FREE Full text] [doi: 10.1016/j.brat.2006.04.003] [Medline: 16759637]
14. Riemann D, Baglioni C, Feige B, Spiegelhalder K. [Insomnia--state of the science]. Nervenarzt 2014 Jan 18;85(1):43-49. [doi: 10.1007/s00115-013-3892-7] [Medline: 24346426]
15. Anothaisintawee T, Reutrakul S, Van Cauter E, Thakkinstian A. Sleep disturbances compared to traditional risk factors for diabetes development: Systematic review and meta-analysis. Sleep Med Rev 2016 Dec;30:11-24. [doi: 10.1016/j.smrv.2015.10.002] [Medline: 26687279]
16. Meng L, Zheng Y, Hui R. The relationship of sleep duration and insomnia to risk of hypertension incidence: a meta-analysis of prospective cohort studies. Hypertens Res 2013 Nov;36(11):985-995 [FREE Full text] [doi: 10.1038/hr.2013.70] [Medline: 24005775]
17. Sofi F, Cesari F, Casini A, Macchi C, Abbate R, Gensini GF. Insomnia and risk of cardiovascular disease: a meta-analysis. Eur J Prev Cardiol 2014 Jan;21(1):57-64. [doi: 10.1177/2047487312460020] [Medline: 22942213]
18. Brownlow JA, Miller KE, Gehrman PR. Insomnia and Cognitive Performance. Sleep Med Clin 2020 Mar;15(1):71-76 [FREE Full text] [doi: 10.1016/j.jsmc.2019.10.002] [Medline: 32005351]
19. Fortier-Brochu, Beaulieu-Bonneau S, Ivers H, Morin C. Insomnia and daytime cognitive performance: a meta-analysis. Sleep Med Rev 2012 Feb;16(1):83-94 [FREE Full text] [doi: 10.1016/j.smrv.2011.03.008] [Medline: 21636297]
20. Hudson AN, Van Dongen HPA, Honn KA. Sleep deprivation, vigilant attention, and brain function: a review. Neuropsychopharmacology 2020 Jan;45(1):21-30 [FREE Full text] [doi: 10.1038/s41386-019-0432-6] [Medline: 31176308]
21. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
22. Hertenstein E, Feige B, Gmeiner T, Kienzler C, Spiegelhalder K, Johann A, et al. Insomnia as a predictor of mental disorders: A systematic review and meta-analysis. Sleep Med Rev 2019 Feb;43:96-105. [doi: 10.1016/j.smrv.2018.10.006] [Medline: 30537570]
23. Li L, Wu C, Gan Y, Qu X, Lu Z. Insomnia and the risk of depression: a meta-analysis of prospective cohort studies. BMC Psychiatry 2016 Nov 05;16(1):375 [FREE Full text] [doi: 10.1186/s12888-016-1075-3] [Medline: 27816065]
24. Wade AG. The societal costs of insomnia. Neuropsychiatr Dis Treat 2010 Dec 20;7:1-18 [FREE Full text] [doi: 10.2147/NDT.S15123] [Medline: 21326650]
25. Daley M, Morin CM, LeBlanc M, Grégoire JP, Savard J. The economic burden of insomnia: direct and indirect costs for individuals with insomnia syndrome, insomnia symptoms, and good sleepers. Sleep 2009 Jan;32(1):55-64 [FREE Full text] [Medline: 19189779]
26. Bramoweth AD, Taylor DJ. Chronic insomnia and health care utilization in young adults. Behav Sleep Med 2012;10(2):106-121. [doi: 10.1080/15402002.2011.587067] [Medline: 22468929]
27. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD, Clinical Guidelines Committee of the American College of Physicians. Management of Chronic Insomnia Disorder in Adults: A Clinical Practice Guideline From the American College of Physicians. Ann Intern Med 2016 Jul 19;165(2):125-133 [FREE Full text] [doi: 10.7326/M15-2175] [Medline: 27136449]
28. Riemann D, Baglioni C, Bassetti C, Bjorvatn B, Dolenc Groselj L, Ellis JG, et al. European guideline for the diagnosis and treatment of insomnia. J Sleep Res 2017 Dec;26(6):675-700 [FREE Full text] [doi: 10.1111/jsr.12594] [Medline: 28875581]

29. Mitchell MD, Gehrman P, Perlis M, Umscheid CA. Comparative effectiveness of cognitive behavioral therapy for insomnia: a systematic review. BMC Fam Pract 2012 May 25;13:40 [FREE Full text] [doi: 10.1186/1471-2296-13-40] [Medline: 22631616]
30. Riemann D, Baum E, Cohrs S, Crönlein T, Hajak G, Hertenstein E, et al. S3-Leitlinie Nicht erholsamer Schlaf/Schlafstörungen. Somnologie 2017 Feb 27;21(1):2-44. [doi: 10.1007/s11818-016-0097-x]
31. Irwin MR, Cole JC, Nicassio PM. Comparative meta-analysis of behavioral interventions for insomnia and their efficacy in middle-aged adults and in older adults 55+ years of age. Health Psychol 2006 Jan;25(1):3-14. [doi: 10.1037/0278-6133.25.1.3] [Medline: 16448292]
32. Trauer JM, Qian MY, Doyle JS, Rajaratnam SMW, Cunnington D. Cognitive Behavioral Therapy for Chronic Insomnia: A Systematic Review and Meta-analysis. Ann Intern Med 2015 Aug 04;163(3):191-204. [doi: 10.7326/M14-2841] [Medline: 26054060]
33. van Straten A, van der Zweerde T, Kleiboer A, Cuijpers P, Morin CM, Lancee J. Cognitive and behavioral therapies in the treatment of insomnia: A meta-analysis. Sleep Med Rev 2018 Apr;38:3-16. [doi: 10.1016/j.smrv.2017.02.001] [Medline: 28392168]
34. Wu JQ, Appleman ER, Salazar RD, Ong JC. Cognitive Behavioral Therapy for Insomnia Comorbid With Psychiatric and Medical Conditions: A Meta-analysis. JAMA Intern Med 2015 Sep;175(9):1461-1472. [doi: 10.1001/jamainternmed.2015.3006] [Medline: 26147487]
35. Edinger JD, Arnedt JT, Bertisch SM, Carney CE, Harrington JJ, Lichstein KL, et al. Behavioral and psychological treatments for chronic insomnia disorder in adults: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. J Clin Sleep Med 2021 Feb 01;17(2):263-298 [FREE Full text] [doi: 10.5664/jcsm.8988] [Medline: 33164741]
36. Wang M, Wang S, Tsai P. Cognitive behavioural therapy for primary insomnia: a systematic review. J Adv Nurs 2005 Jun;50(5):553-564. [doi: 10.1111/j.1365-2648.2005.03433.x] [Medline: 15882372]
37. Manber R, Edinger J, Gress J, San Pedro-Salcedo MG, Kuo T, Kalista T. Cognitive behavioral therapy for insomnia enhances depression outcome in patients with comorbid major depressive disorder and insomnia. Sleep 2008 Apr;31(4):489-495 [FREE Full text] [doi: 10.1093/sleep/31.4.489] [Medline: 18457236]
38. Taylor DJ, Lichstein KL, Weinstock J, Sanford S, Temple JR. A pilot study of cognitive-behavioral therapy of insomnia in people with mild depression. Behav Ther 2007 Mar;38(1):49-57. [doi: 10.1016/j.beth.2006.04.002] [Medline: 17292694]
39. Watanabe N, Furukawa TA, Shimodera S, Morokuma I, Katsuki F, Fujita H, et al. Brief behavioral therapy for refractory insomnia in residual depression: an assessor-blind, randomized controlled trial. J Clin Psychiatry 2011 Dec;72(12):1651-1658. [doi: 10.4088/JCP.10m06130gry] [Medline: 21457679]
40. Kalmbach DA, Cheng P, Arnedt JT, Anderson JR, Roth T, Fellman-Couture C, et al. Treating insomnia improves depression, maladaptive thinking, and hyperarousal in postmenopausal women: comparing cognitive-behavioral therapy for insomnia (CBTI), sleep restriction therapy, and sleep hygiene education. Sleep Med 2019 Mar;55:124-134 [FREE Full text] [doi: 10.1016/j.sleep.2018.11.019] [Medline: 30785053]
41. Koffel E, Bramoweth AD, Ulmer CS. Increasing access to and utilization of cognitive behavioral therapy for insomnia (CBT-I): a narrative review. J Gen Intern Med 2018 Jun 4;33(6):955-962 [FREE Full text] [doi: 10.1007/s11606-018-4390-1] [Medline: 29619651]
42. Manber R, Simpson N. Dissemination of CBT for Insomnia. Curr Sleep Medicine Rep 2016 Jul 9;2(3):136-141. [doi: 10.1007/s40675-016-0048-x]
43. Perils ML, Smith MT. How can we make CBT-I and other BSM services widely available? J Clin Sleep Med 2008 Feb 15;4(1):11-13 [FREE Full text] [Medline: 18350955]
44. Araújo T, Jarrin DC, Leanza Y, Vallières A, Morin CM. Qualitative studies of insomnia: Current state of knowledge in the field. Sleep Med Rev 2017 Feb;31:58-69 [FREE Full text] [doi: 10.1016/j.smrv.2016.01.003] [Medline: 27090821]
45. Cheung JMY, Bartlett DJ, Armour CL, Saini B. Treating Insomnia: A Review of Patient Perceptions Toward Treatment. Behav Sleep Med 2016;14(3):235-266. [doi: 10.1080/15402002.2014.981818] [Medline: 26240937]
46. van Straten A, Cuijpers P. Self-help therapy for insomnia: a meta-analysis. Sleep Med Rev 2009 Feb;13(1):61-71. [doi: 10.1016/j.smrv.2008.04.006] [Medline: 18952469]
47. Wickwire EM. The value of digital insomnia therapeutics: what we know and what we need to know. J Clin Sleep Med 2019 Jan 15;15(1):11-13 [FREE Full text] [doi: 10.5664/jcsm.7558] [Medline: 30621849]
48. Andersson G. Internet interventions: Past, present and future. Internet Interv 2018 Jun;12:181-188 [FREE Full text] [doi: 10.1016/j.invent.2018.03.008] [Medline: 30135782]
49. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
50. Luik AI, van der Zweerde T, van Straten A, Lancee J. Digital Delivery of Cognitive Behavioral Therapy for Insomnia. Curr Psychiatry Rep 2019 Jun 04;21(7):50 [FREE Full text] [doi: 10.1007/s11920-019-1041-0] [Medline: 31161406]
51. van der Zweerde T, Lancee J, Ida Luik A, van Straten A. Internet-Delivered Cognitive Behavioral Therapy for Insomnia: Tailoring Cognitive Behavioral Therapy for Insomnia for Patients with Chronic Insomnia. Sleep Med Clin 2019 Sep;14(3):301-315. [doi: 10.1016/j.jsmc.2019.04.002] [Medline: 31375200]

52. Seyffert M, Lagisetty P, Landgraf J, Chopra V, Pfeiffer PN, Conte ML, et al. Internet-Delivered Cognitive Behavioral Therapy to Treat Insomnia: A Systematic Review and Meta-Analysis. PLoS One 2016;11(2):e0149139 [FREE Full text] [doi: 10.1371/journal.pone.0149139] [Medline: 26867139]
53. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - A systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
54. Behrendt D, Ebert DD, Spiegelhalder K, Lehr D. Efficacy of a Self-Help Web-Based Recovery Training in Improving Sleep in Workers: Randomized Controlled Trial in the General Working Population. J Med Internet Res 2020 Jan 07;22(1):e13346 [FREE Full text] [doi: 10.2196/13346] [Medline: 31909725]
55. Bostock S, Luik AI, Espie CA. Sleep and Productivity Benefits of Digital Cognitive Behavioral Therapy for Insomnia: A Randomized Controlled Trial Conducted in the Workplace Environment. J Occup Environ Med 2016 Jul;58(7):683-689. [doi: 10.1097/JOM.0000000000000778] [Medline: 27257747]
56. de Bruin EJ, Dewald-Kaufmann JF, Oort FJ, Bögels SM, Meijer AM. Differential effects of online insomnia treatment on executive functions in adolescents. Sleep Med 2015 Apr;16(4):510-520 [FREE Full text] [doi: 10.1016/j.sleep.2014.12.009] [Medline: 25796966]
57. Espie CA, Emsley R, Kyle SD, Gordon C, Drake CL, Siriwardena AN, et al. Effect of Digital Cognitive Behavioral Therapy for Insomnia on Health, Psychological Well-being, and Sleep-Related Quality of Life: A Randomized Clinical Trial. JAMA Psychiatry 2019 Jan 01;76(1):21-30 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.2745] [Medline: 30264137]
58. Batterham PJ, Christensen H, Mackinnon AJ, Gosling JA, Thorndike FP, Ritterband LM, et al. Trajectories of change and long-term outcomes in a randomised controlled trial of internet-based insomnia treatment to prevent depression. BJPsych Open 2017 Sep 02;3(5):228-235 [FREE Full text] [doi: 10.1192/bjpo.bp.117.005231] [Medline: 28959453]
59. Christensen H, Batterham PJ, Gosling JA, Ritterband LM, Griffiths KM, Thorndike FP, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. Lancet Psychiatry 2016 Apr;3(4):333-341. [doi: 10.1016/S2215-0366(15)00536-2] [Medline: 26827250]
60. van der Zweerde T, van Straten A, Effting M, Kyle SD, Lancee J. Does online insomnia treatment reduce depressive symptoms? A randomized controlled trial in individuals with both insomnia and depressive symptoms. Psychol. Med 2018 May 11;49(3):501-509. [doi: 10.1017/s0033291718001149]
61. Cheng P, Kalmbach DA, Tallent G, Joseph CL, Espie CA, Drake CL. Depression prevention via digital cognitive behavioral therapy for insomnia: a randomized controlled trial. Sleep 2019 Oct 09;42(10) [FREE Full text] [doi: 10.1093/sleep/zsz150] [Medline: 31535688]
62. Lancee J, van den Bout J, Sorbi MJ, van Straten A. Motivational support provided via email improves the effectiveness of internet-delivered self-help treatment for insomnia: a randomized trial. Behav Res Ther 2013 Dec;51(12):797-805. [doi: 10.1016/j.brat.2013.09.004] [Medline: 24121097]
63. Vedaa, Hagatun S, Kallestad H, Pallesen S, Smith OR, Thorndike FP, et al. Long-Term Effects of an Unguided Online Cognitive Behavioral Therapy for Chronic Insomnia. J Clin Sleep Med 2019 Jan 15;15(1):101-110 [FREE Full text] [doi: 10.5664/jcsm.7580] [Medline: 30621837]
64. Hagatun S, Vedaa, Nordgreen T, Smith ORF, Pallesen S, Havik OE, et al. The Short-Term Efficacy of an Unguided Internet-Based Cognitive-Behavioral Therapy for Insomnia: A Randomized Controlled Trial With a Six-Month Nonrandomized Follow-Up. Behav Sleep Med 2019;17(2):137-155. [doi: 10.1080/15402002.2017.1301941] [Medline: 28345961]
65. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile Phone-Delivered Cognitive Behavioral Therapy for Insomnia: A Randomized Waitlist Controlled Trial. J Med Internet Res 2017 Apr 11;19(4):e70 [FREE Full text] [doi: 10.2196/jmir.6524] [Medline: 28400355]
66. Lorenz N, Heim E, Roetger A, Birrer E, Maercker A. Randomized Controlled Trial to Test the Efficacy of an Unguided Online Intervention with Automated Feedback for the Treatment of Insomnia. Behav Cogn Psychother 2019 May;47(3):287-302 [FREE Full text] [doi: 10.1017/S1352465818000486] [Medline: 30185239]
67. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a Web-Based Cognitive Behavior Therapy for Insomnia Intervention With 1-Year Follow-up: A Randomized Clinical Trial. JAMA Psychiatry 2017 Jan 01;74(1):68-75. [doi: 10.1001/jamapsychiatry.2016.3249] [Medline: 27902836]
68. Vincent N, Lewycky S. Logging on for better sleep: RCT of the effectiveness of online treatment for insomnia. Sleep 2009 Jun;32(6):807-815 [FREE Full text] [doi: 10.1093/sleep/32.6.807] [Medline: 19544758]
69. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
70. Choi YK, Demiris G, Lin S, Iribarren SJ, Landis CA, Thompson HJ, et al. Smartphone Applications to Support Sleep Self-Management: Review and Evaluation. J Clin Sleep Med 2018 Oct 15;14(10):1783-1790 [FREE Full text] [doi: 10.5664/jcsm.7396] [Medline: 30353814]

71. Rajabi Majd N, Broström A, Ulander M, Lin C, Griffiths MD, Imani V, et al. Efficacy of a Theory-Based Cognitive Behavioral Technique App-Based Intervention for Patients With Insomnia: Randomized Controlled Trial. J Med Internet Res 2020 Apr 01;22(4):e15841 [FREE Full text] [doi: 10.2196/15841] [Medline: 32234700]
72. Crönlein T, Langguth B, Popp R, Lukesch H, Pieh C, Hajak G, et al. Regensburg Insomnia Scale (RIS): a new short rating scale for the assessment of psychological symptoms and sleep in insomnia; study design: development and validation of a new short self-rating scale in a sample of 218 patients suffering from insomnia and 94 healthy controls. Health Qual Life Outcomes 2013 Apr 22;11:65 [FREE Full text] [doi: 10.1186/1477-7525-11-65] [Medline: 23601161]
73. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
74. Trockel M, Manber R, Chang V, Thurston A, Taylor CB. An e-mail delivered CBT for sleep-health program for college students: effects on sleep quality and depression symptoms. J Clin Sleep Med 2011 Jun 15;7(3):276-281 [FREE Full text] [doi: 10.5664/JCSM.1072] [Medline: 21677898]
75. Okajima I, Tanizawa N, Harata M, Suh S, Yang C, Li SX, et al. Can an E-Mail-Delivered CBT for Insomnia Validated in the West Be Effective in the East? A Randomized Controlled Trial. Int J Environ Res Public Health 2021 Dec 24;19(1):186 [FREE Full text] [doi: 10.3390/ijerph19010186] [Medline: 35010445]
76. Minddistrict. URL: https://minddistrict.com [accessed 2023-02-20]
77. Borbély AA. A two process model of sleep regulation. Hum Neurobiol 1982;1(3):195-204. [Medline: 7185792]
78. Johansson O, Michel T, Andersson G, Paxling B. Experiences of non-adherence to Internet-delivered cognitive behavior therapy: A qualitative study. Internet Interventions 2015 May;2(2):137-142 [FREE Full text] [doi: 10.1016/j.invent.2015.02.006]
79. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
80. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry 2007;29(5):388-395. [doi: 10.1016/j.genhosppsych.2007.06.004] [Medline: 17888804]
81. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J Affect Disord 2004 Jul;81(1):61-66. [doi: 10.1016/S0165-0327(03)00198-8] [Medline: 15183601]
82. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the working alliance inventory. Psychotherapy Research 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
83. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2010;17(3):231-239. [doi: 10.1002/cpp.658] [Medline: 20013760]
84. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research & Practice 1979;16(3):252-260. [doi: 10.1037/h0085885]
85. Beunckens C, Molenberghs G, Kenward MG. Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials. Clin Trials 2005;2(5):379-386. [doi: 10.1191/1740774505cn119oa] [Medline: 16315646]
86. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
87. Morris SB. Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. Organizational Research Methods 2007 Jul 23;11(2):364-386. [doi: 10.1177/1094428106291059]
88. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
89. Blom K, Jernelöv S, Lindefors N, Kaldo V. Facilitating and hindering factors in Internet-delivered treatment for insomnia and depression. Internet Interv 2016 May;4:51-60 [FREE Full text] [doi: 10.1016/j.invent.2016.03.004] [Medline: 30135790]
90. Lancee J, van den Bout J, van Straten A, Spoormaker VI. Internet-delivered or mailed self-help treatment for insomnia?: a randomized waiting-list controlled trial. Behav Res Ther 2012 Jan;50(1):22-29. [doi: 10.1016/j.brat.2011.09.012] [Medline: 22055281]
91. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2010;17(3):231-239. [doi: 10.1002/cpp.658] [Medline: 20013760]
92. Espie CA, Kyle SD, Williams C, Ong JC, Douglas NJ, Hames P, et al. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. Sleep 2012 Jun 01;35(6):769-781 [FREE Full text] [doi: 10.5665/sleep.1872] [Medline: 22654196]
93. Lappalainen P, Langrial S, Oinas-Kukkonen H, Muotka J, Lappalainen R. ACT for sleep - Internet-delivered self-help ACT for sub-clinical and clinical insomnia: A randomized controlled trial. Journal of Contextual Behavioral Science 2019 Apr;12:119-127 [FREE Full text] [doi: 10.1016/j.jcbs.2019.04.001]

94. Ho FY, Chung K, Yeung W, Ng TH, Kwan K, Yung K, et al. Self-help cognitive-behavioral therapy for insomnia: a meta-analysis of randomized controlled trials. Sleep Med Rev 2015 Feb;19:17-28. [doi: 10.1016/j.smrv.2014.06.010] [Medline: 25104471]
95. Buysse DJ, Ancoli-Israel S, Edinger JD, Lichstein KL, Morin CM. Recommendations for a standard research assessment of insomnia. Sleep 2006 Sep;29(9):1155-1173. [doi: 10.1093/sleep/29.9.1155] [Medline: 17040003]
96. Gellis LA, Lichstein KL, Scarinci IC, Durrence HH, Taylor DJ, Bush AJ, et al. Socioeconomic status and insomnia. J Abnorm Psychol 2005 Feb;114(1):111-118. [doi: 10.1037/0021-843X.114.1.111] [Medline: 15709817]
97. Lallukka T, Sares-Jäske L, Kronholm E, Sääksjärvi K, Lundqvist A, Partonen T, et al. Sociodemographic and socioeconomic differences in sleep duration and insomnia-related symptoms in Finnish adults. BMC Public Health 2012 Jul 28;12:565 [FREE Full text] [doi: 10.1186/1471-2458-12-565] [Medline: 22839359]
98. Drake CL. The Promise of Digital CBT-I. Sleep 2016 Jan 01;39(1):13-14 [FREE Full text] [doi: 10.5665/sleep.5304] [Medline: 26564140]
99. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
100. Espie CA. "Stepped care": a health technology solution for delivering cognitive behavioral therapy as a first line insomnia treatment. Sleep 2009 Dec;32(12):1549-1558 [FREE Full text] [doi: 10.1093/sleep/32.12.1549] [Medline: 20041590]
101. Horsch C, Lancee J, Beun RJ, Neerincx MA, Brinkman W. Adherence to Technology-Mediated Insomnia Treatment: A Meta-Analysis, Interviews, and Focus Groups. J Med Internet Res 2015 Sep 04;17(9):e214 [FREE Full text] [doi: 10.2196/jmir.4115] [Medline: 26341671]
102. Wildeboer G, Kelders SM, van Gemert-Pijnen JEWC. The relationship between persuasive technology principles, adherence and effect of web-Based interventions for mental health: A meta-analysis. Int J Med Inform 2016 Dec;96:71-85. [doi: 10.1016/j.ijmedinf.2016.04.005] [Medline: 27117057]
103. van Straten A, Lancee J. Digital cognitive behavioural therapy for insomnia: the answer to a major public health issue? Lancet Digit Health 2020 Aug;2(8):e381-e382 [FREE Full text] [doi: 10.1016/S2589-7500(20)30167-9] [Medline: 33328041]

Abbreviations

CBT-I: cognitive behavioral therapy for insomnia
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th edition
ITT: intention-to-treat
PHQ-9: 9-item Patient Health Questionnaire
REDCap: Research Electronic Data Capture
RIS: Regensburg Insomnia Scale
WAI-SR: Working Alliance Inventory-Short Revised

Edited by J Torous; submitted 30.05.22; peer-reviewed by B Bjorvatn, RJ Beun; comments to author 27.07.22; revised version received 15.09.22; accepted 30.09.22; published 21.03.23.

Please cite as:
Vollert B, Müller L, Jacobi C, Trockel M, Beintner I
Effectiveness of an App-Based Short Intervention to Improve Sleep: Randomized Controlled Trial
JMIR Ment Health 2023;10:e39052
URL: https://mental.jmir.org/2023/1/e39052
doi:10.2196/39052
PMID:36943337

Original Paper

Situating Meditation Apps Within the Ecosystem of Meditation Practice: Population-Based Survey Study

Sin U Lam[1,2*], MA; Qiang Xie[1,2*], MEd; Simon B Goldberg[1,2], PhD

[1]Department of Counseling Psychology, University of Wisconsin-Madison, Madison, WI, United States

[2]Center for Healthy Minds, University of Wisconsin-Madison, Madison, WI, United States

[*]these authors contributed equally

Corresponding Author:
Simon B Goldberg, PhD
Department of Counseling Psychology
University of Wisconsin-Madison
335 Education Building
1000 Bascom Mall
Madison, WI, 53706
United States
Phone: 1 6082624807
Email: sbgoldberg@wisc.edu

Abstract

Background: Meditation apps have the potential to increase access to evidence-based strategies to promote mental health. However, it is currently unclear how meditation apps are situated within the broader landscape of meditation practice and what factors may influence engagement with them.

Objective: This study aimed to clarify the prevalence and correlates of meditation app use in a population-based sample of individuals with lifetime exposure to meditation in the United States. In addition, we sought to identify the concerns and desired features of meditation apps among those with lifetime exposure to meditation.

Methods: A total of 953 participants completed an initial screening survey. Of these 953 participants, 434 (45.5%) reported lifetime exposure to meditation and completed a follow-up survey (434/470, 92.3% response rate) assessing their meditation app use, anxiety, depression, loneliness, initial motivation for meditation, and concerns about and desired features of meditation apps.

Results: Almost half (434/953, 45.5%) of the participants who completed the screening survey reported lifetime exposure to meditation. Among those with lifetime exposure to meditation (ie, meditators), more than half (255/434, 58.8%) had used meditation apps at least once in their lives, and 21.7% (94/434) used meditation apps weekly or daily (ie, active users). Younger age, higher anxiety, and a mental health motivation for practicing meditation were associated with lifetime exposure to meditation apps. Among meditators, those with lifetime exposure to meditation apps were more likely to report concerns about apps, including concerns regarding the cost and effectiveness of apps, time required for use, technical issues with apps, and app user-friendliness. Meditators who used meditation apps weekly or daily (ie, active users) were younger, less likely to be men and non-Latinx White individuals and have lower income, and more likely to have an initial spiritual motivation for meditation. Active users reported more concerns regarding usability and technical problems and were less likely to report disinterest in apps. Headspace and Calm were the most frequently used apps. Tips and reminders for practice, encouragement of “mini” practices, and mental health content were the most desired features. Participants were less interested in social features (eg, the ability to communicate with other users or teachers).

Conclusions: Meditation apps are commonly used by meditators in the United States, with a higher use among certain demographic groups. Future studies may increase user engagement in meditation apps by addressing concerns (eg, cost and effectiveness) and incorporating desired features (eg, tips and reminders for practice).

(JMIR Ment Health 2023;10:e43565) doi:10.2196/43565

KEYWORDS

mindfulness; meditation; consumer behavior; user engagement; mobile health; digital health; mobile phone

Introduction

Background

There have been dramatic innovations in digital technology over the past 10 years. According to the Pew Research Center, 85% of Americans owned a smartphone as of 2021 compared with only 35% in 2011 [1]. Smartphones are increasingly used to deliver health interventions [2]. Smartphone-delivered interventions may be particularly helpful in expanding access to health services and integrating health support into people's daily lives [3,4]. The potential for smartphone-delivered interventions has been obvious during the COVID-19 pandemic, with the increasing need for mental health care coupled with a shortage of mental health services [3]. There is growing evidence supporting the efficacy of smartphone-delivered interventions for both physical and mental health [5-10].

Meditation is an umbrella term for various forms of mental training focused on the cultivation of attentional and affective regulatory skills that promote well-being [11]. Meditation has become increasingly popular in the United States (US) over the past 10 years. According to the National Health Interview Survey (NHIS) conducted by the Centers for Disease Control and Prevention, the percentage of adults in the US who used meditation in the last 12 months increased from 6.5% in 2012 to 21.1% in 2017 [12,13]. Although there is a wide variety of types of meditation, for the purposes of the NHIS and this study, meditation was operationalized as mindfulness, mantra, and spiritual meditation, which are 3 popular forms [14]. Research suggests that meditation may be beneficial for mental and physical health [15-20].

Advancements in digital technology may have contributed to the increased popularity of meditation over the past 10 years. Meditation practices figure prominently in the mental health app landscape. A recent study found that 2 popular meditation apps (Headspace and Calm) alone accounted for 96% of daily active mental health app users [21]. Moreover, there is meta-analytic evidence suggesting that app-delivered meditation interventions significantly reduce psychological distress and improve quality of life [22].

Although meditation apps potentially increase access to evidence-based psychological strategies and demonstrate promising effects on health outcomes, they are not without limitations. Notably, meditation apps, as with other smartphone-delivered interventions, have notoriously poor user retention, which can limit the interventions' potential benefits in the long term and may undermine the validity of studies evaluating these interventions [23,24]. Given the potentially negative impact of poor retention, it is important to investigate why people do or do not continue using meditation apps.

Recent studies on meditation apps and other smartphone-delivered interventions have begun exploring perceptions of these interventions, the demographic and clinical characteristics of users, and the role of these factors in dropout and engagement patterns. For example, recent studies suggest that app users frequently report concerns about the security and privacy of apps, which may be barriers to continued use [25,26]. Huberty et al [27] investigated the user characteristics and use patterns of the Calm meditation app among paid subscribers. They found that participating subscribers were mainly White individuals and women, the most common reasons for trying Calm were sleep difficulties and psychological distress, and those with sleep difficulties used Calm more frequently. Although an intriguing first look at associations between user variables and use, the findings of Huberty et al [27] may or may not generalize to those who use other meditation apps given that different meditation apps may have different user bases and features that may or may not affect the findings (eg, the Liberate app, which was specifically designed for Black individuals [28]).

To date, the role of meditation apps within the broader landscape of meditation practice in the US has not been clarified. This is an important area to clarify given the promising effects of meditation training on health [17,29] and the ubiquity of meditation practices within mental health apps [21]. It is currently unclear how frequently meditators are exposed to meditation through apps, who is being exposed in this way, and what factors are associated with persistence with app-based meditation training. In addition, the concerns that meditators may have regarding meditation apps and the features that may be desired by potential users have not been investigated. Insights gained on these topics could collectively be used to inform the ongoing development and dissemination of app-based meditation interventions and maximize the public health impact of these interventions.

Objectives

To address these questions, we conducted a web-based survey using a population-based sampling method in the US. We had 4 primary aims. First, we sought to provide estimates of the prevalence and patterns of meditation app use in a demographically representative sample of meditators. In this study, *meditators* refers to people with lifetime exposure to NHIS-defined meditation (ie, mindfulness, mantra, or spiritual meditation) [14]. We focused on meditators instead of the general population (ie, meditators and nonmeditators) as we were interested in how meditation apps are situated within the broader landscape of meditation practice and factors that might influence engagement with meditation apps. Second, we aimed to examine the demographic (eg, gender and race or ethnicity) and psychological (eg, anxiety and depression) correlates of meditation app use. Third, we sought to understand which meditation apps are being used by meditators. Finally, we aimed to characterize the concerns and desired features of meditation apps among meditators. Given the limited existing literature, this study was exploratory and we had no a priori hypotheses.

Methods

Ethics Approval

The study procedures were approved by the Institutional Review Board of the University of Wisconsin–Madison (reference 2020-1368).

Participants

We recruited participants using the Prolific platform (Prolific Academic Ltd) [30] in November and December 2020. Prolific is a recruitment platform that has been shown to include participants who are more diverse, less dishonest, and less familiar with research materials than other web-based recruitment platforms (eg, Amazon Mechanical Turk) [31,32]. Using Prolific's representative sampling procedure, we recruited participants based on their age, sex, and race in proportion to the US Census data.

In this sample, 96% (953/993) of the participants completed our screening survey, passed an attention check ("I have been randomly selecting responses on this survey"), and indicated whether they had practiced meditation in their lifetime. A comparison of the overall sample with the US adult population [33] showed that this sample was more educated (482/953, 50.6% with a bachelor's degree or higher vs 32.1% in the 2015-2019 US Census), older (median age 44 years vs 38 years), and wealthier (median income US $40,000 vs US $34,103). Although some racial and ethnic groups (Asian, Black, and multiracial) were represented in proportion to the US population, non-Latinx White participants were overrepresented (673/953, 70.6% vs 60.7%), and Latinx participants were underrepresented (57/953, 6% vs 18%). The low Latinx representation was most likely the result of Prolific matching based on race but not ethnicity (for sample demographics, see Multimedia Appendix 1).

Almost half (470/953, 49.3%) of the participants who underwent screening reported exposure to meditation (ie, mindfulness, mantra, or spiritual meditation) at some point in their lifetime. Meditators (ie, those with lifetime exposure to meditation) were invited to complete a follow-up survey that assessed various aspects of their meditation practice, including their experience with meditation apps. Those who completed the follow-up survey and passed a second attention check ("Please select the leftmost response") formed the primary analytic sample for this study (n=434). This sample represented most of those invited to complete the follow-up survey (434/470, 92.3% response rate). We compared the demographics of those who completed the follow-up survey with the demographics of those who were invited but did not complete the follow-up survey using a correlation coefficient for ease of interpretation. Of note, a correlation between 2 dichotomous variables is a special case of the generalized Pearson correlation coefficient [34]. The results showed that non-Latinx White participants completed the follow-up survey at a higher rate (r=0.15; P=.001), although no significant differences were found for other demographic variables (age, gender, education, or income; r=−0.05 to 0.05; P>.28 in all cases).

Measures

Demographics

Participants were asked to provide their age, gender identity, race and ethnicity, highest degree of education, and annual income (for the items assessed, see Multimedia Appendix 2). The following demographic variables were dichotomized as covariates in our models: gender (not men as the reference group), race and ethnicity (racial and ethnic minority as the reference group), education (not college graduate as the reference group), and annual income (income below the US population per capita median [US $34,103] as the reference group [33]).

Meditation App Use

Participants who completed the follow-up survey (ie, meditators) were asked about their meditation practice frequency and experience with meditation apps. We assessed the specific meditation apps that meditators used, the specific apps they used the most, and their meditation app use frequency (1=*never*, 2=*several times per year*, 3=*monthly*, 4=*weekly*, and 5=*daily*). Participants were asked to indicate any apps they had used. Options for text entry were provided for participants to indicate meditation apps that were not on the list. The list of meditation apps was adopted from the study by Carlo et al [35] and included 22 popular meditation apps (eg, Headspace, Calm, and Insight Timer; Multimedia Appendix 3).

Psychological Measures

On the basis of prior work indicating that individuals with elevated mental health symptoms were more likely to access meditation apps [27] and evidence that elevated symptoms may be associated with a greater risk of adverse responses to meditation, which could decrease persistence [36,37], we assessed the symptoms of depression, anxiety, and loneliness.

The Patient-Reported Outcomes Measurement Information System Depression and Anxiety scales were used to assess depression (eg, "I felt worthless") and anxiety (eg, "I felt fearful") symptoms [38]. Participants responded to each item using a 5-point Likert-type scale (1=*never*; 5=*always*). Higher total scores reflected higher levels of anxiety and depressive symptoms. Internal consistency was adequate in this sample (Cronbach α=.93 and .90 for the Patient-Reported Outcomes Measurement Information System Depression and Anxiety scales, respectively).

The 5-item National Institutes of Health Toolbox Loneliness Scale [39] was used to assess loneliness. Participants were asked to rate their experience of loneliness in the past week (eg, "I feel alone") on a 5-point Likert-type scale (1=*never*; 5=*always*). Internal consistency was adequate (Cronbach α=.94) in this sample.

Meditation Motivation

Participants were asked to respond to items adapted from the study by Pepping et al [40] to assess motivation for meditation practice. Participants indicated which factors motivated them to begin their meditation practice. Options included physical health, emotional health or stress reduction, sociocultural and spiritual reasons, and "other" to indicate reasons not listed previously (Multimedia Appendix 3).

Concerns About Meditation Apps

To assess participants' concerns about meditation apps, we adopted 10 items from the studies by Stoyanov et al [41], Torous et al [42,43], and Kenny et al [44]. Example items include "cost of apps," "I am not interested in them," "they do not target or help with my goals," and "technical problems" (Multimedia

Appendix 3). Participants could select multiple concerns. We evaluated associations between app use and individual concerns, as well as with the total number of concerns.

Desired Features of Meditation Apps

Participants were asked to indicate to what extent they were interested in a list of different meditation app features on a 6-point Likert-type scale (1=*not at all*; 6=*a great deal*). The sample list of suggested items was generated based on relevant criteria from the Mobile App Rating Scale [41] as well as features included in existing apps or that could be included in future apps. Sample items included were "ability to connect with other users," "having content related to depression and anxiety," "having tips for daily life practice," and "the ability to text with a meditation coach."

Data Analysis

We first calculated the percentage of meditators (ie, those with lifetime exposure to meditation) who indicated having ever used a meditation app and those reporting active app use (weekly or daily) to provide an estimate of the prevalence of meditation app use among meditators. We then examined sociodemographic (age, gender, race and ethnicity, and income), psychological (depression, anxiety, and loneliness), and motivational (initial motivation for practice) characteristics and concerns regarding meditation apps as correlates of lifetime exposure to meditation apps and active app use. Of note, these analyses were conducted with participants who reported lifetime exposure to meditation and completed the follow-up survey (434/953, 45.5%). Correlations were used to provide a comparable effect size across models given that correlations with dichotomous outcomes or predictors are special cases of the Pearson coefficient [34]. A multiple regression model with non-Latinx as the reference group was conducted to examine associations between active app use and specific racial and ethnic subgroups. In total, 2 sets of sensitivity analyses were conducted: one controlling for demographics and one removing outliers (values 3 SDs from the mean). We used partial correlations for sensitivity analyses by controlling for demographics. Instances where the results changed in the sensitivity analyses were noted. Given that we examined several potential correlates of meditation app use, we controlled for false discovery rate (FDR) using the method by Benjamini and Hochberg [45]. *P* values reported in the text were FDR adjusted [45].

Results

Demographics

The meditator sample (ie, those with lifetime exposure to meditation) comprised predominantly women (237/434, 54.6%), with 43.5% (189/434) men and 1.8% (8/434) nonbinary or transgender-identifying participants. Most participants (315/434, 72.6%) identified as non-Latinx White individuals, 12% (52/434) identified as African American individuals, 5.3% (23/434) identified as Latinx individuals, 6.7% (29/434) identified as Asian individuals, 0.5% (2/434) identified as Native American individuals, and 3% (13/434) identified as multiracial. Meditators were, on average, aged 43.77 (SD 15.53) years. The average income was US $54,389.56 (SD US $60,126.77; median US $40,000). Less than half (187/434, 43.1%) of the participants had an annual income below the median for the US population (US $34,103) [33].

Sample descriptive statistics of meditators in this study (434/953, 45.5%) are shown in Multimedia Appendix 1. No continuous variables were found to deviate from normality outside the recommended ranges (ie, skewness <2.00 and kurtosis <7.00 [46]). Specifically, skewness ranged from −1.01 to 0.70 and kurtosis ranged from −1.17 to 0.75 in this study. As reported by Goldberg et al [36], the distribution of lifetime hours of meditation practice was as follows: 0 to 10 (82/434, 18.9%), 11 to 100 (179/434, 41.2%), 101 to 500 (71/434, 16.4%), 501 to 1000 (39/434, 9%), 1001 to 5000 (31/434, 7.1%), and ≥5001 (32/434, 7.4%).

Meditation App Use

Among meditators (ie, those with lifetime exposure to meditation) who completed the follow-up survey (434/953, 45.5%), more than half (255/434, 58.8%) indicated having used a meditation app in their lifetime. The frequency of current app use among meditators was 7.1% (31/434) daily, 14.5% (63/434) weekly, 10.6% (46/434) monthly, and 20.7% (90/434) several times per year, with the remaining participants reporting either no lifetime exposure to a meditation app (179/434, 41.2%) or no use in the past year (25/434, 5.8%). Mediators with current daily or weekly app use were considered active app users.

Participants with lifetime exposure to meditation apps indicated their most frequently used apps. The most frequently used meditation apps were Headspace (76/255, 29.8%), Calm (73/255, 28.6%), The Mindfulness App (26/255, 10.2%), and Insight Timer (26/255, 10.2%). For those who reported practicing meditation with their most frequently used app weekly, average practice time per week was 48.35 (SD 57.78; range 5 to 420) minutes. For those who reported practicing meditation with their most frequently used app daily, average practice time per day was 24.58 (SD 17.34; range 2 to 90) minutes.

Active app users (weekly or daily use) were predominantly women (61/94, 65%), with 34% (32/94) being men and 1% (1/94) being nonbinary or transgender-identifying individuals. Most active app users (56/94, 60%) identified as non-Latinx White individuals, 22% (21/94) identified as African American individuals, 5% (5/94) identified as Latinx American individuals, 5% (5/94) identified as Asian American individuals, 1% (1/94) identified as Native American individuals, and 6% (6/94) identified as multiracial individuals. The average age of active meditation app users was 40.39 (SD 13.72) years, and the average income was US $67,457 (SD US $84,516; median US $47,500).

Correlates of Lifetime Exposure to Meditation Apps

Younger participants were more likely to have tried a meditation app (r=−0.24; P<.001). None of the other sociodemographic variables (ie, gender, race and ethnicity, educational background, and income conditions) was associated with lifetime exposure to meditation apps (Table 1 and Figure 1). Anxiety was correlated with increased likelihood of having used a meditation app (r=0.20; P<.001). Depression was also correlated with

increased likelihood of having used a meditation app (r=0.13; P=.005), although this association was no longer significant after controlling for demographic variables. Loneliness was not significantly associated with meditation app use (r=0.09; P=.051). Those who initially practiced meditation for mental health or stress reduction motives were more likely to have used a meditation app (r=0.20; P<.001). In contrast, individuals who first tried meditation practice for social, cultural, or religious reasons were less likely to try meditation apps (r=−0.12; P=.02). Other meditation motives were not associated with lifetime exposure to meditation apps: trying meditation for physical health purposes (r=0.04; P=.36), for general spiritual or self-transformation (r=−0.05; P=.31), and for achieving other ultimate goals (eg, enlightenment or awakening; r=0.02; P=.70).

Participants reporting a larger total number of concerns regarding meditation apps were more likely to have used a meditation app (r=0.18; P=.02). Similarly, most concerns were associated with increased likelihood of having used a meditation app: cost of apps (r=0.25; P<.001), time required for use (r=0.29; P<.001), technical issues (r=0.17; P<.001), not being user-friendly (r=0.13; P<.001), and not being targeted for personal goals (r=0.13; P<.001). The only concern that was negatively associated with lifetime exposure to a meditation app was being disinterested in them (r=−0.32; P<.001). The remaining concerns (doubt about effectiveness, not recommended by a health care provider, trustworthiness of apps as a source of information, and concerns about the security of health data) were not associated with lifetime exposure to meditation apps (r=−0.02 to 0.09; P>.05 in all cases).

Table 1. Correlations between participant factors and meditation app use (n=434).

Variable	Any lifetime use			Active app use		
	r	*P* value	FDR[a]-adjusted *P* value	*r*	*P* value	FDR-adjusted *P* value
Demographics						
Age	−0.24	<.001[b]	<.001[b]	−0.11	.02[c]	.03[c]
Men	−0.05	.32	.27	−0.10	.04[c]	.02[c]
Non-Latinx White	−0.08	.08	.76	−0.15	.001[d]	.02[c]
Bachelor's degree	0.07	.12	.09	0.03	.56	.72
Low income	−0.06	.25	.16	−0.12	.01[c]	.005[d]
Psychological factors						
Depression	0.13	.005[d]	.07	0.02	.61	.59
Anxiety	0.20	<.001[b]	.002[d]	0.05	.30	.46
Loneliness	0.09	.05	.29	0.01	.79	.79
Initial motivation for meditation						
Physical motivation[e]	0.04	.36	.32	0.09	.05	.03[c]
Mental motivation[f]	0.20	<.001[b]	.003[d]	0.07	.16	.53
Cultural motivation[g]	−0.12	.02[c]	.047[c]	0.02	.73	.80
Spiritual motivation[h]	−0.05	.31	.36	0.11	.03[c]	.04[c]
Awakening motivation[i]	0.02	.70	.48	0.07	.18	.11
Concerns about meditation apps						
Total concerns[j]	0.18	<.001[b]	.001[d]	−0.01	.78	.67
Cost concerns	0.25	<.001[b]	<.001[b]	0.04	.46	.54
Time concerns	0.29	<.001[b]	<.001[b]	0.001	.98	.66
Effectiveness concerns[k]	0.03	.48	.91	−0.05	.29	.26
Recommendation concerns[l]	−0.02	.72	.83	−0.06	.22	.35
Interest concerns[m]	−0.32	<.001[b]	<.001[b]	−0.23	<.001[b]	<.001[b]
Security concerns[n]	0.09	.05	.15	0.04	.47	.85
Target concerns[o]	0.17	<.001[b]	.002[d]	0.05	.29	.45
Usability concerns[p]	0.13	.006[d]	.02[c]	0.15	.002[d]	.004[d]
Technical problem concerns[q]	0.13	.005[d]	.02[c]	0.13	.005[d]	.01[c]
Trust concerns[r]	0.07	.17	.15	−0.02	.74	.86

[a]FDR: false discovery rate.

[b]*P*<.001.

[c]*P*<.05.

[d]*P*<.01.

[e]Physical motivation: physical health motivations for practice.

[f]Mental motivation: mental or emotional health or stress reduction motivations for practice.

[g]Cultural motivation: social, cultural, or religious identity motivations for practice.

[h]Spiritual motivation: general spiritual or self-transformation motivations for practice.

[i]Awakening motivation: enlightenment, awakening, nirvana, or other ultimate goal motivations for practice.

[j]Total concerns: the total number of concerns.

[k]Effectiveness concerns: unsure whether apps are effective.

[l]Recommendation concerns: not recommended by a health care provider.

[m]Interest concerns: not interested in using apps.

[n]Security concerns: concerns regarding the security of health data.

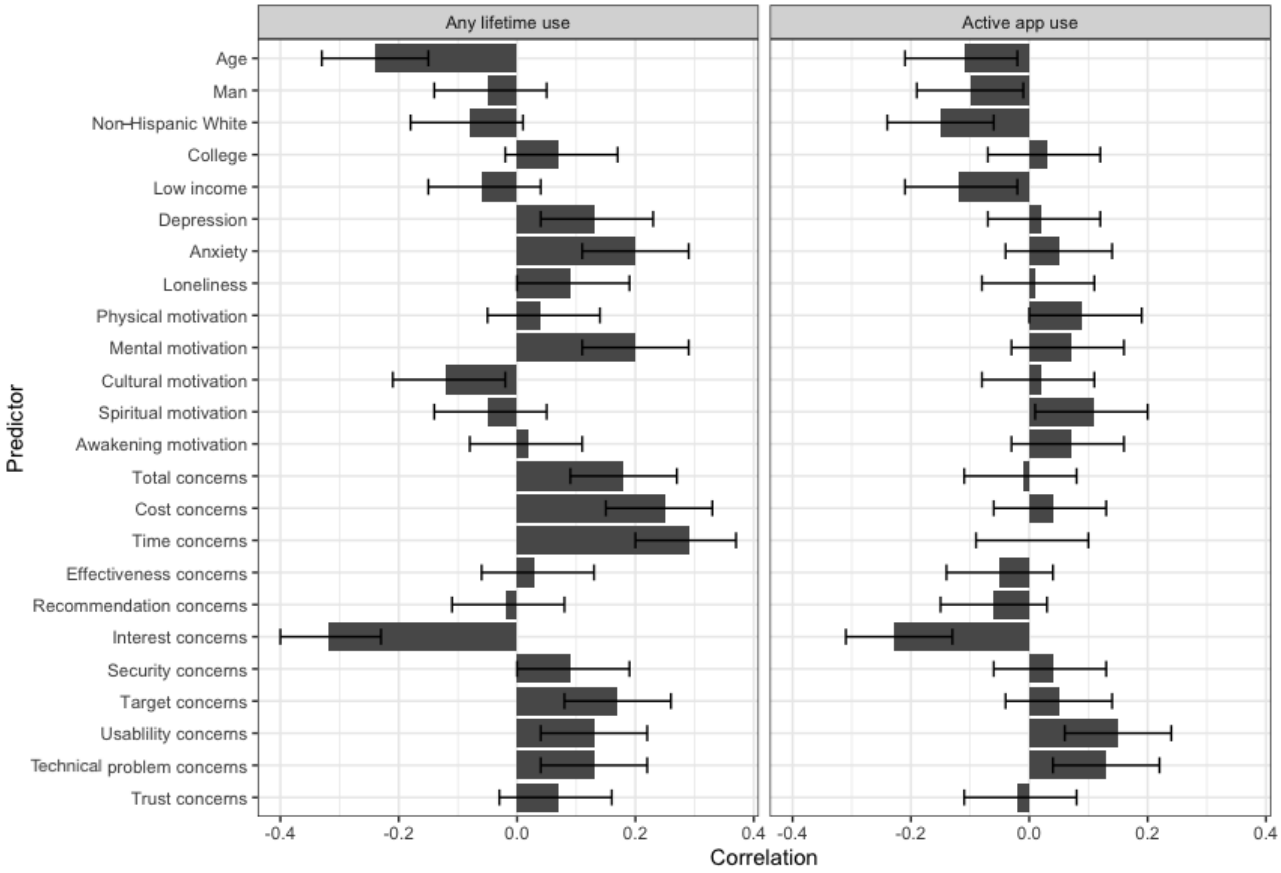
[o]Target concerns: not targeting or helping with personal goals.

[p]Usability concerns: apps not being user-friendly.

[q]Technical problem concerns: experiencing technical problems.

[r]Trust concerns: apps not being a trustworthy source of information.

Figure 1. Correlations between participant factors and meditation app use (ie, lifetime and active use; n=434). The bars indicate correlation coefficients with 95% CIs. See Table 1 for these data in tabular format. Active app use: weekly or daily use; awakening motivation: enlightenment, awakening, nirvana, or other ultimate goal motivations for practice; cultural motivation: social, cultural, or religious identity motivations for practice; effectiveness concerns: unsure whether apps are effective; interest concerns: not interested in using apps; mental motivation: mental or emotional health or stress reduction motivations for practice; physical motivation: physical health motivations for practice; recommendation concerns: not recommended by a health care provider; security concerns: concerns regarding the security of health data; target concerns: not targeting or helping with personal goals; technical problem concerns: experiencing technical problems; trust concerns: apps not being a trustworthy source of information; usability concerns: apps not being user-friendly.



Correlates of Active App Use

Correlational analysis results surviving FDR *P* value correction indicated that active app users (weekly or daily) were less likely to be non-Latinx White (r=−0.15; P<.001). Active app users were also less likely to indicate that they were disinterested in meditation apps (r=−0.23; P<.001) and more likely to endorse concerns regarding apps not being user-friendly (r=0.15; P=.002) and having technical problems (r=0.13; P=.005). Other demographic factors such as age, gender, education, and income were not associated with active app use. The associations between active app use and other concerns regarding meditation apps, the total number of concerns, and motives for meditation practices did not survive *P* value correction (Table 1).

To further examine the racial and ethnic demographics of active meditation app users, a multiple regression model was conducted using non-Latinx White participants as a reference group. The results showed that Black (β=.18; P<.001) and multiracial (β=.12; P=.01) participants were more likely to be active app users relative to non-Latinx White participants. This effect persisted after controlling for age, gender, education, and income status (β=.16 and .11 and P=.001 and .02 for Black and multiracial participants relative to non-Latinx White participants, respectively). There were no significant differences between other racial and ethnic groups and non-Latinx White participants in terms of active use of meditation apps (Table 2).

Table 2. Associations between demographic variables and active app use (n=434).

Demographics	β (SE[a])	95% CI	*P* value
Race and ethnicity[b]			
Black	.16 (0.05)	0.07 to 0.25	.001
Latinx	.01 (0.05)	−0.09 to 0.10	.90
Asian	−.04 (0.05)	−0.13 to 0.06	.46
Native American	.03 (0.05)	−0.06 to 0.13	.47
Multiracial	.11 (0.05)	0.02 to 0.21	.02
Age	−.11 (0.05)	−0.21 to −0.01	.03
Men	−.11 (0.05)	−0.20 to −0.02	.02
Low income	−.16 (0.05)	−0.26 to −0.06	.002
College	.00 (0.05)	−0.10 to −0.10	.95

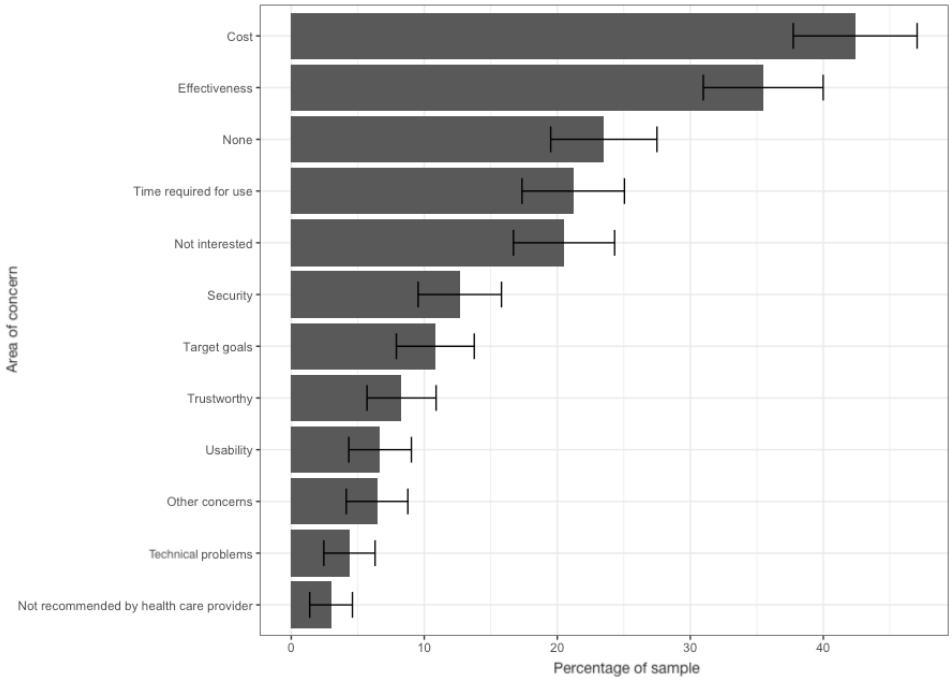
[a]SE rounded to 2 digits.

[b]Non-Latinx White participants as a reference group.

Concerns About Meditation Apps

The most common concern regarding meditation apps among meditators was the cost of the apps (184/434, 42.4%; Figure 2). Doubts regarding app effectiveness (154/434, 35.5%), time required for use (92/434, 21.2%), and lack of interest (89/434, 20.5%) were also commonly endorsed. Fewer participants endorsed concerns about app usability (ie, not being user-friendly; 29/434, 6.7%), technical problems (19/434, 4.4%), and apps not being recommended by a health care provider (13/434, 3%; Figure 2; Multimedia Appendix 4).

Figure 2. Concerns about meditation apps among meditators (n=434). The bars indicate the percentage of the sample reporting a given concern along with 95% CIs. Effectiveness: unsure whether apps are effective; security: concerns regarding the security of health data; target goals: not targeting or helping with personal goals; technical problems: experiencing technical problems; trustworthy: apps not being a trustworthy source of information; usability: apps not being user-friendly.

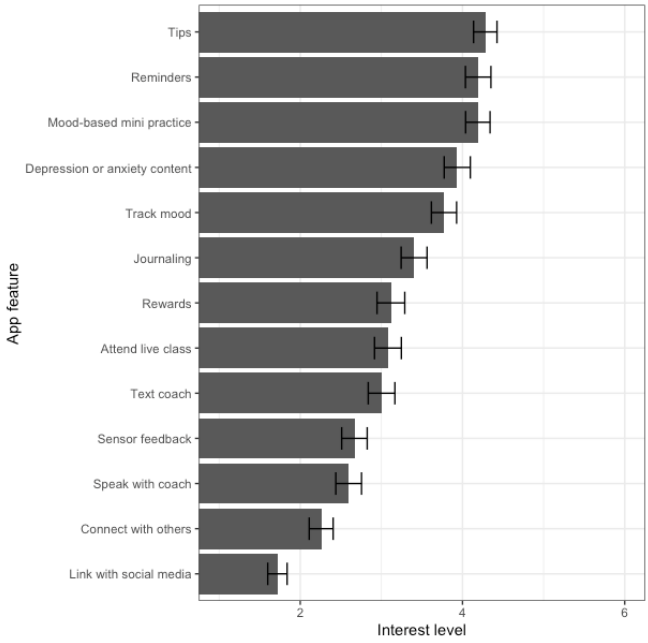


Desired Features of Meditation Apps

Meditators reported the extent to which they desired various features (1=*not at all*; 6=*a great deal*) within meditation apps (Figure 3; Multimedia Appendix 4). Among the list of desired features, the most highly rated included having tips for daily life practice (mean 4.28, SD 1.53), the ability to set reminders to practice (mean 4.19, SD 1.67), encouragement to try "mini" meditation practices based on one's mood (mean 4.19, SD 1.60), providing content related to depression and anxiety (mean 3.93, SD 1.71), and functions to track mood for customized practices (mean 3.77, SD 1.63). Less highly rated features included automated feedback based on phone sensors (mean 2.67, SD 1.67) and the ability to speak with a meditation coach by phone or video (mean 2.60, SD 1.68). The lowest-rated app features were the ability to connect with other users (mean 2.26, SD

1.57) and the ability to link use to social media (eg, Facebook, Twitter, Instagram, and TikTok; mean 1.72, SD 1.27).

Figure 3. Desired features of meditation apps among meditators (n=434). Interest level was rated from 1 (not at all) to 6 (a great deal). The bars indicate the mean ratings with 95% CIs. Connect with others: ability to connect with other users; depression or anxiety content: content related to depression and anxiety; journaling: opportunity to journal or complete reflections about the experience; link with social media: ability to link app use to social media; mood-based mini practice: encouragement to try mini meditation practices based on mood; reminders: ability to set practice reminders; rewards: rewards (eg, trophies) for practicing for a certain number of days; sensor feedback: automated feedback based on sensors in phone (eg, location, SMS text messages, and camera); speak with coach: ability to speak with a meditation coach; text coach: ability to text with a meditation coach; tips: tips for daily life practice; track mood: complete questionnaires that track mood to customize practices.



Discussion

Principal Findings

This study sought to characterize the current status of meditation apps within the broader context of meditation practice using a population-based sampling method. We examined the prevalence of meditation app use, correlates of use, and concerns and desired features of meditation apps among meditators (ie, those with lifetime exposure to meditation). To contextualize our sample, it is worth comparing sample demographics with those of meditators from nationally representative surveys. Data on the demographics of those with past-year exposure to meditation were reported for the 2012 NHIS [47]. Of note, this is not a perfect comparison given the differing definitions of meditation practice (ie, past-year exposure in the NHIS vs lifetime exposure in this sample) as well as the timing of data collection (2012 vs 2020). However, both samples were of similar age (average age of 43.77, SD 15.53 years in this sample vs most participants being aged between 25 and 64 years in the 2012 NHIS), were mostly college educated (247/434, 56.9% in this sample vs 61.19% in the 2012 NHIS), comprised mostly women (237/434, 54.6% in this sample vs 51.35% in the 2012 NHIS), and had a similar proportion of Black individuals (52/434, 11.98% in this sample vs 12.37% in the 2012 NHIS). In contrast, this sample had fewer Latinx individuals (23/434, 5.3% vs 15.40% in the 2012 NHIS), which may have been due to Prolific matching on race but not ethnicity, as noted previously. The largest difference between the samples was in the proportion of individuals who reported practicing meditation, with 49.3% (470/953) of this screening survey sample reporting lifetime exposure to NHIS-defined meditation categories versus 4.1% with past-year exposure in the 2012 NHIS. This difference may be due to Prolific participants differing from the general population in ways not captured by Prolific's representative sampling feature. It may also be attributed to the difference in the time frame of the measures (ie, lifetime vs past-year exposure) or reflect the increase in meditation use over the past decade [48].

A striking finding that emerged was the frequency of app use among meditators. Most meditators (255/434, 58.8%) had used a meditation app at least once in their lives, and approximately 1 in 5 (94/434, 21.7%) used meditation apps weekly or more frequently. Thus, it appears that meditation apps are a major component of meditation practice for meditators in the US. The wide use of meditation apps among meditators in the US mirrors the dramatic rise in the adoption of mobile technology generally (eg, smartphones and social media) [49,50] as well as the prominence of meditation content within mental health apps [21]. We also found that active app users were predominantly women (61/94, 65%) and non-Latinx White (56/94, 60%) individuals. This is consistent with the study by Huberty et al [27], who reported that participating subscribers of the Calm app were also predominantly women (79.94%) and White (81.41%) individuals.

Several participant characteristics were associated with increased likelihood of lifetime exposure to meditation apps and active app use. Younger age, higher anxiety, and an initial mental health motivation were associated with greater likelihood of lifetime meditation app use. It may not be surprising that younger people are more likely to try meditation apps given

that younger people are more likely to own smartphones and use the internet [51]. The finding that those with higher anxiety and mental health motivations are more likely to try meditation apps is consistent with prior work indicating that mental health concerns are associated with meditation use generally [52].

An intriguing finding was a pattern of increased concerns regarding meditation apps among lifetime meditation app users. We saw this pattern both for the total number of concerns and for endorsement of several specific concerns, including cost, time required, technical issues, user-friendliness, and not being targeted for personal goals. This finding was unexpected given the assumption that these concerns would discourage participants from using meditation apps. On the basis of the observed pattern of findings, these concerns are presumably the result rather than the cause of meditation app use, with individuals who have used these tools finding them lacking in these specific ways. This may be a reasonable possibility given that individuals who have used meditation apps may have a more realistic (and concerning) sense of the cost, time required, and user-friendliness of these apps. In contrast, the only concern that was negatively associated with lifetime exposure to meditation apps was a lack of interest in using them.

Patterns differed somewhat when examining correlates of active app use (weekly or daily). Active users were also younger but less likely to be men, have a low income, and be non-Latinx White individuals. The association with race and ethnicity is particularly intriguing given evidence that racial and ethnic minority populations face more barriers to receiving mental health services [53-55] and tend to access mental health care less frequently than non-Latinx White individuals [56]. Higher active use of meditation apps among racial and ethnic minority participants supports the notion that meditation apps, along with other technology-based interventions, may reduce barriers to use (eg, transportation, time, cost, and stigma) [57]. Unlike associations with lifetime app use, the total number of concerns and concerns regarding cost, time, and the apps being targeted to users' concerns were not associated with the likelihood of active app use. Having an interest in apps and concerns regarding usability and technical problems remained associated with active app use. Thus, it appears that cost and time concerns may not be barriers for meditators to use meditation apps actively (ie, on a weekly or daily basis).

Mediators expressed various concerns, with concerns regarding app cost and effectiveness being the most common. The concerns about cost mirror the findings of a previous national survey indicating that cost is one of the most important reasons for not downloading health apps [58]. The fact that many meditators (154/434, 35.5%) had concerns about the effectiveness of apps is also notable, particularly given the growing body of meta-analytic evidence suggesting that meditation apps are indeed effective in reducing common mental health symptoms and improving well-being [22]. The fact that many meditators (and perhaps the public generally) are not aware of this research evidence highlights the need for public health education efforts. Such efforts could take cues from health care organizations such as the Veterans Affairs system that, for over a decade, have highlighted the evidence-based nature of the interventions they promote [59].

In terms of desired features, meditators were interested in lighter-weight features, such as tips and reminders for practice and encouragement to engage in brief "mini" practices. This finding aligns with previous research documenting participants' interest in reminders in health apps [60]. Meditators were also interested in mental health content (depression, anxiety, and mood tracking), which aligned with links between both mental health concerns and motivations and lifetime exposure to meditation apps. In contrast, interest in social components was low. Specifically, meditators were less interested in linking use to social media and connecting with other users and meditation teachers through texting or calling. This finding contradicts prior work demonstrating that social factors such as subjective norms (ie, perception of others' attitudes on meditation practice) predict meditation app use [61]. A potential explanation for this discrepancy may be the difference in relationship closeness. Specifically, participants in this study reported low interest in connecting with individuals with whom they were likely less close (ie, meditation teachers and other app and social media users), whereas participants in the study by Crandall et al [61] reported their significant others' perceptions of meditation practice. It is possible that participants perceive meditation as a private or even solitary health behavior, which limits their interest in linking use to social media and connecting with other users and meditation teachers. At the same time, it is important to acknowledge that even features rated as less desirable in this study may in fact be welcome or helpful to certain meditation app users. For instance, several popular meditation apps include these low-rated components (eg, ability to connect with others in Insight Timer and ability to connect with meditation teachers in Ten Percent Happier) [62,63].

Limitations

Several limitations of this study are noteworthy. First, we conducted a web-based survey, which restricted participants to those with access to the internet. Moreover, our study focused on individuals with lifetime exposure to meditation as defined by the NHIS (ie, mindfulness, mantra, or spiritual meditation), and our sample demographics did not align perfectly with the general US demographics. Therefore, findings may not be generalizable to meditation-naive individuals, those who practice other types of meditation, and the general US population. In addition, the data for app use were collected based on self-report measures, which were susceptible to retrospective and social desirability biases [64]. The results may have been different if app use data had been tracked objectively [65]. Another important limitation was that we assessed a limited number of potential concerns and desired features. It is possible that some important concerns and desired features were not adequately captured. Similarly, we examined only a small set of potential predictors (eg, demographics, anxiety, depression, loneliness, motivation, and concerns about apps) of meditation app use. There are surely many important participant and app-specific characteristics that were not evaluated.

Future Directions

Given the widespread use of meditation apps among meditators, an important future direction is the additional assessment of meditation app use in national surveys that investigate

complementary and integrative health use (eg, NHIS) [66]. Such data could provide a more trustworthy depiction of the role of these technologies than the one this study could provide. A second future direction would involve systematically examining whether the desired features (eg, tips and reminders for practice) are included in publicly available apps (eg, Headspace and Calm). Such a study could empirically assess which features are associated with user ratings of and persistence in using these apps. For example, Huberty et al [27] evaluated the mood check-in feature of the Calm app and found that the use of this feature was associated with higher meditation practice frequency in Calm users, particularly for inactive users. Further observational and experimental studies can be conducted to identify baseline characteristics related to persistence in using meditation apps, as has been done previously [61]. Given that cost is a potentially major barrier to the use of meditation apps, efforts should be made to develop viable cost models for meditation apps. Potential ways of addressing this issue include reimbursement through insurance and increasing awareness of free apps [67]. The intriguing finding that racial and ethnic minority meditators may be more likely to use apps more actively than non-Latinx White participants supports ongoing efforts to develop culturally congruent digital tools [68]. Regarding concerns about the effectiveness of meditation apps, it will be essential to continue conducting rigorous randomized controlled trials to evaluate app effectiveness, with the results of these trials communicated to potential users through public health campaigns. In addition, consumer confidence in digital health technologies may increase when these technologies receive approval from the Food and Drug Administration [69].

Acknowledgments

SL and QX contribute equally to this work as co-first authors and are listed in alphabetical order by last name. This study was supported by the National Center for Complementary and Integrative Health Grant K23AT010879 (SBG), the Hope for Depression Research Foundation Defeating Depression Award (SBG), the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation (SBG), and by funding from the Wisconsin Center for Education Research (SBG).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full and follow-up survey sample demographics.
[DOCX File , 40 KB - mental_v10i1e43565_app1.docx]

Multimedia Appendix 2
Screening survey items.
[DOCX File , 34 KB - mental_v10i1e43565_app2.docx]

Multimedia Appendix 3
Follow-up survey items.
[DOCX File , 40 KB - mental_v10i1e43565_app3.docx]

Multimedia Appendix 4
Descriptive statistics for meditators (ie, lifetime exposure to meditation) who completed the follow-up survey (n=434).
[DOCX File , 45 KB - mental_v10i1e43565_app4.docx]

References

1. Mobile fact sheet. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-10-08]
2. Powell AC, Landman AB, Bates DW. In search of a few good apps. JAMA 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
3. Figueroa CA, Aguilera A. The need for a mental health technology revolution in the COVID-19 pandemic. Front Psychiatry 2020 Jun 3;11:523 [FREE Full text] [doi: 10.3389/fpsyt.2020.00523] [Medline: 32581891]
4. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
5. Dounavi K, Tsoumani O. Mobile health applications in weight management: a systematic literature review. Am J Prev Med 2019 Jun;56(6):894-903 [FREE Full text] [doi: 10.1016/j.amepre.2018.12.005] [Medline: 31003801]

6. Gandhi S, Chen S, Hong L, Sun K, Gong E, Li C, et al. Effect of mobile health interventions on the secondary prevention of cardiovascular disease: systematic review and meta-analysis. Can J Cardiol 2017 Feb;33(2):219-231. [doi: 10.1016/j.cjca.2016.08.017] [Medline: 27956043]
7. Goldberg SB, Lam SU, Simonsson O, Torous J, Sun S. Mobile phone-based interventions for mental health: a systematic meta-review of 14 meta-analyses of randomized controlled trials. PLOS Digit Health 2022;1(1):e0000002 [FREE Full text] [doi: 10.1371/journal.pdig.0000002] [Medline: 35224559]
8. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
9. McCarroll R, Eyles H, Ni Mhurchu C. Effectiveness of mobile health (mHealth) interventions for promoting healthy eating in adults: a systematic review. Prev Med 2017 Dec;105:156-168. [doi: 10.1016/j.ypmed.2017.08.022] [Medline: 28882743]
10. Xie Q, Torous J, Goldberg SB. E-mental health for people with personality disorders: a systematic review. Curr Psychiatry Rep 2022 Oct 16;24(10):541-552 [FREE Full text] [doi: 10.1007/s11920-022-01360-1] [Medline: 35972718]
11. Lutz A, Slagter HA, Dunne JD, Davidson RJ. Attention regulation and monitoring in meditation. Trends Cogn Sci 2008 Apr;12(4):163-169 [FREE Full text] [doi: 10.1016/j.tics.2008.01.005] [Medline: 18329323]
12. 2012 National Health Interview Survey (NHIS) adult complementary and alternative medicine public use file (ALT). National Center for Health Statistics. 2013 Jun 11. URL: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2012/althealt_freq.pdf [accessed 2022-10-08]
13. 2017 National Health Interview Survey (NHIS) Sample Adult Public Use File (samadult). National Center for Health Statistics. 2018 Jun 15. URL: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2017/samadult_freq.pdf [accessed 2022-10-08]
14. Stussman BJ, Bethell CD, Gray C, Nahin RL. Development of the adult and child complementary medicine questionnaires fielded on the National Health Interview Survey. BMC Complement Altern Med 2013 Nov 23;13(1):328 [FREE Full text] [doi: 10.1186/1472-6882-13-328] [Medline: 24267412]
15. Creswell JD, Lindsay EK, Villalba DK, Chin B. Mindfulness training and physical health: mechanisms and outcomes. Psychosom Med 2019 Apr;81(3):224-232 [FREE Full text] [doi: 10.1097/PSY.0000000000000675] [Medline: 30806634]
16. Dunn TJ, Dimolareva M. The effect of mindfulness-based interventions on immunity-related biomarkers: a comprehensive meta-analysis of randomised controlled trials. Clin Psychol Rev 2022 Mar;92:102124. [doi: 10.1016/j.cpr.2022.102124] [Medline: 35078038]
17. Goldberg SB, Riordan KM, Sun S, Davidson RJ. The empirical status of mindfulness-based interventions: a systematic review of 44 meta-analyses of randomized controlled trials. Perspect Psychol Sci 2022 Jan 16;17(1):108-130 [FREE Full text] [doi: 10.1177/1745691620968771] [Medline: 33593124]
18. Gonçalves JPB, Lucchetti G, Menezes PR, Vallada H. Religious and spiritual interventions in mental health care: a systematic review and meta-analysis of randomized controlled clinical trials. Psychol Med 2015 Jul 23;45(14):2937-2949. [doi: 10.1017/s0033291715001166]
19. Lynch J, Prihodova L, Dunne PJ, Carroll Á, Walsh C, McMahon G, et al. Mantra meditation for mental health in the general population: a systematic review. Eur J Integrative Med 2018 Oct;23:101-108. [doi: 10.1016/j.eujim.2018.09.010]
20. Xie Q, Guan Y, Hofmann SG, Jiang T, Liu X. The potential mediating role of anxiety sensitivity in the impact of mindfulness training on anxiety and depression severity and impairment: a randomized controlled trial. Scand J Psychol 2023 Feb 19;64(1):21-29. [doi: 10.1111/sjop.12860] [Medline: 35851956]
21. Wasil AR, Gillespie S, Patel R, Petre A, Venturo-Conerly KE, Shingleton RM, et al. Reassessing evidence-based content in popular smartphone apps for depression and anxiety: developing and applying user-adjusted analyses. J Consult Clin Psychol 2020 Nov;88(11):983-993. [doi: 10.1037/ccp0000604] [Medline: 32881542]
22. Gál É, tefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Jan 15;279:131-142. [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
23. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
24. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
25. Torous J, Wisniewski H, Liu G, Keshavan M. Mental health mobile phone app usage, concerns, and benefits among psychiatric outpatients: comparative survey study. JMIR Ment Health 2018 Nov 16;5(4):e11715 [FREE Full text] [doi: 10.2196/11715] [Medline: 30446484]
26. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]

27. Huberty J, Vranceanu A, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019 Nov 03;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]
28. Owings-Fonner N. Wellness apps designed for people of color. American Psychological Association. 2022 Jun. URL: https://www.apaservices.org/practice/business/technology/tech-column/wellness-apps [accessed 2023-03-09]
29. Galante J, Galante I, Bekkers M, Gallacher J. Effect of kindness-based meditation on health and well-being: a systematic review and meta-analysis. J Consult Clin Psychol 2014 Dec;82(6):1101-1114. [doi: 10.1037/a0037249] [Medline: 24979314]
30. Stanton K, Carpenter RW, Nance M, Sturgeon T, Villalongo Andino M. A multisample demonstration of using the prolific platform for repeated assessment and psychometric substance use research. Exp Clin Psychopharmacol 2022 Aug;30(4):432-443. [doi: 10.1037/pha0000545] [Medline: 35025585]
31. Palan S, Schitter C. Prolific.ac—a subject pool for online experiments. J Behavioral Experimental Finance 2018 Mar;17:22-27. [doi: 10.1016/j.jbef.2017.12.004]
32. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. J Experimental Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
33. Quick facts. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/US/PST045221 [accessed 2022-10-08]
34. Cohen J, Cohen P, West S, Aiken L. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd edition). Milton Park, Abingdon, Oxfordshire: Routledge; 2013.
35. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019 Jun 17;2(1):54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
36. Goldberg SB, Lam SU, Britton WB, Davidson RJ. Prevalence of meditation-related adverse effects in a population-based sample in the United States. Psychother Res 2022 Mar 02;32(3):291-305 [FREE Full text] [doi: 10.1080/10503307.2021.1933646] [Medline: 34074221]
37. Schlosser M, Sparby T, Vörös S, Jones R, Marchant NL. Unpleasant meditation-related experiences in regular meditators: prevalence, predictors, and conceptual considerations. PLoS One 2019;14(5):e0216643 [FREE Full text] [doi: 10.1371/journal.pone.0216643] [Medline: 31071152]
38. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, PROMIS Cooperative Group. Item banks for measuring emotional distress from the patient-reported outcomes measurement information system (PROMIS®): depression, anxiety, and anger. Assessment 2011 Sep;18(3):263-283 [FREE Full text] [doi: 10.1177/1073191111411667] [Medline: 21697139]
39. Cyranowski JM, Zill N, Bode R, Butt Z, Kelly MA, Pilkonis PA, et al. Assessing social support, companionship, and distress: National Institute of Health (NIH) Toolbox Adult Social Relationship Scales. Health Psychol 2013 Mar;32(3):293-301 [FREE Full text] [doi: 10.1037/a0028586] [Medline: 23437856]
40. Pepping CA, Walters B, Davis PJ, O'Donovan A. Why do people practice mindfulness? An investigation into reasons for practicing mindfulness meditation. Mindfulness 2016 Feb 4;7(2):542-547. [doi: 10.1007/s12671-016-0490-3]
41. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
42. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]
43. Torous J, Kiang MV, Lorme J, Onnela J. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
44. Kenny R, Dooley B, Fitzgerald A. Feasibility of "CopeSmart": a telemental health app for adolescents. JMIR Ment Health 2015 Aug 10;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
45. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. J Educational Behavioral Stat 2016 Nov 23;27(1):77-83. [doi: 10.3102/10769986027001077]
46. Curran PJ, West SG, Finch JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychological Method 1996 Mar;1(1):16-29. [doi: 10.1037/1082-989x.1.1.16]
47. Burke A, Lam CN, Stussman B, Yang H. Prevalence and patterns of use of mantra, mindfulness and spiritual meditation among adults in the United States. BMC Complement Altern Med 2017 Jun 15;17(1):316 [FREE Full text] [doi: 10.1186/s12906-017-1827-8] [Medline: 28619092]
48. Lam SU, Riordan KM, Simonsson O, Davidson RJ, Goldberg SB. Who sticks with meditation? Rates and predictors of persistence in a population-based sample in the USA. Mindfulness 2023 Jan 06;14(1):66-78. [doi: 10.1007/s12671-022-02061-9] [Medline: 36777474]
49. Perrin A. Social media usage: 2005-2015. Pew Research Center. 2015 Oct 8. URL: https://www.pewresearch.org/internet/2015/10/08/social-networking-usage-2005-2015/ [accessed 2022-10-08]

50. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. Pew Research Center. 2016 Feb 22. URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2022-10-08]
51. Silver L. 2. In emerging economies, smartphone adoption has grown more quickly among younger generations. Pew Research Center. 2019 Feb 5. URL: https://www.pewresearch.org/global/2019/02/05/in-emerging-economies-smartphone-adoption-has-grown-more-quickly-among-younger-generations/ [accessed 2022-10-08]
52. Cramer H, Hall H, Leach M, Frawley J, Zhang Y, Leung B, et al. Prevalence, patterns, and predictors of meditation use among US adults: a nationally representative survey. Sci Rep 2016 Nov 10;6(1):36760 [FREE Full text] [doi: 10.1038/srep36760] [Medline: 27829670]
53. Hines-Martin V, Malone M, Kim S, Brown-Piper A. Barriers to mental health care access in an African American population. Issues Ment Health Nurs 2003 Jul 09;24(3):237-256. [doi: 10.1080/01612840305281] [Medline: 12623684]
54. Miranda R, Soffer A, Polanco-Roman L, Wheeler A, Moore A. Mental health treatment barriers among racial/ethnic minority versus white young adults 6 months after intake at a college counseling center. J Am Coll Health 2015 Feb 18;63(5):291-298. [doi: 10.1080/07448481.2015.1015024] [Medline: 25692661]
55. Sorkin DH, Murphy M, Nguyen H, Biegler KA. Barriers to mental health care for an ethnically and racially diverse sample of older adults. J Am Geriatr Soc 2016 Oct 26;64(10):2138-2143 [FREE Full text] [doi: 10.1111/jgs.14420] [Medline: 27565017]
56. Cook BL, Trinh N, Li Z, Hou SS, Progovac AM. Trends in racial-ethnic disparities in access to mental health care, 2004-2012. Psychiatr Serv 2017 Jan 01;68(1):9-16 [FREE Full text] [doi: 10.1176/appi.ps.201500453] [Medline: 27476805]
57. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
58. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
59. Karlin BE, Cross G. From the laboratory to the therapy room: national dissemination and implementation of evidence-based psychotherapies in the U.S. Department of Veterans Affairs Health Care System. Am Psychol 2014 Jan;69(1):19-33. [doi: 10.1037/a0033888] [Medline: 24001035]
60. Thornton LK, Kay-Lambkin FJ. Specific features of current and emerging mobile health apps: user views among people with and without mental health problems. Mhealth 2018 Dec;4:56 [FREE Full text] [doi: 10.21037/mhealth.2018.11.04] [Medline: 30701174]
61. Crandall A, Cheung A, Young A, Hooper AP. Theory-based predictors of mindfulness meditation mobile app usage: a survey and cohort study. JMIR Mhealth Uhealth 2019 Mar 22;7(3):e10794 [FREE Full text] [doi: 10.2196/10794] [Medline: 30900992]
62. Best NI, Durham CF, Woods-Giscombe C, Waldrop J. Combating compassion fatigue with mindfulness practice in military nurse practitioners. J Nurse Practitioner 2020 May;16(5):e57-e60. [doi: 10.1016/j.nurpra.2020.02.023]
63. Truhlar L, Durand C, Cooper M, Goldsmith C. Exploring the effects of a smartphone-based meditation app on stress, mindfulness, well-being, and resilience in pharmacy students. Am J Health Syst Pharm 2022 Nov 22;79(23):2159-2165. [doi: 10.1093/ajhp/zxac240] [Medline: 36070430]
64. Schwarz N. Self-reports: how the questions shape the answers. Am Psychologist 1999 Feb;54(2):93-105. [doi: 10.1037/0003-066X.54.2.93]
65. Parry DA, Davidson BI, Sewall CJ, Fisher JT, Mieczkowski H, Quintana DS. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. Nat Hum Behav 2021 Nov;5(11):1535-1547. [doi: 10.1038/s41562-021-01117-5] [Medline: 34002052]
66. Clarke TC, Barnes PM, Black LI, Stussman BJ, Nahin RL. Use of yoga, meditation, and chiropractors among U.S. adults aged 18 and over. NCHS Data Brief 2018 Nov(325):1-8 [FREE Full text] [Medline: 30475686]
67. Raber I, McCarthy CP, Yeh RW. Health insurance and mobile health devices: opportunities and concerns. JAMA 2019 May 14;321(18):1767-1768. [doi: 10.1001/jama.2019.3353] [Medline: 30973573]
68. Spanhel K, Balci S, Feldhahn F, Bengel J, Baumeister H, Sander LB. Cultural adaptation of internet- and mobile-based interventions for mental disorders: a systematic review. NPJ Digit Med 2021 Aug 25;4(1):128 [FREE Full text] [doi: 10.1038/s41746-021-00498-1] [Medline: 34433875]
69. Shuren J, Patel B, Gottlieb S. FDA regulation of mobile medical apps. JAMA 2018 Jul 24;320(4):337-338. [doi: 10.1001/jama.2018.8832] [Medline: 29971339]

Abbreviations

FDR: false discovery rate
NHIS: National Health Interview Survey

Edited by J Torous; submitted 15.10.22; peer-reviewed by JD Creswel, T Ranjan; comments to author 07.02.23; revised version received 11.03.23; accepted 13.03.23; published 28.04.23.

Please cite as:
Lam SU, Xie Q, Goldberg SB
Situating Meditation Apps Within the Ecosystem of Meditation Practice: Population-Based Survey Study
JMIR Ment Health 2023;10:e43565
URL: https://mental.jmir.org/2023/1/e43565
doi:10.2196/43565
PMID:37115618

Original Paper

Effects of Induced Mindfulness at Night on Repetitive Negative Thinking: Ecological Momentary Assessment Study

Amanda Sommerhoff[1], MSc; Thomas Ehring[1], PhD; Keisuke Takano[2], PhD

[1]Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

[2]Human Informatics and Interaction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

Corresponding Author:
Keisuke Takano, PhD
Human Informatics and Interaction Research Institute
National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Higashi
Tsukuba, Ibaraki, 305-8566
Japan
Phone: 81 29 849 1456
Email: keisuke.takano@aist.go.jp

Abstract

Background: Repetitive negative thinking (RNT) is a cognitive risk factor for various disorders. Although brief mindfulness-based interventions (MBIs; lasting 20-30 minutes or shorter) are effective tools to reduce RNT, the effect of a minimal (5-minute) MBI remains largely unknown.

Objective: We investigated the acute changes in RNT induced by a 10-day minimal MBI (body scan before sleeping) using an ecological momentary assessment (EMA) administered during the MBI training phase. In addition, we examined longer-term effects on the postintervention and 2-month follow-up assessments for questionnaire-based RNT and psychological distress.

Methods: A total of 68 participants (community sample, aged 18-55 years; n=58, 85% women) were randomly allocated to either the intervention group (n=35, 51%) or the no-training control group (n=33, 49%). Both groups completed a 10-day EMA phase of RNT, during which only the intervention group performed a daily 5-minute body scan before sleeping.

Results: The intervention group showed a significantly larger reduction in questionnaire-based RNT than the control group at the follow-up assessment (for growth-curve modeling analysis [GMA], d_{GMA}=−0.91; P<.001), but this effect was not observed during the EMA phase or at the postintervention assessment. Furthermore, the intervention group showed significantly larger decreases in stress both at the postintervention (d_{GMA}=−0.78; P<.001) and follow-up (d_{GMA}=−0.60; P<.001) assessments than the control group. We found no intervention effects on depressive and anxiety symptoms.

Conclusions: A 5-minute body scan before sleeping reduces RNT and stress when continued for at least 10 days; however, the results suggest that this effect only appears with some time lag because no acute changes during and immediately after the intervention emerged for RNT.

(JMIR Ment Health 2023;10:e44365) doi:10.2196/44365

KEYWORDS

mindfulness; repetitive negative thinking; stress; daily life; ecological momentary assessment; mobile phone

Introduction

Background

It is estimated that approximately one-third of the global population is affected by mental disorders [1], which are associated with high direct and indirect costs for the individual and society [2,3]. In addition, subclinical levels of psychopathology lead to considerable dysfunction and impairment [4]. However, there is limited access to mental health services provided by mental health professionals [4], necessitating a self-management strategy to maintain and improve mental health and prevent the onset of mental disorders [5].

As a promising treatment approach for a variety of mental health outcomes, mindfulness-based interventions (MBIs) have increasingly captured the interest of the scientific community

over the past 3 decades [6]. Standardized MBIs such as mindfulness-based stress reduction (MBSR) [7,8] and mindfulness-based cognitive therapy (MBCT) [9] have been established as effective prevention and treatment approaches for various mental health outcomes in clinical [10,11] and nonclinical populations [12-14]. Importantly, MBIs are shown to be effective not only in improving symptoms but also in reducing vulnerability factors such as depressive rumination and repetitive negative thinking (RNT) [15,16].

However, programs such as MBSR and MBCT require substantial time investment from practitioners [12]. Hence, brief MBIs have been developed to enhance the accessibility of MBIs.

Although brief MBIs are suitable for addressing the accessibility issue, a consensus has not yet been reached concerning how much exercise should be included in a single session and for how many days the intervention should be continued, which poses an obstacle for the widespread implementation of MBIs as a mental health tool [17,18]. A systematic review [19] found that even a minimal set of brief MBIs (ie, 5-20 minutes in duration, mostly mindfulness audios) in 1 session can have a positive effect on negative mood, anxiety, or associated factors such as rumination in healthy and clinical populations.

However, the longevity of the effects of brief MBIs is unclear [19,20], and only a few studies have examined the real-time effects of MBIs using ecological momentary assessment (EMA) designs [21-23]. In an EMA study, participants are typically asked to provide systematic self-reports of their everyday lives repeatedly during their waking hours [24]. This web-based ecologically valid assessment approach improves temporal resolution, allowing researchers to capture even brief changes in symptoms. A study administering EMA during MBSR [22] suggested that the intervention improved mindfulness skills, depression, and anxiety symptoms throughout the intervention. Interestingly, these effects emerged only in the EMA-assessed symptoms but not in the symptoms reported in the pre- and postintervention phases. Another study [23] conducted EMA during an 8-week MBI program in which participants were instructed to choose from meditations varying in length from 5 to 20 minutes. The results indicated that the effects of the MBI appeared immediately, highlighting a pronounced decline in anxiety and sleep problems and an acute increase in happiness within the first weeks of the intervention program [23]. Furthermore, the total number of minutes meditated was not a statistically significant predictor of outcomes at the 8-week assessment point [23]. There is also an EMA study [25] that targeted changes in a cognitive dysfunction (rumination) over 42 days of an MBI, which consisted of 20-minute daily meditation practice sessions. The changes in rumination seemed to follow a 3-phase pattern: during the first week of the MBI, rumination decreased rapidly; between days 10 and 30, it remained stable; and after 30 days of practice, it decreased again [25].

To expand these earlier EMA-mindfulness works, this study used EMA to examine the effect of a minimal (ie, 5-minute) MBI on a marked cognitive vulnerability factor for psychopathology, that is, RNT. We targeted RNT as the main outcome here because the theory underlying MBCT [9] highlights the role of this maladaptive thinking style as a putative mechanism of change for an MBI; specifically, it has been proposed that MBIs enhance awareness of, and disengagement from, RNT, which reduces symptoms and prevents a relapse of depression [9]. RNT is defined as excessive and repetitive thinking about one's current concerns, problems, past experiences, or worries about the future [26]—conceptually encompassing depressive rumination and anxious worry. RNT is an important risk and maintenance factor for various mental disorders, including depression and anxiety [26], and has been a target of prevention and intervention [27-29]. Although there is no consensus concerning the mechanisms underlying an MBI to date [30], a growing body of literature suggests that alterations in RNT are a significant contributing factor to the effects of an MBI [31,32].

Although RNT is typically operationalized as a ruminative tendency (or trait factor), it has been shown that RNT has some temporal variability that may hint at how RNT-focused interventions can be implemented. An EMA study [33] found that RNT levels have a U-shaped diurnal variation—being the highest in the morning and evening. This pattern was replicated by another EMA study, which also identified individual differences owing to depressive symptoms, that is, individuals with higher levels of depressive symptoms tend to experience higher levels of ruminative thinking in the evening [34]. Evening rumination seems to be triggered by a lack of social distractors, such as interaction with others [35], and is associated with increased autonomic arousal [36] and poor sleep quality [37-39]. Therefore, the evening is a critical time window to effectively implement psychological interventions targeting RNT.

Objectives

In this study, participants performed a 5-minute body scan before sleeping, which we expected to prevent the occurrence of, and help disengage from, RNT. As supporting evidence for an MBI at night, it has been suggested that a 6-minute mindfulness induction before sleeping improves athletes' sleep quality (although evidence is lacking for RNT) [40]. Therefore, we aimed to provide direct evidence of the effectiveness of a minimal MBI before sleeping to reduce RNT. To establish this effect, we used a 2×3 factorial design with 2 groups (MBI group vs no-training control group) and 3 assessment points (baseline, after the intervention, and 2-month follow-up), as well as continuous EMA to capture the acute in-training changes in RNT throughout the MBI. Our primary outcome was RNT, and we also assessed psychological distress symptoms (ie, depressive, anxiety, and stress symptoms) as secondary outcomes at each assessment. First, we hypothesized that the intervention group would show a significant decrease in EMA-assessed RNT as the MBI progressed, whereas the control group would maintain their initial levels of RNT over time (hypothesis 1). Second, we hypothesized that the intervention group would show a significant reduction in questionnaire-assessed RNT from baseline to the postintervention and follow-up assessments (hypothesis 2).

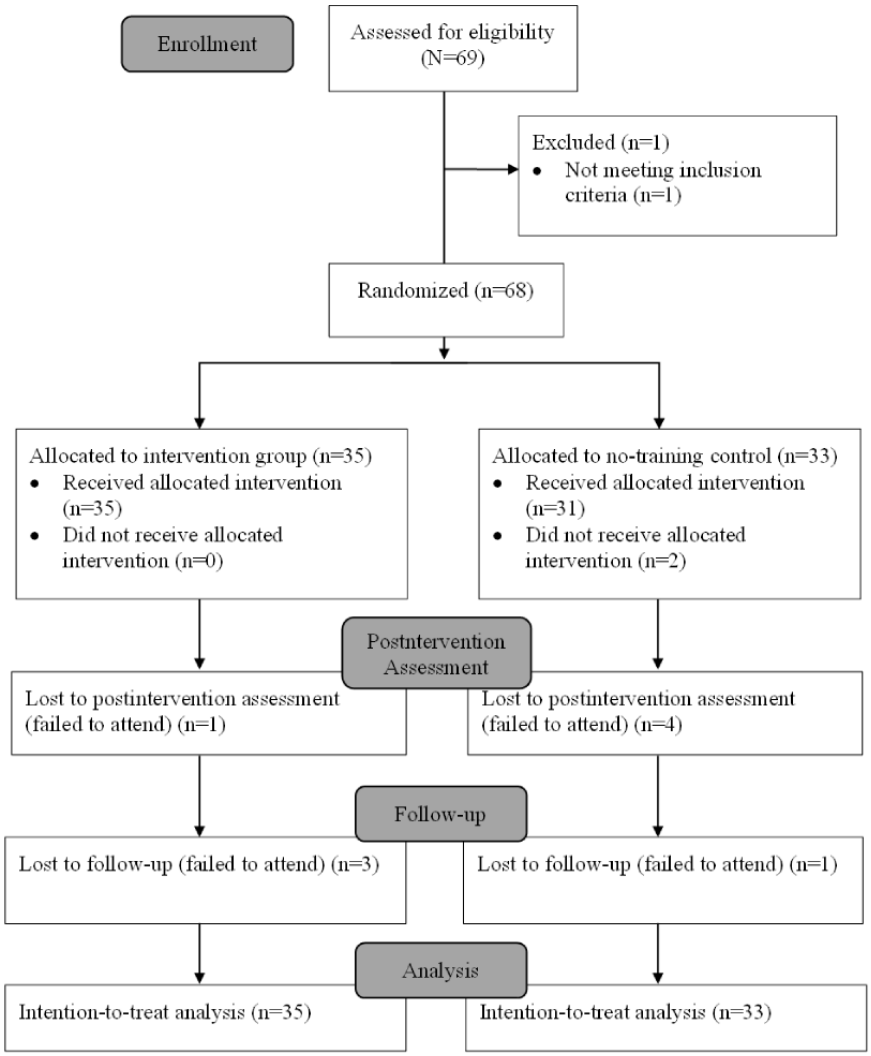
Methods

Participants

Participants were recruited at the university campus of Ludwig-Maximilians-Universität München (LMU Munich) and from the Munich metropolitan area on the web (Facebook and eBay Kleinanzeigen [advertisements in the personal columns]), as well as via flyers and by using the university's email distribution list (covering graduate and undergraduate students). Participants were required to be aged at least 18 years and fluent in German and were excluded if they had practiced mindfulness daily in the past 3 months before the study [41] or had a neurological disorder or a serious disorder for which mindfulness practice may have adverse effects (cardiac function disorders, hypotension, bronchial asthma, migraines, seizure disorders such as epilepsy, attention-deficit/hyperactivity disorder, and dementia). Each participant was informed of the purpose of the study at the recruitment as follows: "The aim of this study is to investigate the effect of a short mindfulness exercise before bedtime on mood, sleep quality, and cognitions. You will be allocated to one of two groups. One group performs a short mindfulness exercise every day whereas the other group does not change anything in their everyday life." Eligible participants (N=69) were then assessed for the inclusion criteria (Figure 1). Of these 69 participants, 1 (1%) was excluded because of an asthma condition. The remaining participants were randomly allocated to either the intervention group (35/68, 51%) or the control group (33/68, 49%) following a predetermined random allocation sequence generated by Random [42].

An a priori power analysis was conducted using G*Power [43] to determine the sample size. Querstret et al [14] reported an effect of Hedges g=−1.13 in the postintervention levels of rumination in nonclinical samples, which gives the required sample size as 28, with the assumptions of α=.05 and power=0.80.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



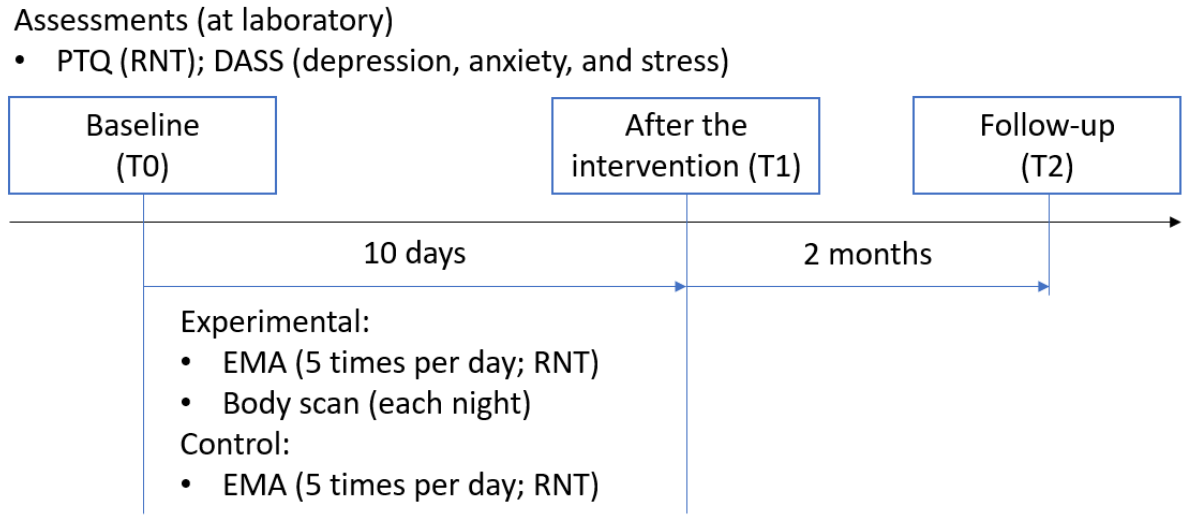
Procedure

Participants attended the preintervention assessment (T0) at the laboratory of LMU Munich, where they provided written informed consent and completed questionnaires for demographics and baseline symptomatology. Note that data were collected within a larger study (Takano, K, unpublished data, October 2022) using a test battery, including other measures, that are not reported here. Participants were randomly allocated to either the intervention or the no-training control

group following the randomization sequence that had been created before data collection. No blinding procedures were performed. Participants were instructed to use the EMA app (MobileQ; KU Leuven), which they had installed on their Android smartphone before the appointment (in case of technical issues or if they did not own an Android smartphone, the participants received a smartphone for the duration of the study). Upon complying with this instruction, participants read through all EMA questions, and any unclear points regarding the study procedure were resolved by an experimenter. The intervention group received further instructions for the 5-minute MBI. The day after the T0 assessment, all participants started the 10-day EMA and, if they were allocated to the intervention group, the MBI training sessions. Participants completed the same questionnaires as the ones administered at T0 on a web-based platform the day after the completion of the 10-day EMA (T1, after the intervention) and 2 months later (T2, follow-up). Furthermore, the T2 assessment included 1 question asking whether participants had performed mindfulness practice during the period between the postintervention and follow-up assessments (refer to Figure 2 for the flow of the study).

After completion of the T2 assessment, the participants were debriefed and provided with either course credit or monetary compensation. For monetary compensation, participants received €8 (US $8.8) per hour for assessments and had the chance to win 1 of 10 €20 (US $22) vouchers, depending on their compliance during the EMA. For student participants, course credit was given based on compliance during the EMA. In addition, all participants who completed both T1 and T2 assessments had the opportunity to win a €50 (US $55) voucher. The data collection started in November 2019 and ended in May 2020, and we did not change any aspect of the study protocol even after the COVID-19 pandemic.

Figure 2. Flow of the study. DASS: Depression Anxiety Stress Scale; EMA: ecological momentary assessment; PTQ: Perseverative Thinking Questionnaire; RNT: repetitive negative thinking.



Ethics Approval

The ethics committee of the Department of Psychology at LMU Munich approved the study protocol (approval ID: 11_2019_Takano_b).

MBI Practice

Participants in the intervention group performed a 5-minute guided version of the body scan derived from MBSR [7] as an MBI each night throughout the EMA. Participants were instructed to practice the body scan before going to sleep as their final activity of the day and to mindfully guide their attention from their feet to the head in a nonjudgmental manner. It has been shown that guided exercises support participants in practicing mindfulness [44], and the body scan is considered a good starting point for mindfulness practice [45]. In addition, it is one of the most popular meditation exercises among novices [46]. Participants in the control group were instructed to go about their day as usual.

Measures

RNT Assessment

RNT was assessed using the 15-item Perseverative Thinking Questionnaire (PTQ) [47]. Three subscales assess RNT as a transdiagnostic process, with 9 items covering the three core characteristics of RNT: (1) repetitiveness, (2) intrusiveness, and (3) difficulties in disengaging from thoughts; for each core characteristic, 3 questions are asked concerning the unproductiveness and mental capacity captured by RNT. Items were rated on a scale ranging from 0=*never* to 4=*almost always*. The total sum score for all 15 items was computed. The PTQ is a valid measurement instrument for RNT [47]. In this study, internal consistency was excellent, with Cronbach α=.93 for the sum score and internal consistencies of Cronbach α=.92 (core characteristics), Cronbach α=.82 (unproductiveness), and Cronbach α=.84 (mental capacity) for each respective subscale.

Psychological Distress Symptoms

Depressive, anxiety, and stress symptoms were measured using the German short form [48] of the Depression Anxiety Stress Scale (DASS) [49] for nonclinical populations. Participants were asked to rate 21 items within the time frame of the previous

week on a scale ranging from 0=*did not apply to me at all* to 3=*applied to me very much or most of the time*. The total score for each subscale was computed (depression, anxiety, and stress: 7 items each). The DASS showed good internal consistencies: Cronbach α=.83 (depression), Cronbach α=.77 (anxiety), and Cronbach α=.83 (stress).

EMA Method

The EMA was conducted using a smartphone app, MobileQ [50]. Participants received 5 prompts daily for 10 consecutive days. In response to each prompt, participants had to start a short survey within 90 seconds. At each prompt, participants were asked to rate 4 items adapted from the PTQ [51] on a scale ranging from 0=*not at all* to 7=*very much* concerning their current RNT: that is, "The same thoughts keep going through my mind again and again" (repetitiveness), "Thoughts come to my mind without me wanting them to" (intrusiveness), "I get stuck on certain issues and can't move on" (uncontrollability), and "I feel disturbed by negative thoughts" (distress). The distress item was adapted from the original wording [51] to fit in the context of this study. The mean score of all 4 items was computed. In addition, the first prompt of each day included an item intended as an adherence check for the body scan, where participants were asked to indicate whether they had practiced the 5-minute body scan the previous night (compare with the study by Hülsheger et al [52]). A duration of 10 days was chosen because RNT has been shown to rapidly decrease during the first week of mindfulness practice [25]. The number of daily assessments was based on general guidelines [53]. Furthermore, a sampling design of 5 daily assessments across 10 days is recommended [51] because this sampling frequency has been shown to yield the best trade-off between participant burden and information obtained by EMA. Each EMA included items other than RNT, such as momentary levels of affect (rated on a slider ranging from 0=*unpleasant* to 100=*pleasant*). We found no significant effect of the MBI on the affect (valence) ratings, and thus we focused exclusively on RNT in the *Results* section. The first EMA prompt of each day was sent at 9 AM, and the remaining 4 prompts were scheduled at random times between 10 AM and 10 PM, each prompt separated by an average of 2 hours. The estimated time to answer all questions was 2 minutes.

Data Analyses

Multilevel modeling was performed on the intention-to-treat sample with (restricted) maximum likelihood (ML) estimation implemented by the R packages *lme4* [54] and *lmerTest* [55]. This analytic approach has some advantages in modeling the random effect structure and handling missing data. For hypothesis 1 (RNT would decrease over the 10-day MBI training phase), a multilevel model was estimated on the RNT, as assessed by the EMA. We assumed a 2-level nested structure, with assessment-level variables nested within the person-level variables. The model is formulated as follows:

Prompt level:

$$RNT_{ij} = \beta_{0j} + \beta_{1j}Time_{ij} + r_{ij} \quad \textbf{(1)}$$

Person level:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Group_j + u_{0j} \quad \textbf{(2)}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Group_j + u_{1j} \quad \textbf{(3)}$$

RNT_{ij} is the momentary level of RNT for participant i for the prompt j of the EMA. *Group* is a dummy code indicating group allocation, with 0 for the control group and 1 for the intervention group. The residual is denoted by r. The intercept and slopes (β_{0j} and β_{1j}) were allowed to vary across participants with person-level random effects (u_{0j} and u_{1j}), and the cross-level interaction effect (γ_{11}) represents the group differences in the (linear) change over the MBI training phase. In other words, we tested whether the rate of change in RNT (β_{1j}) would differ between the 2 groups, and this difference was modeled by the interaction effect γ_{11}. We performed model selection using information criteria (with the ML estimator) to see whether a polynomial function would better approximate the changing pattern of RNT than a linear function of time. However, we found that the linear function was better than the quadratic and cubic functions, and thus we decided to focus exclusively on the linear change in the *Results* section.

To test the effects of our MBI at the T1 and T2 assessments (ie, hypothesis 2), we estimated a similar 2-level model for each of the 4 outcome measures (but with the restricted ML estimator): the PTQ and DASS depression, anxiety, and stress subscales. Each model is formulated as follows:

Assessment level:

$$Outcome_{ij} = \beta_{0j} + \beta_{1j}DT1_{ij} + \beta_{2j}DT2_{ij} + r_{ij} \quad \textbf{(4)}$$

Person level:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Group_j + u_{0j} \quad \textbf{(5)}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Group_j + u_{1j} \quad \textbf{(6)}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}Group_j + u_{2j} \quad \textbf{(7)}$$

$Outcome_{ij}$ is the outcome value of participant i at assessment j. The time variables *DT*1 and *DT*2 are dummy codes representing the differences between the T0 and T1 assessments and between the T0 and T2 assessments, respectively. Both the intercept and slopes (β_{0j}, β_{1j}, and β_{2j}) were allowed to vary across participants with person-level random effects (u_{0j}, u_{1j}, and u_{2j}, respectively). The variance of random effects was fixed at 0 if the estimates exceeded 0. The 2 interaction effects γ_{11} and γ_{21} were of particular interest because they represented group differences in the changes between the 2 given assessments.

Post hoc (simple slope) tests were conducted for any significant interaction effects between time and group. Effect sizes were defined in the framework of the growth-curve modeling analysis (GMA) [56], that is, d_{GMA} was calculated by first multiplying the fixed effect by time (ie, the average change in a given time window) and then dividing the product by the raw SD of the outcome. This metric is comparable with Cohen *d*. For each analysis, we tested whether the results were unchanged even after controlling for covariates (ie, age, gender, and spontaneous engagement in mindfulness practice up to the follow-up assessment).

Results

Descriptive Data and Compliance

The final sample consisted of 68 participants aged 18 to 55 (mean 26.79, SD 9.41) years (n=58, 85% women). The descriptive statistics for the T0, T1, and T2 assessments are presented in Tables 1 and 2. The mean compliance among the 2 groups during the EMA was 76.73% (SD 16%). Participants in the intervention group (35/68, 51%) had a slightly higher mean compliance (78%, SD 16%) than participants in the control group (31/68, 49%; 75%, SD 16%). Of the 31 participants in the control group, 2 (6%) had no available EMA data (n=1, 50% did not start the EMA, and n=1, 50% did not respond to any EMA signals). Participants showed excellent adherence to the intervention; they performed our 5-minute body scan for a mean 86.98% (SD 11%) of the days during the MBI training phase. Only 1 (3%) of the 35 participants performed <70% of the MBI training sessions. At the T2 assessment, of the 31 participants in the intervention group, 12 (39%) reported that they had stopped mindfulness practice after the 10-day training period, 18 (58%) had continued the practice at least once a month, and 1 (3%) had practiced mindfulness for most of the days. Note that the self-guided continuation of the MBI exercises was a spontaneous reaction because we did not instruct the participants to continue the practice. In the control group, of the 28 participants, 11 (39%) reported having engaged in spontaneous mindfulness practice after the T1 assessment, and 2 (7%) reported having practiced mindfulness on most of the days.

Table 1. Demographics and baseline characteristics (n=68).

Variable	Control group (n=33)	Intervention group (n=35)	Group difference		
			t test (*df*)	Chi-square (*df*)	*P* value
Age (years), mean (SD)	27.76 (10.69)	25.89 (8.07)	0.81 (66)	N/A[a]	.42
Gender, n (%)					
Man	4 (12)	6 (17)	N/A	0.1 (1)	.81
Woman	29 (88)	29 (83)	N/A	0.1 (1)	.81
EMA[b] compliance (%), mean (SD)	78 (16)[c]	75 (16)	0.59 (64)	N/A	.55
T0 (baseline) measures, mean (SD)					
PTQ[d]	31.48 (9.27)	28.97 (10.45)	1.05 (66)	N/A	.30
DASS-D[e]	3.94 (3.57)	3.81 (3.96)	0.14 (63)	N/A	.89
DASS-A[f]	3.48 (3.31)	2.69 (2.97)	1.05 (66)	N/A	.30
DASS-S[g]	5.94 (4.04)	6.71 (3.91)	−0.78 (64)	N/A	.44

[a]N/A: not applicable.

[b]EMA: ecological momentary assessment.

[c]Of the 33 participants in the control group, 2 (6%) had no EMA data.

[d]PTQ: Perseverative Thinking Questionnaire.

[e]DASS-D: Depression Anxiety Stress Scale, depression subscale.

[f]DASS-A: Depression Anxiety Stress Scale, anxiety subscale.

[g]DASS-S: Depression Anxiety Stress Scale, stress subscale.

Table 2. Descriptives at the postintervention and follow-up assessments.

Variables	Postintervention assessment, mean (SD)		Follow-up assessment, mean (SD)	
	Control group (n=29)	Intervention group (n=34)	Control group (n=28)	Intervention group (n=31)
PTQ[a]	27.55 (11.33)	22.68 (8.79)	28.00 (9.66)	20.32 (8.56)
DASS-D[b]	3.19 (3.64)	2.52 (3.13)	3.39 (3.02)	2.65 (3.12)
DASS-A[c]	2.39 (2.77)	1.33 (2.03)	2.44 (2.55)	1.58 (2.20)
DASS-S[d]	5.59 (5.00)	3.88 (2.78)	5.22 (3.47)	4.35 (3.15)

[a]PTQ: Perseverative Thinking Questionnaire.

[b]DASS-D: Depression Anxiety Stress Scale, depression subscale.

[c]DASS-A: Depression Anxiety Stress Scale, anxiety subscale.

[d]DASS-S: Depression Anxiety Stress Scale, stress subscale.

Hypothesis 1: EMA-Assessed RNT Throughout the MBI Training Phase

To test the temporal changes in RNT throughout the MBI, a multilevel model was estimated with *group*, *time* (EMA prompt), and their interaction as predictors of EMA-assessed RNT (Table 3). The results showed no significant main or interaction effects, suggesting that the level of RNT did not change over time and that there was no significant group difference in the rate of change (Figure 3). As an exploratory analysis, we examined the effects of *group* and *time* on the levels of RNT reported after 8 PM. The results showed neither a significant main effect nor a significant interaction effect. It is unlikely that the MBI specifically reduces the evening levels of RNT. However, note that the last EMA signal was sent before 10 PM each day, whereas a body scan was performed right before going to sleep (typically after the last EMA signal was sent). Our design might be suboptimal to capture the immediate effect of a body scan on RNT at night.

Table 3. Multilevel model predicting ecological momentary assessment–assessed repetitive negative thinking[a].

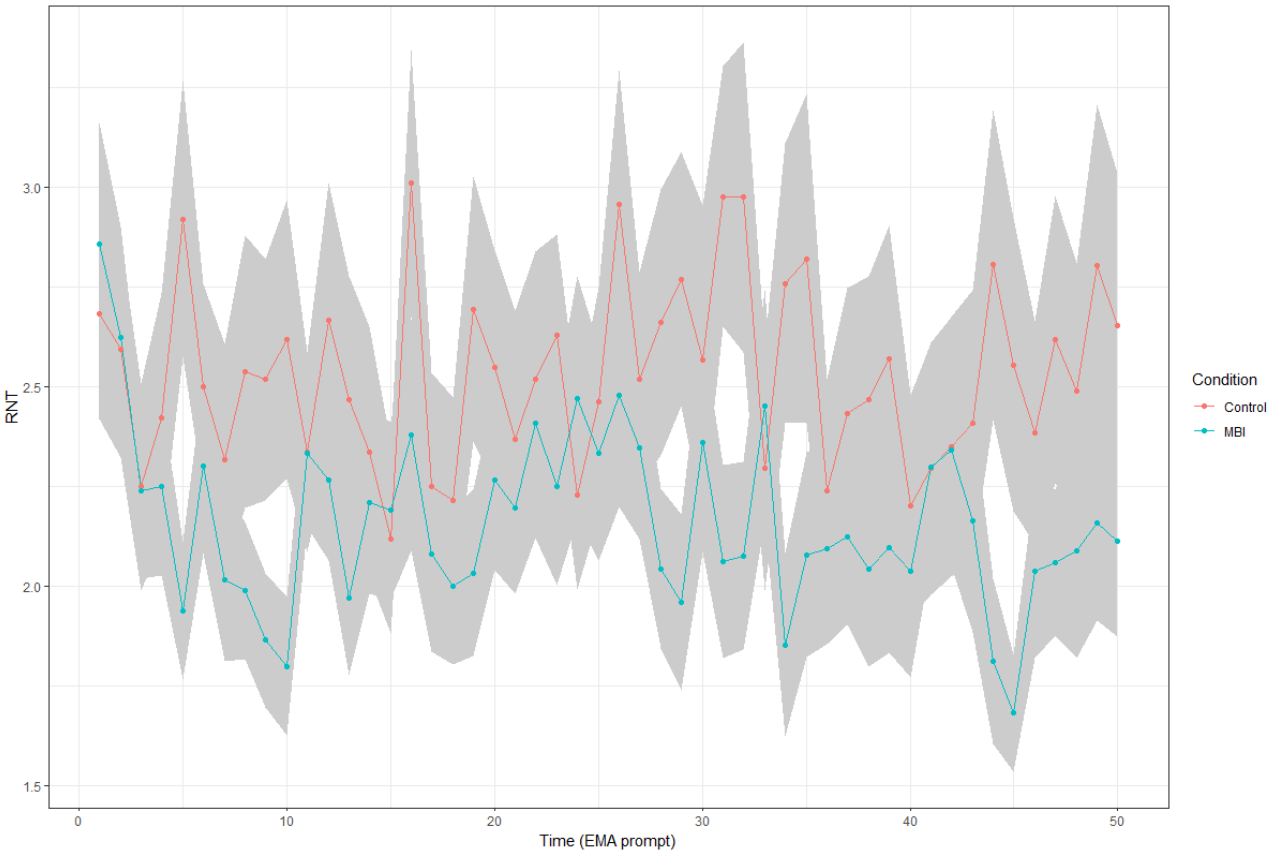
Predictor	*B* (95% CI; SE)	*t* test (*df*)	*P* value
Intercept	2.43 (2.08 to 2.77; 0.18)	13.84 (66.8)	<.001
Time[b]	0.00 (−0.01 to 0.01; 0.00)	0.61 (67.6)	.54
Group[c]	−0.17 (−0.64 to 0.31; 0.24)	−0.69 (66.7)	.49
Time:group	−0.01 (−0.02 to 0.00; 0.01)	−1.02 (67.1)	.31

[a]There were 2448 observations across 68 participants.

[b]Ecological momentary assessment prompt ranging from 1 to 50.

[c]Mindfulness-based intervention group versus no-training control group.

Figure 3. Development of repetitive negative thinking (RNT) over the course of the ecological momentary assessment (EMA). Mean RNT scores are illustrated for the no-training control group and the intervention group; the latter performed a 10-day 5-minute mindfulness-based intervention (MBI) at night. The figure covers the 10-day MBI training phase, during which EMA data were collected. The time spans 50 prompts across the 10-day EMA. The gray field indicates the SE.



Hypothesis 2: Effects From Pre- to Postintervention and Follow-Up Assessments

To test hypothesis 2, separate multilevel models were estimated for each outcome measure of the PTQ and the DASS. For the PTQ, the results (Table 4) showed significant main effects of time (both for the postintervention assessment [DT1] and the follow-up assessment [DT2]), qualified by a significant interaction between *group* and DT2 (but not DT1). Both groups exhibited an equally significant reduction from the T0 to T1

assessments (intervention group: B=−6.40, SE 1.29; $t_{57.2}$=−4.98; P<.001; d_{GMA}=−0.66; control group: B=−4.18). However, the intervention group showed a larger reduction from the T0 to T2 assessments than the control group (intervention group: B=−8.70, SE 1.35; $t_{95.4}$=−6.43; P<.001; d_{GMA}=−0.91; control group: B=−3.31). These results suggest that the PTQ score continued to decrease even after the completion of the MBI (Figure 4A). The results were overall unchanged even after controlling for age, gender, and engagement in mindfulness practice, which spontaneously continued up to the follow-up assessment. As an exploratory analysis, we tested the mean differences in the PTQ scores at the follow-up assessment between participants who engaged in mindfulness practice and those who did not. We identified no significant difference for the MBI group (mean 21.6, SD 9.45; mean 18.3, SD 6.80; t_{29}=−1.07; P=.29). However, for the control group, participants who spontaneously performed mindfulness practice after the T1 assessment up to the follow-up assessment showed lower PTQ scores than those who did not (mean 22.36, SD 8.35; mean 31.64, SD 8.82; t_{26}=2.78; P=.01).

For the DASS, we found no significant main or interaction effects on the depression and anxiety subscales (Table 4). However, the stress subscale had significant interactions between group and the 2 time dummies, suggesting that the intervention group experienced a larger reduction in stress from the T0 to T1 assessments (B=−2.65, SE 0.55; $t_{89.6}$=−4.80; P<.001; d_{GMA}=−0.78) and from the T0 to T2 assessments (B=−2.11, SE 0.55; $t_{70.9}$=−3.83; P<.001; d_{GMA}=−0.60) than the control group (T0 to T1 assessments: B=−0.05; T0 to T2 assessments: B=−0.49; Figures 4B-4D). We added age and gender as covariates to the model, which did not change the results. Another covariate, namely continued engagement in mindfulness up to the follow-up assessment, had no substantial influence on the interaction between group and DT1 (B=−1.90, SE 0.82; $t_{108.3}$=−2.33; P=.02), but this covariate reduced the effect slightly at DT2 (B=−1.33, SE 0.80; $t_{108.3}$=−1.66; P=.01). This may point to the possibility that the sustained stress reduction at the follow-up assessment is attributed to the spontaneous engagement in mindfulness practice after the postintervention assessment. However, an exploratory analysis identified no significant difference in the levels of stress at the follow-up assessment between participants who continued mindfulness practice and those who did not (mean 4.63, SD 3.53; mean 3.92, SD 2.50; t_{29}=0.61; P=.55) for the MBI group. For the control group, participants who spontaneously performed mindfulness practice after the T1 assessment up to the follow-up assessment showed slightly lower stress levels than those who did not (mean 3.60, SD 2.07; mean 6.18, SD 3.81; t_{25}=1.96; P=.06).

Table 4. Multilevel models predicting the effect of a 5-minute mindfulness-based intervention at nighttime on repetitive negative thinking and psychological distress symptoms (n=68)[a].

Outcome and predictor	*B* (95% CI; SE)	*t* test (*df*)	*P* value
PTQ[b]			
Intercept	31.48 (28.19 to 34.78; 1.68)	18.74 (96.0)	<.001
DT1[c]	−4.18 (−6.88 to −1.48; 1.38)	−3.04 (59.0)	.002
DT2[d]	−3.31 (−6.09 to −0.53; 1.42)	−2.33 (94.8)	.02
Group	−2.51 (−7.10 to 2.08; 2.34)	−1.07 (96.0)	.28
Group[e]:DT1	−2.21 (−5.91 to 1.48; 1.88)	−1.18 (58.2)	.24
Group:DT2	−5.39 (−9.24 to −1.55; 1.96)	−2.75 (95.1)	.006
DASS-D[f]			
Intercept	3.94 (2.76 to 5.12; 0.60)	6.53 (99.6)	<.001
DT1	−0.56 (−1.64 to 0.51; 0.55)	−1.03 (113.4)	.30
DT2	−0.46 (−1.50 to 0.58; 0.53)	−0.86 (112.7)	.39
Group	−0.19 (−1.86 to 1.47; 0.85)	−0.23 (102.9)	.82
Group:DT1	−0.58 (−2.05 to 0.89; 0.75)	−0.77 (113.2)	.44
Group:DT2	−0.41 (−1.88 to 1.05; 0.75)	−0.55 (113.2)	.58
DASS-A[g]			
Intercept	3.48 (2.55 to 4.42; 0.48)	7.30 (97.4)	<.001
DT1	−1.00 (−1.81 to −0.18; 0.42)	−2.39 (114.7)	.02
DT2	−0.72 (−1.55 to 0.11; 0.42)	−1.71 (115.1)	.09
Group	−0.80 (−2.10 to 0.51; 0.67)	−1.20 (97.4)	.23
Group:DT1	−0.26 (−1.38 to 0.86; 0.57)	−0.46 (114.2)	.65
Group:DT2	−0.25 (−1.39 to 0.89; 0.58)	−0.43 (114.7)	.67
DASS-S[h] (n=67)			
Intercept	5.94 (4.64 to 7.24; 0.66)	9.00 (89.3)	<.001
DT1	−0.05 (−1.22 to 1.11; 0.59)	−0.09 (84.5)	.93
DT2	−0.49 (−1.63 to 0.65; 0.58)	−0.85 (71.1)	.40
Group	0.70 (−1.09 to 2.50; 0.92)	0.77 (90.1)	.44
Group:DT1	−2.59 (−4.18 to −1.01; 0.81)	−3.20 (86.9)	.001
Group:DT2	−1.62 (−3.19 to −0.05; 0.80)	−2.02 (71.0)	.04

[a]There were 184 to 190 observations across participants.

[b]PTQ: Perseverative Thinking Questionnaire.

[c]DT1: Dummy-coded time variables representing the differences between the T0 (baseline) and T1 (day after the intervention) assessments.

[d]DT2: Dummy-coded time variables representing the differences between the T0 (baseline) and T2 (2-month follow-up) assessments.

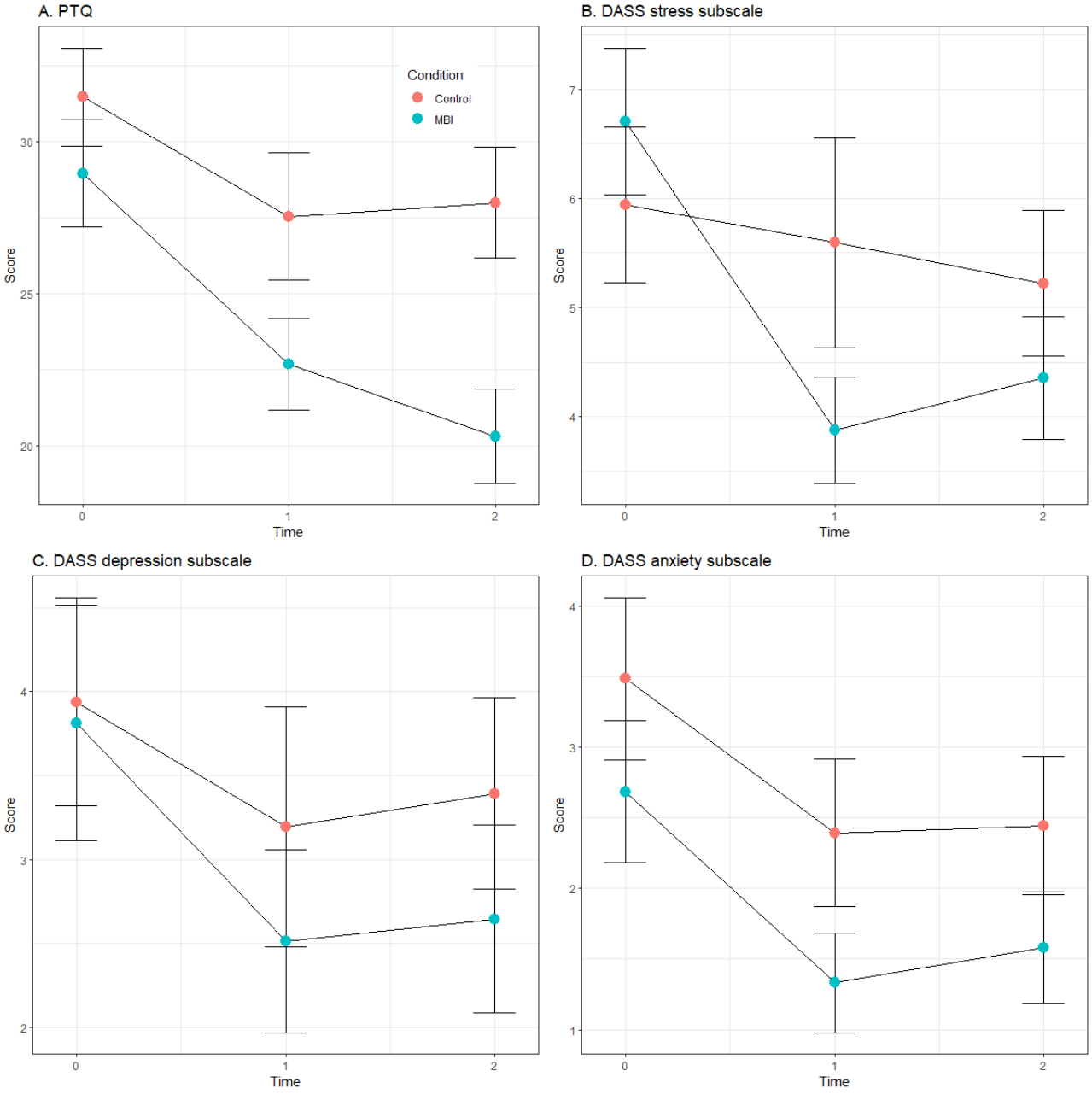
[e]Mindfulness-based intervention group versus no-training control group.

[f]DASS-D: Depression Anxiety Stress Scale, depression subscale.

[g]DASS-A: Depression Anxiety Stress Scale, anxiety subscale.

[h]DASS-S: Depression Anxiety Stress Scale, stress subscale.

Figure 4. Changes in repetitive negative thinking and psychological distress (A: repetitive negative thinking, B: stress, C: depressive symptoms, and D: anxiety) from baseline to the postintervention and 2-month follow-up assessments. Changes are illustrated for the no-training control group and the intervention group; the latter performed a 10-day 5-minute mindfulness-based intervention (MBI) at night. The error bars indicate SEs. Time: 0=baseline assessment, 1=postintervention assessment, and 2=2-month follow-up assessment. DASS: Depression Anxiety Stress Scale; PTQ: Perseverative Thinking Questionnaire.



Discussion

Principal Findings

Brief MBIs are a promising tool for improving mental health in an accessible way. However, little is known about the real-time effect of minimal (ie, 5-minute) MBIs on RNT, which is assumed to be the underlying reason for the effectiveness of an MBI. This study investigated whether a minimal MBI (a 5-minute body scan performed daily for 10 days) before sleeping would have positive effects on RNT and psychological distress symptoms (ie, depressive, anxiety, and stress symptoms) by comparing an intervention group with a no-training control group. First, we examined RNT in real time using EMA during the MBI training phase (hypothesis 1). Contrary to our expectations, the results showed no significant group differences in RNT during the training phase. Second, we examined the effects of the 5-minute body scan immediately after the intervention (T1) and at the 2-month follow-up assessment (T2; hypothesis 2). We found that both groups had significant reductions in questionnaire-based RNT at the postintervention assessment (although there was no group difference); however, the intervention group showed a significantly larger reduction than the control group at the follow-up assessment. As hypothesized, only the intervention group showed a decrease in stress at both postintervention and 2-month follow-up time points. We did not find any intervention effects on depressive or anxiety symptoms.

Taking a closer look at hypothesis 1, the results regarding the momentary levels of RNT, as assessed by EMA, showed that a minimal MBI may not induce acute changes in RNT during the training phase. This finding is inconsistent with the theory [9] and contradicts previous empirical evidence where RNT declined rapidly during the first week of an MBI [25]. However, given that the previous EMA study [25] used more intensive training (ie, a 20-minute daily MBI), it is possible that our 5-minute body scan was not sufficient to observe immediate changes in RNT.

Examining hypothesis 2, we found that a minimal MBI administered before sleeping is effective in decreasing questionnaire-based RNT, although significant group differences emerged only at the T2 assessment and not at the T1 assessment. Although the effect of the intervention started to appear early, it was only after continued improvement during the follow-up phase that the questionnaire-based RNT scores were significantly different from the control condition, that is, the effect seems to have accumulated gradually, and the difference from the control condition became visible not immediately during and after the intervention but at the 2-month follow-up assessment with increased mindfulness practice experience. This explanation is supported by the fact that 61% (19/31) of the participants in the intervention group continued mindfulness practice even after the MBI training phase. However, the exploratory analysis revealed no statistically significant difference in RNT levels at the T2 assessment between participants who continued the MBI and those who did not, which contradicts the idea that the MBI effect accumulates through spontaneous practice. Another aspect that should be mentioned is that the control group showed a significant decrease in RNT at the postintervention assessment. This decrease may suggest that the EMA worked as a self-monitoring technique, which helped participants to realize how frequently they had engaged in RNT. In addition, control participants who spontaneously started mindfulness practice after the T1 assessment showed lower levels of RNT and stress than those who did not. Taken together, it seems that the spontaneous use of mindfulness practice plays an important role to maintain and enhance the effect of mindfulness practice, but we cannot draw a solid conclusion only from these data. Furthermore, because a previous EMA study [25] documented a second significant decline in RNT between the 30th and 42nd day of mindfulness practice, it may be interesting to extend the active training phase of the 5-minute MBI, depending upon participant compliance.

Examining psychological distress symptoms, we found that participants in the intervention group showed significantly larger decreases in their stress levels than those in the control group at both the T1 and T2 assessments. This finding is in line with the body of evidence reporting the beneficial effects of an MBI on stress [12,13,57]. However, the effect of our minimal MBI was not significant for depressive or anxiety symptoms. Given that our sample showed low levels of depressive and anxiety symptoms at baseline, our minimal MBI sought to diminish what had already been low, which could be responsible for the nonsignificant effects on the depressive and anxiety outcomes (Table 1). Our sample only showed slightly higher means of the DASS depression and anxiety subscales scores than those of a large nonclinical sample in the literature: 2.83 for depressive symptoms, 1.88 for anxiety symptoms, and 4.73 for stress symptoms [58]. By contrast, the means of DASS scores in a clinical sample were 6.3 for depressive symptoms, 4.1 for anxiety symptoms, and 5.7 for stress symptoms [48]. Future studies may benefit from recruiting participants with a broader range of subclinical psychological distress symptoms.

Moreover, stress may be more acutely influenced by mindfulness practice than depressive and anxiety symptoms via autonomic arousal: in a study measuring heart rate variability over a 10-day MBI, perceived stress and acute physiological stress decreased for the mindfulness practice [57]. By contrast, long-term cognitive changes may be required to alleviate depressive and anxiety symptoms. On the one hand, more practice may be needed for cognitive change because, initially, awareness is brought to unpleasant experiences, and it is only in the second step that coping takes place through the acceptance of these experiences and the development of a nonjudgmental attitude [6]. On the other hand, a different type of MBI may be needed to trigger cognitive changes because there is evidence showing that the mindfulness practice used in this study (ie, body scan) is inferior to other mindfulness practices (eg, sitting meditation and yoga) in its ability to enhance a nonjudgmental stance [59], and programs such as MBCT combine mindfulness meditation and cognitive techniques to elicit cognitive change (compare with the study by Segal et al [9]). Nonetheless, regarding the observed effects on RNT and stress, our findings are promising, bearing in mind that RNT is a target for the prevention and treatment of highly prevalent disorders such as depression and anxiety [27], which are considered a consequence of ongoing stress [12].

We observed excellent compliance rates for MBI completion in this study, underscoring the feasibility of our 5-minute body scan. In addition, we noticed continued interest in training in our sample because the majority of the participants (19/31, 61%) spontaneously maintained mindfulness practice after the active training phase. Thus, our data suggest that a 5-minute MBI at night is feasible and that the minimal time investment of 5 minutes can be seen as getting a foot in the door of mindfulness practice, leaving room for the possibility that those who experience benefits may eventually increase their practice times.

Similar to the findings of other studies in the field, our findings need to be interpreted considering a number of limitations. First, we did not use a blinded experimenter, and participants in both groups were informed about the aim of the study. Therefore, we cannot exclude the possibility that the results are biased by social desirability and a tendency to please the experimenter. This may explain the frequent (and somewhat unexpected) engagements in mindfulness practice among control participants, who might have been primed by the study information and assessments. Second, our data had an imbalance in terms of gender because the majority of the participants (58/68, 85%) in our study were women. Another potential issue concerning the sample characteristics is that we did not selectively recruit individuals with high levels of RNT or psychological distress. This is because we wanted to show the effectiveness of a minimal MBI used daily in a nonclinical population as a method of RNT reduction and stress management. However, the

downside is that the effect might be underestimated because many of the participants showed low-to-moderate levels of RNT and distress at baseline. In addition, it should be noted that the sample size was relatively small (refer to the power analysis details in the *Participants* subsection of the *Methods* section), and this may have prevented us from detecting a significant effect of the MBI on EMA-assessed RNT. Third, adherence to the MBI was only measured via self-report, leaving the possibility that the actual amount of mindfulness practice differed from reported adherence. Fourth, we chose to implement an MBI at night; however, we were unable to incorporate conditions examining the effect of the MBI at other times of the day. Therefore, the opposite of our proposed rationale may be true. Although we have argued that targeting RNT at night when it is at its highest [33,34] would increase MBI effectiveness, it is possible that it was in fact harder to disengage from RNT at that time point. Thus, an MBI might be especially beneficial when one is in a state of mind where the attentional mindful focus can be attained throughout the practice without having to fight RNT. Fifth and last, we used a no-training control group, which provides relatively weak evidence for the efficacy of the intervention compared with an active control group.

Conclusions

Notwithstanding the aforementioned limitations, the results of this study suggest that a 5-minute body scan before sleeping can effectively decrease RNT and stress. This minimal dosage makes an MBI more accessible than the traditional or full-package program to a wide range of audiences and would therefore be particularly suited for treating subclinical RNT and stress because it does not cost too much effort for training. However, future research may benefit from a more rigorous study design comparing the effects of different mindfulness exercises (eg, body scan, sitting meditation, and yoga) at different times of the day (eg, morning, noon, and night) to establish concrete recommendations for individual practitioners. This would enable the optimization of an MBI that could effectively promote mental health both acutely and in the long term.

Acknowledgments

This study was supported by the German Research Foundation (TA 1466/1-1).

Conflicts of Interest

None declared.

References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, CDBE2010Study Group. Cost of disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011 Oct;21(10):718-779. [doi: 10.1016/j.euroneuro.2011.08.008] [Medline: 21924589]
3. Mack S, Jacobi F, Beesdo-Baum K, Gerschler A, Strehle J, Höfler M, et al. Functional disability and quality of life decrements in mental disorders: results from the mental health module of the German health interview and examination survey for adults (DEGS1-MH). Eur Psychiatry 2015 Sep;30(6):793-800. [doi: 10.1016/j.eurpsy.2015.06.003] [Medline: 26169476]
4. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
5. Versluis A, Verkuil B, Spinhoven P, van der Ploeg MM, Brosschot JF. Changing mental health and positive psychological well-being using ecological momentary interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jun 27;18(6):e152 [FREE Full text] [doi: 10.2196/jmir.5642] [Medline: 27349305]
6. Creswell JD. Mindfulness interventions. Annu Rev Psychol 2017 Jan 03;68:491-516. [doi: 10.1146/annurev-psych-042716-051139] [Medline: 27687118]
7. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York, NY: Delacourt Press; 1990.
8. Kabat-Zinn J. Mindfulness-based interventions in context: past, present, and future. Clin Psychol Sci Pract 2003;10(2):144-156 [FREE Full text] [doi: 10.1093/clipsy.bpg016]
9. Segal ZV, Williams JM, Teasdale JD. Mindfulness-Based Cognitive Therapy for Depression: A New Approach to Preventing Relapse. 2nd edition. New York, NY: The Guilford Press; 2013.
10. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. J Consult Clin Psychol 2010 Apr;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
11. Kuyken W, Warren FC, Taylor RS, Whalley B, Crane C, Bondolfi G, et al. Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials. JAMA Psychiatry 2016 Jun 01;73(6):565-574 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.0076] [Medline: 27119968]

12. Sharma M, Rush SE. Mindfulness-based stress reduction as a stress management intervention for healthy individuals: a systematic review. J Evid Based Complementary Altern Med 2014 Oct;19(4):271-286. [doi: 10.1177/2156587214543143] [Medline: 25053754]
13. Khoury B, Sharma M, Rush SE, Fournier C. Mindfulness-based stress reduction for healthy individuals: a meta-analysis. J Psychosom Res 2015 Jun;78(6):519-528. [doi: 10.1016/j.jpsychores.2015.03.009] [Medline: 25818837]
14. Querstret D, Morison L, Dickinson S, Cropley M, John M. Mindfulness-based stress reduction and mindfulness-based cognitive therapy for psychological health and well-being in nonclinical samples: a systematic review and meta-analysis. Int J Stress Manag 2020 Nov;27(4):394-411 [FREE Full text] [doi: 10.1037/str0000165]
15. Mao L, Li P, Wu Y, Luo L, Hu M. The effectiveness of mindfulness-based interventions for ruminative thinking: a systematic review and meta-analysis of randomized controlled trials. J Affect Disord 2023 Jan 15;321:83-95. [doi: 10.1016/j.jad.2022.10.022] [Medline: 36302490]
16. Feruglio S, Matiz A, Grecucci A, Pascut S, Fabbro F, Crescentini C. Differential effects of mindfulness meditation conditions on repetitive negative thinking and subjective time perspective: a randomized active-controlled study. Psychol Health 2021 Nov;36(11):1275-1298. [doi: 10.1080/08870446.2020.1836178] [Medline: 33094652]
17. Dimidjian S, Segal ZV. Prospects for a clinical science of mindfulness-based intervention. Am Psychol 2015 Oct;70(7):593-620 [FREE Full text] [doi: 10.1037/a0039589] [Medline: 26436311]
18. Goldberg SB, Hanley AW, Baldwin SA, Bernstein A, Garland EL. Does mindfulness practice promote psychological functioning or is it the other way around? A daily diary study. Psychotherapy (Chic) 2020 Sep;57(3):310-322 [FREE Full text] [doi: 10.1037/pst0000286] [Medline: 32352811]
19. Howarth A, Smith JG, Perkins-Porras L, Ussher M. Effects of brief mindfulness-based interventions on health-related outcomes: a systematic review. Mindfulness 2019 May 31;10(10):1957-1968 [FREE Full text] [doi: 10.1007/s12671-019-01163-1]
20. Schumer MC, Lindsay EK, Creswell JD. Brief mindfulness training for negative affectivity: a systematic review and meta-analysis. J Consult Clin Psychol 2018 Jul;86(7):569-583 [FREE Full text] [doi: 10.1037/ccp0000324] [Medline: 29939051]
21. Huffziger S, Ebner-Priemer U, Eisenbach C, Koudela S, Reinhard I, Zamoscik V, et al. Induced ruminative and mindful attention in everyday life: an experimental ambulatory assessment study. J Behav Ther Exp Psychiatry 2013 Sep;44(3):322-328 [FREE Full text] [doi: 10.1016/j.jbtep.2013.01.007] [Medline: 23466521]
22. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. J Psychiatr Res 2016 Apr;75:116-123 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.01.011] [Medline: 26851494]
23. Lahtinen O, Salmivalli C. The relationship between mindfulness meditation and well-being during 8 weeks of ecological momentary assessment. Mindfulness 2019 Nov 18;11(1):255-263 [FREE Full text] [doi: 10.1007/s12671-019-01248-x]
24. Larson R, Csikszentmihalyi M. The experience sampling method. In: Csikszentmihalyi M, editor. Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi. Dordrecht, The Netherlands: Springer; 2014:21-34.
25. Andreotti E, Congard A, Le Vigouroux S, Dauvier B, Illy J, Poinsot R, et al. Rumination and mindlessness processes: trajectories of change in a 42-day mindfulness-based intervention. J Cogn Psychother 2018 Jun;32(2):127-139. [doi: 10.1891/0889-8391.32.2.127] [Medline: 32746402]
26. Ehring T, Watkins ER. Repetitive negative thinking as a transdiagnostic process. Int J Cogn Ther 2008 Sep;1(3):192-205 [FREE Full text] [doi: 10.1521/ijct.2008.1.3.192]
27. Topper M, Emmelkamp PM, Watkins E, Ehring T. Prevention of anxiety disorders and depression by targeting excessive worry and rumination in adolescents and young adults: a randomized controlled trial. Behav Res Ther 2017 Mar;90:123-136. [doi: 10.1016/j.brat.2016.12.015] [Medline: 28049069]
28. Monteregge S, Tsagkalidou A, Cuijpers P, Spinhoven P. The effects of different types of treatment for anxiety on repetitive negative thinking: a meta‐analysis. Clin Psychol Sci Pract 2020 Jan 06;27(2):6-16 [FREE Full text] [doi: 10.1111/cpsp.12316]
29. Spinhoven P, Klein N, Kennis M, Cramer AO, Siegle G, Cuijpers P, et al. The effects of cognitive-behavior therapy for depression on repetitive negative thinking: a meta-analysis. Behav Res Ther 2018 Jul;106:71-85. [doi: 10.1016/j.brat.2018.04.002] [Medline: 29699700]
30. Alsubaie M, Abbott R, Dunn B, Dickens C, Keil TF, Henley W, et al. Mechanisms of action in mindfulness-based cognitive therapy (MBCT) and mindfulness-based stress reduction (MBSR) in people with physical and/or psychological conditions: a systematic review. Clin Psychol Rev 2017 Jul;55:74-91. [doi: 10.1016/j.cpr.2017.04.008] [Medline: 28501707]
31. van der Velden AM, Kuyken W, Wattar U, Crane C, Pallesen KJ, Dahlgaard J, et al. A systematic review of mechanisms of change in mindfulness-based cognitive therapy in the treatment of recurrent major depressive disorder. Clin Psychol Rev 2015 Apr;37:26-39 [FREE Full text] [doi: 10.1016/j.cpr.2015.02.001] [Medline: 25748559]
32. Gu J, Strauss C, Bond R, Cavanagh K. How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. Clin Psychol Rev 2015 Apr;37:1-12. [doi: 10.1016/j.cpr.2015.01.006] [Medline: 25689576]

33. Moberly NJ, Watkins ER. Ruminative self-focus and negative affect: an experience sampling study. J Abnorm Psychol 2008 May;117(2):314-323 [FREE Full text] [doi: 10.1037/0021-843X.117.2.314] [Medline: 18489207]
34. Takano K, Tanno Y. Diurnal variation in rumination. Emotion 2011 Oct;11(5):1046-1058. [doi: 10.1037/a0022757] [Medline: 21517158]
35. Takano K, Sakamoto S, Tanno Y. Ruminative self-focus in daily life: associations with daily activities and depressive symptoms. Emotion 2013 Aug;13(4):657-667. [doi: 10.1037/a0031867] [Medline: 23527502]
36. Takano K, Ueno M, Tanno Y. Self-focused thinking predicts nighttime physiological de-arousal. Biol Psychol 2014 Mar;97:9-14. [doi: 10.1016/j.biopsycho.2014.01.001] [Medline: 24444702]
37. Takano K, Sakamoto S, Tanno Y. Repetitive thought impairs sleep quality: an experience sampling study. Behav Ther 2014 Jan;45(1):67-82. [doi: 10.1016/j.beth.2013.09.004] [Medline: 24411116]
38. Zoccola PM, Dickerson SS. Extending the recovery window: effects of trait rumination on subsequent evening cortisol following a laboratory performance stressor. Psychoneuroendocrinology 2015 Aug;58:67-78. [doi: 10.1016/j.psyneuen.2015.04.014] [Medline: 25965871]
39. Zoccola PM, Dickerson SS, Zaldivar FP. Rumination and cortisol responses to laboratory stressors. Psychosom Med 2008 Jul;70(6):661-667. [doi: 10.1097/PSY.0b013e31817bbc77] [Medline: 18606726]
40. Li C, Kee YH, Lam LS. Effect of brief mindfulness induction on university athletes' sleep quality following night training. Front Psychol 2018 Apr 12;9:508 [FREE Full text] [doi: 10.3389/fpsyg.2018.00508] [Medline: 29706914]
41. Timm C, Rachota-Ubl B, Beddig T, Zamoscik VE, Ebner-Priemer U, Reinhard I, et al. Mindfulness-based attention training improves cognitive and affective processes in daily life in remitted patients with recurrent depression: a randomized controlled trial. Psychother Psychosom 2018;87(3):184-186. [doi: 10.1159/000488862] [Medline: 29768276]
42. Random.org: true random number service. RANDOM.ORG. URL: https://www.random.org/ [accessed 2023-06-04]
43. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
44. Birtwell K, Williams K, van Marwijk H, Armitage CJ, Sheffield D. An exploration of formal and informal mindfulness practice and associations with wellbeing. Mindfulness (N Y) 2019;10(1):89-99 [FREE Full text] [doi: 10.1007/s12671-018-0951-y] [Medline: 30662573]
45. Ussher M, Spatz A, Copland C, Nicolaou A, Cargill A, Amini-Tabrizi N, et al. Immediate effects of a brief mindfulness-based body scan on patients with chronic pain. J Behav Med 2014 Feb;37(1):127-134. [doi: 10.1007/s10865-012-9466-5] [Medline: 23129105]
46. Ribeiro L, Atchley RM, Oken BS. Adherence to practice of mindfulness in novice meditators: practices chosen, amount of time practiced, and long-term effects following a mindfulness-based intervention. Mindfulness (N Y) 2018 Apr;9(2):401-411 [FREE Full text] [doi: 10.1007/s12671-017-0781-3] [Medline: 30881517]
47. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): validation of a content-independent measure of repetitive negative thinking. J Behav Ther Exp Psychiatry 2011 Jun;42(2):225-232 [FREE Full text] [doi: 10.1016/j.jbtep.2010.12.003] [Medline: 21315886]
48. Nilges P, Essau C. Die depressions-angst-stress-skalen. Schmerz 2015 Jul 24;29(6):649-657 [FREE Full text] [doi: 10.1007/s00482-015-0019-z]
49. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the beck depression and anxiety inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
50. Meers K, Dejonckheere E, Kalokerinos EK, Rummens K, Kuppens P. mobileQ: a free user-friendly application for collecting experience sampling data. Behav Res Methods 2020 Aug;52(4):1510-1515. [doi: 10.3758/s13428-019-01330-1] [Medline: 31898294]
51. Rosenkranz T, Takano K, Watkins ER, Ehring T. Assessing repetitive negative thinking in daily life: development of an ecological momentary assessment paradigm. PLoS One 2020 Apr 20;15(4):e0231783 [FREE Full text] [doi: 10.1371/journal.pone.0231783] [Medline: 32310979]
52. Hülsheger UR, Feinholdt A, Nübold A. A low-dose mindfulness intervention and recovery from work: effects on psychological detachment, sleep quality, and sleep duration. J Occup Organ Psychol 2015 Mar 22;88(3):464-489 [FREE Full text] [doi: 10.1111/joop.12115]
53. Mehl MR, Conner TS. Handbook of Research Methods for Studying Daily Life. New York, NY: The Guilford Press; 2012.
54. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48 [FREE Full text] [doi: 10.18637/jss.v067.i01]
55. Kuznetsova A, Brockhoff PB, Christensen RH. lmerTest package: tests in linear mixed effects models. J Stat Softw 2017;82(13):1-26 [FREE Full text] [doi: 10.18637/jss.v082.i13]
56. Feingold A. Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. Psychol Methods 2009 Mar;14(1):43-53 [FREE Full text] [doi: 10.1037/a0014699] [Medline: 19271847]
57. Kirk U, Axelsen JL. Heart rate variability is enhanced during mindfulness practice: a randomized controlled trial involving a 10-day online-based mindfulness intervention. PLoS One 2020 Dec 17;15(12):e0243488 [FREE Full text] [doi: 10.1371/journal.pone.0243488] [Medline: 33332403]

58. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
59. Sauer-Zavala SE, Walsh EC, Eisenlohr-Moul TA, Lykins EL. Comparing mindfulness-based intervention strategies: differential effects of sitting meditation, body scan, and mindful yoga. Mindfulness 2012 Aug 12;4(4):383-388 [FREE Full text] [doi: 10.1007/s12671-012-0139-9]

Abbreviations

DASS: Depression Anxiety Stress Scale
EMA: ecological momentary assessment
GMA: growth-curve modeling analysis
MBCT: mindfulness-based cognitive therapy
MBI: mindfulness-based intervention
MBSR: mindfulness-based stress reduction
ML: maximum likelihood
PTQ: Perseverative Thinking Questionnaire
RNT: repetitive negative thinking

Edited by J Torous; submitted 21.11.22; peer-reviewed by Q Chen, H Krampe; comments to author 07.04.23; revised version received 26.05.23; accepted 31.05.23; published 19.07.23.

Please cite as:
Sommerhoff A, Ehring T, Takano K
Effects of Induced Mindfulness at Night on Repetitive Negative Thinking: Ecological Momentary Assessment Study
JMIR Ment Health 2023;10:e44365
URL: https://mental.jmir.org/2023/1/e44365
doi:10.2196/44365
PMID:37467038

Editorial

Real Life Experiences as Head of Science

Jennifer Huberty[1,2], PhD

[1]Fit Minded LLC, Phoenix, AZ, United States
[2]Mays Cancer Center, UT Health San Antonio MD Anderson, San Antonio, TX, United States

Corresponding Author:
Jennifer Huberty, PhD
Mays Cancer Center
UT Health San Antonio MD Anderson
7979 Wurzbach Rd
San Antonio, TX, 78229
United States
Phone: 1 602 935 6986
Email: jen@jenhubertyphd.com

Abstract

As the field of digital health grows and evolves, there is a critical need for evidence and theory-based interventions in health care. The need for integration of science into business is more critical than ever. To develop sustainable and scalable products, companies need scientists who understand the industry, can develop scientific strategy that aligns with business priorities, and are able to apply science cross-functionally. In this article, I share the insights I have gained about the roles and responsibilities of industry scientists and the lessons I have learned after spending 5 years as the head of science for a digital health start-up that turned unicorn.

(JMIR Ment Health 2023;10:e43820) doi:10.2196/43820

KEYWORDS

industry; digital health; business; digital health; technology; science; research; CEO; industry; founder; growth

Introduction

The field of digital health is growing and evolving rapidly, and the current market (ie, wise and skeptical customers, medical community paying attention to new tools, and advancements of regulatory bodies to ensure user safety) demands evidence and theory-based interventions in health care [1,2]. Thus, the need for scientists who can integrate science in business is more critical than ever [3,4]. To develop sustainable and scalable products, companies need scientists who can navigate the industry, develop scientific strategy that aligns with business priorities, and apply science cross-functionally. To date, there have been limited opportunities for researchers to gain relevant training for conducting science in industry. Academia and industry operate very differently, both wanting to develop solutions for better health but with different timelines, priorities, and incentives [5]. For example, in academia, developments are often theory-based and driven by meticulously designed research in carefully controlled settings to establish evidence of effectiveness. In contrast, in industry, development happens rapidly and iteratively, driven primarily by the market, often without theoretical foundations and with little to no supporting evidence [5].

Castro-Sweet and colleagues [4] recently published recommendations to better prepare scientists for meaningful careers in the industry. These included suggestions to network between trainee and industry; facilitate industry scientists as guest speakers for courses, seminars, or the like within academic settings; and find ways to shift the culture in academia to normalize industry positions [4]. However, there is also a distinct lack of resources, discussion, and established best practices on how to integrate science into business. More systematic sharing of real-world experiences and lessons learned would be advantageous to both the scientists who are currently working, or will work, in industry, and the companies that employ them. Below I share the insights I have gained about the roles and responsibilities of industry scientists and the lessons I have learned after spending 5 years as the head of science for a digital health start-up that turned unicorn.

Roles and Responsibilities

One of the most important roles as a scientist in digital health is the ability to develop scientific strategy aligned with the company's business priorities. This consideration is quite similar to how scientists often align their research with federal funding priorities. The science strategy must ensure that the company

has the evidence and support to market, sell, and/or partner in any way that is related to company goals and how the company wants to grow. As such, initially, a company would utilize data that are already available to them and then report the data (eg, publications) to support their claims. Following what is learned, the science would be developed with more sophisticated designs that will produce more rigorous evidence that is important for supporting the company's mission. Furthermore, in addition to supporting such missions, industry-based scientists have unique opportunities to continue to advance science in their respective fields owing to the novel data they have access to while working in industry.

Often in industry, science is viewed as a randomized controlled trial or something that is very time and resource intensive. It is the role of the industry scientist to teach the company about other ways in which science can be used to help them. For example, cross-sectional surveys, validated psychological measures, and other tools based on behavior change models can be used to help companies learn more about their customers, their customers' problems, and if their products are helping solve those problems. Companies generally have a mission and a vision to help their customer with a specific need, whereas the customer may be using their product for a different need, or the product may be helping the customer a way in which they did not expect. These data, collected through surveys and especially when aligned with engagement data, can help with marketing, sales and informing product. Additionally, the data can be published to enhance visibility for the company.

Industry-based scientists must also understand what it means to work cross-functionally (ie, across teams internal to the company, such as product, content, and marketing), and they must be able to execute and communicate science to these different audiences. A scientist must understand what role the other teams play in the company and how science might optimally support those teams' goals and objectives. Science should enhance roadmaps of other teams, while aiming to minimize additional work. Examples of cross-functional work includes developing cross-sectional surveys with marketing teams to learn more about the customer, working with membership to recruit participants (eg, current users), conducting brief educational sessions for the product and content teams about behavior change and evidence-based strategies for behavior change, developing external partnerships with universities and medical centers to conduct research, and guiding sales teams on how to use science to sell products.

It is also necessary for the scientist to understand institutional review board (IRB) processes and be very comfortable communicating with commercial IRBs. How data are used and shared and who may need IRB approval will vary with each situation. Clear communication and clarification when needed from a commercial IRB is an essential role of a scientist in industry.

Other responsibilities that are important for an industry scientist in more of a leadership role are building and leading a scientific advisory board (SAB); developing a process for external investigator proposals or partnerships that includes a way to bring revenue into the company; writing and communicating about scientific papers and grants; and building an internal system to conduct rigorous, ethical science as the company grows. Overall, an industry scientist must think with half a business brain and half a science brain; they must understand how to design and implement science that provides evidence for the company, and simultaneously contribute to scientific knowledge.

Lessons Learned

Here I address some of the most important lessons I learned in a leadership role as head of science at a digital health start-up. First, the selection of the SAB matters. It is important to vet and engage leading scientists who are passionate about the industry in which they are working. More importantly, the scientists that serve on the SAB should be willing to do the work to help grow the company's scientific agenda. Many times, companies use SABs to showcase that they are working with *big names* in science, but in reality, the SAB members are only attending monthly meetings and do not make meaningful contribution(s) to the company's scientific priorities. If an SAB is established with very active members that contribute to manuscripts, networking, advising external partners, and grant writing (if applicable), there is bound to be faster growth of science.

Second, scientists working in industry must be comfortable with change. Digital health start-ups are constantly reinventing themselves. They have business goals to meet, as well as investors with expectations to deal with. For example, a scientist could be working on research aligned with the current month's priority and find out halfway through data collection that the company is shifting priorities, which means the direction of research needs to be flexible, with a possibility to reset to align with new priorities, deliverables, and timelines. Scientists must also be able to answer perpetual questions, for example: *When will the data be ready? When can we share this with the press?* Therefore, scientists must make sure there are both short- and long-term projects and results that they can generate.

Third, one of the best things about being a scientist in industry who dreams of being funded by the National Institutes of Health (NIH) is that this dream can still be realized. The NIH does fund for-profit companies, and there are also ways for industry to partner with academics and receive funds to conduct collaborative research. Getting NIH funded for conducting research in industry is also beneficial for companies who can use granting success for marketing, partnerships, and sales.

Fourth, science can enhance business growth, providing a competitive edge. This is especially true for companies operating in business to business and health care markets. For companies, health care organizations, and payers to invest in a behavioral and/or digital health product, they must feel confident that the product works and that it will save money in the long term. Therefore, it is essential to obtain scientific evidence. As head of science, I have seen many sales and partnerships happen because the company was able to demonstrate published evidence for the impact and cost-effectiveness of its product.

Finally, convincing CEOs and investors (and the rest of the company) of the value of science can sometimes be challenging. It is difficult to calculate the direct return of science for a company. For example, when a paper is published, there are costs to the company, including scientists' time, IRB approval, publication processing fees, and so on. The return on that investment is not direct revenue but rather the paper provides visibility and validation of the value of the product. This generates leads for sales or partnerships, and eventually leads to increased revenue. Overall, science is versatile and contributes to revenue generation through marketing and press coverage, sales, and partnerships, but a scientist must be able to communicate that to leadership in order for these contributions to be fully appreciated.

The integration of science into industry is a nascent concept. Academic and industry environments operate differently, and a scientist must be both science- and business savvy. Training is important and, hopefully, as the supply and demand for trained scientists in digital health and health care continues to grow, training for science in industry will improve. In the meantime, learning from industry-based scientists with real-life experiences can help new scientists learn to shift the way they think and be effective at integrating science into business.

Acknowledgments

I am thankful to all my mentors and colleagues who have supported me to follow my heart and work in industry. This includes, but is not limited to, Linda Larkey, PhD; Dun Wang; Jacqueline Kerr, PhD; and Cynthia Castro-Sweet, PhD. Also, thank you to John Torous, MD, MBI, who inspired me to write this editorial.

Conflicts of Interest

None declared.

References

1. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. J Bus Res 2021 Feb;123:557-567. [doi: 10.1016/j.jbusres.2020.10.030]
2. Khirasaria R, Singh V, Batta A. Exploring digital therapeutics: the next paradigm of modern health-care industry. Perspect Clin Res 2020;11(2):54-58 [FREE Full text] [doi: 10.4103/picr.PICR_89_19] [Medline: 32670828]
3. Goldstein CM, Minges KE, Schoffman DE, Cases MG. Preparing tomorrow's behavioral medicine scientists and practitioners: a survey of future directions for education and training. J Behav Med 2017 Feb;40(1):214-226 [FREE Full text] [doi: 10.1007/s10865-016-9758-2] [Medline: 27365056]
4. Castro Sweet C, Kaye L, Alabduljabbar M, Myers V. Training the next generation of behavioral medicine scientists to accelerate digital health. Transl Behav Med 2022 Aug 17;12(8):834-840. [doi: 10.1093/tbm/ibac050] [Medline: 35866841]
5. Hingle M, Patrick H, Sacher PM, Sweet CC. The intersection of behavioral science and digital health: the case for academic-industry partnerships. Health Educ Behav 2019 Feb;46(1):5-9. [doi: 10.1177/1090198118788600] [Medline: 30041556]

Abbreviations

IRB: institutional review board
NIH: National Institutes of Health
SAB: scientific advisory board

Edited by J Torous; submitted 25.10.22; peer-reviewed by J Torous, M Jones Bell; accepted 20.11.22; published 09.01.23.

Please cite as:
Huberty J
Real Life Experiences as Head of Science
JMIR Ment Health 2023;10:e43820
URL: https://mental.jmir.org/2023/1/e43820
doi:10.2196/43820
PMID:36622751

Corrigenda and Addenda

Correction: Digital Phenotyping for Differential Diagnosis of Major Depressive Episode: Narrative Review

Eric Ettore[1], MD; Philipp Müller[2], PhD; Jonas Hinze[3], MSc; Matthias Riemenschneider[3], MD, PhD; Michel Benoit[4], MD, PhD; Bruno Giordana[4], MD; Danilo Postin[5], MSc; Rene Hurlemann[5], MD, PhD; Amandine Lecomte[6], MSc; Michel Musiol[6], PhD; Hali Lindsay[2], MSc; Philippe Robert[7], MD, PhD; Alexandra König[8], PhD

[1]Department of Psychiatry and Memory Clinic, University Hospital of Nice, Nice, France
[2]Research Department Cognitive Assistants, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Saarbrücken, Germany
[3]Department of Psychiatry and Psychotherapy, Saarland University Medical Center, Hombourg, Germany
[4]Department of Psychiatry, Hopital Pasteur, University Hospital of Nice, Nice, France
[5]Department of Psychiatry, School of Medicine and Health Sciences, Carl von Ossietzky University of Oldenburg, Bad Zwischenahn, Germany
[6]Research Department Sémagramme Team, Institut national de recherche en informatique et en automatique, Nancy, France
[7]Research Department, Cognition-Behaviour-Technology Lab, University Côte d'Azur, Nice, France
[8]Research Department Stars Team, Institut national de recherche en informatique et en automatique, Sophia Antipolis - Valbonne, France

Corresponding Author:
Eric Ettore, MD
Department of Psychiatry and Memory Clinic
University Hospital of Nice
30 Voie Romaine
Nice, 06000
France
Phone: 33 633306394
Email: ettore.e@chu-nice.fr

Related Article:

Correction of: https://mental.jmir.org/2023/1/e37225

(JMIR Ment Health 2023;10:e46831) doi:10.2196/46831

In "Digital Phenotyping for Differential Diagnosis of Major Depressive Episode: Narrative Review" (JMIR Ment Health 2023;10:e37225. doi: 10.2196/37225. PMID: 36689265; PMCID: PMC9903183) the authors made three additions to the authorship list.

The authorship is currently listed as:

Eric Ettore, Philipp Müller, Jonas Hinze, Michel Benoit, Bruno Giordana, Danilo Postin, Amandine Lecomte, Hali Lindsay, Philippe Robert, Alexandra König

And will be changed to read as follows:

Eric Ettore, Philipp Müller, Jonas Hinze, Matthias Riemenschneider, Michel Benoit, Bruno Giordana, Danilo Postin, Rene Hurlemann, Amandine Lecomte, Michel Musiol, Hali Lindsay, Philippe Robert, Alexandra König

The authors who have been recently added are affiliated with the following institutions, which are already correctly presented in the affiliation list:

Matthias Riemenschneider[3]:

Department of Psychiatry and Psychotherapy, Saarland University Medical Center, Hombourg, Germany.

Rene Hurlemann[5]:

Department of Psychiatry, School of Medicine and Health Sciences, Carl von Ossietzky University of Oldenburg, Bad Zwischenahn, Germany.

Michel Musiol[6]:

Research Department Sémagramme Team, Institut national de recherche en informatique et en automatique, Nancy, France.

The correction will appear in the online version of the paper on the JMIR Publications website on April 12, 2023, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 27.02.23; this is a non–peer-reviewed article;accepted 27.02.23; published 12.04.23.

Please cite as:
Ettore E, Müller P, Hinze J, Riemenschneider M, Benoit M, Giordana B, Postin D, Hurlemann R, Lecomte A, Musiol M, Lindsay H, Robert P, König A
Correction: Digital Phenotyping for Differential Diagnosis of Major Depressive Episode: Narrative Review
JMIR Ment Health 2023;10:e46831
URL: https://mental.jmir.org/2023/1/e46831
doi:10.2196/46831
PMID:37043763

Corrigenda and Addenda

Correction: Psychiatric Treatment Conducted via Telemedicine Versus In-Person Modality in Posttraumatic Stress Disorder, Mood Disorders, and Anxiety Disorders: Systematic Review and Meta-Analysis

Ali Abbas Shaker[1,2], BEng, MD; Stephen F Austin[1,3], MSc, PhD; Ole Jakob Storebø[1,3], MSc, PhD; Julie Perrine Schaug[1], MSc; Alaa Ayad[1], MSc; John Aasted Sørensen[4], MSc, PhD; Kristine Tarp[5,6], MSc, PhD; Henrik Bechmann[4], MSc, PhD; Erik Simonsen[2,7], PhD, MD

[1]Psychiatric Department, Region Zealand Psychiatry, Psychiatric Research Unit, Slagelse, Denmark

[2]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

[3]Department of Psychology, University of Southern Denmark Faculty of Health Sciences, Odense, Denmark

[4]Department of Engineering Technology and Didactics, Research unit: AI, Mathematics and Software, Technical University of Denmark, Ballerup, Denmark

[5]Research Unit for Digital Psychiatry, Mental Health Services in the Region of Southern Denmark, Odense, Denmark

[6]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

[7]Mental Health Services East, Copenhagen University Hospital – Psychiatry Region Zealand, Roskilde, Denmark

Corresponding Author:
Ali Abbas Shaker, BEng, MD
Psychiatric Department, Region Zealand Psychiatry, Psychiatric Research Unit
Fælledvej 6
Slagelse, 4200
Denmark
Phone: 45 58536070
Email: ashak@regionsjaelland.dk

Related Article:

Correction of: https://mental.jmir.org/2023/1/e44790

(JMIR Ment Health 2023;10:e52269) doi:10.2196/52269

In "Psychiatric Treatment Conducted via Telemedicine Versus In-Person Modality in Posttraumatic Stress Disorder, Mood Disorders, and Anxiety Disorders: Systematic Review and Meta-Analysis" (JMIR Ment Health 2023;10:e44790), the authors noted one error.

In Figures 5 and 6, the labels on the forest plot, "Favors telemedicine" and "Favors in-person," were swapped. In the revised figures, these labels are presented correctly, as follows:

Figure 5. Forest plot (patient satisfaction).

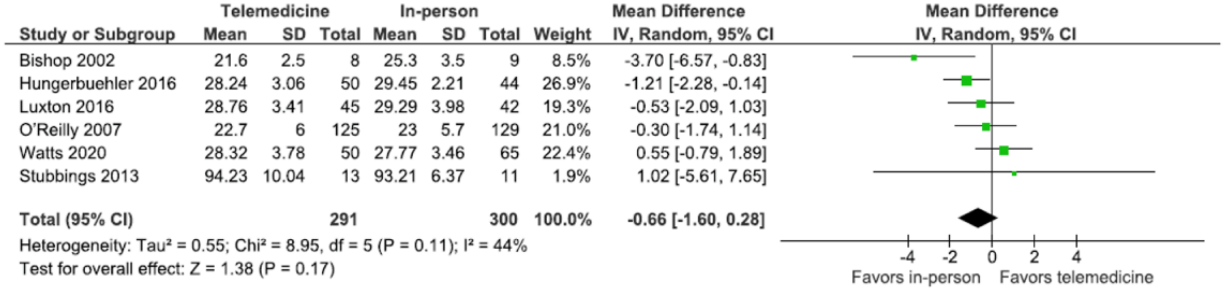
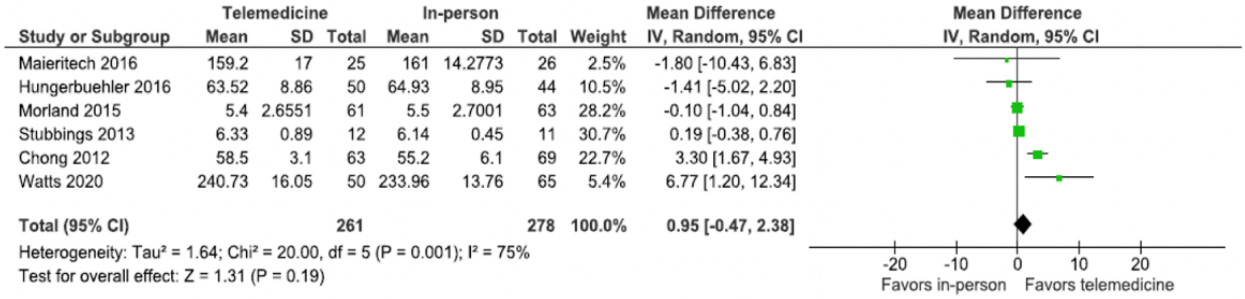


Figure 6. Forest plot (working alliance).



The correction will appear in the online version of the paper on the JMIR Publications website on September 18, 2023 together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 29.08.23; this is a non–peer-reviewed article;accepted 05.09.23; published 18.09.23.

Please cite as:
Shaker AA, Austin SF, Storebø OJ, Schaug JP, Ayad A, Sørensen JA, Tarp K, Bechmann H, Simonsen E
Correction: Psychiatric Treatment Conducted via Telemedicine Versus In-Person Modality in Posttraumatic Stress Disorder, Mood Disorders, and Anxiety Disorders: Systematic Review and Meta-Analysis
JMIR Ment Health 2023;10:e52269
URL: https://mental.jmir.org/2023/1/e52269
doi:10.2196/52269
PMID:37722110

Original Paper

Digital Practices by Citizens During the COVID-19 Pandemic: Findings From an International Multisite Study

Hannah Ramsden Marston[1*], PhD; Pei-Chun Ko[2*], PhD; Vishnunarayan Girishan Prabhu[3*], PhD; Shannon Freeman[4*], PhD; Christopher Ross[4*], MA; Iryna Sharaievska[5*], PhD; Matthew HEM Browning[5*], PhD; Sarah Earle[1*], PhD; Loredana Ivan[6*], PhD; Rubal Kanozia[7*], PhD; Halime Öztürk Çalıkoğlu[8*], PhD; Hasan Arslan[8*], PhD; Burcu Bilir-Koca[8*], PhD; Paula Alexandra Silva[9*], PhD; Sandra C Buttigieg[10*], PhD; Franziska Großschädl[11*], PhD; Gerhilde Schüttengruber[11*], PhD

[1]School of Health, Wellbeing and Social Care, The Open University, Milton Keynes, United Kingdom
[2]School of Social Sciences, Monash University, Melbourne, Australia
[3]Systems Engineering & Engineering Management, University of North Carolina, Clemson, SC, United States
[4]School of Nursing, University of Northern British Columbia, Prince George, BC, Canada
[5]Department of Parks, Recreation and Tourism Management, Clemson University, Clemson, SC, United States
[6]Communication Department, National University of Political Studies and Public Administration, Bucharest, Romania
[7]Department of Mass Communication and Media Studies, Central University of Punjab, Bathinda, India
[8]Department of Educational Sciences, Canakkale Onsekiz Mart University, Çanakkale, Turkey
[9]Department of Informatics Engineering, Center for Informatics and Systems at the University of Coimbra, University of Coimbra, Coimbra, Portugal
[10]Department of Health Systems Management and Leadership, University of Malta, Msida, Malta
[11]Institute of Nursing Science and Age and Care Research Group, Medical University Graz, Graz, Austria
[*]all authors contributed equally

Corresponding Author:
Hannah Ramsden Marston, PhD
School of Health, Wellbeing and Social Care
The Open University
Horlock Building
Walton Hall
Milton Keynes, MK7 6AA
United Kingdom
Phone: 44 7815507547
Email: Hannah.Marston@open.ac.uk

Abstract

Background: The COVID-19 pandemic brought digital practices and engagement to the forefront of society, which were based on behavioral changes associated with adhering to different government mandates. Further behavioral changes included transitioning from working in the office to working from home, with the use of various social media and communication platforms to maintain a level of social connectedness, especially given that many people who were living in different types of communities, such as rural, urban, and city spaces, were socially isolated from friends, family members, and community groups. Although there is a growing body of research exploring how technology is being used by people, there is limited information and insight about the digital practices employed across different age cohorts living in different physical spaces and residing in different countries.

Objective: This paper presents the findings from an international multisite study exploring the impact of social media and the internet on the health and well-being of individuals in different countries during the COVID-19 pandemic.

Methods: Data were collected via a series of online surveys deployed between April 4, 2020, and September 30, 2021. The age of respondents varied from 18 years to over 60 years across the 3 regions of Europe, Asia, and North America. On exploring the associations of technology use, social connectedness, and sociodemographic factors with loneliness and well-being through bivariate and multivariate analyses, significant differences were observed.

Results: The levels of loneliness were higher among respondents who used social media messengers or many social media apps than among those who did not use social media messengers or used ≤1 social media app. Additionally, the levels of loneliness

were higher among respondents who were not members of an online community support group than among those who were members of an online community support group. Psychological well-being was significantly lower and loneliness was significantly higher among people living in small towns and rural areas than among those living in suburban and urban communities. Younger respondents (18-29 years old), single adults, unemployed individuals, and those with lower levels of education were more likely to experience loneliness.

Conclusions: From an international and interdisciplinary perspective, policymakers and stakeholders should extend and explore interventions targeting loneliness experienced by single young adults and further examine how this may vary across geographies. The study findings have implications across the fields of gerontechnology, health sciences, social sciences, media communication, computers, and information technology.

International Registered Report Identifier (IRRID): RR2-10.3389/fsoc.2020.574811

(JMIR Ment Health 2023;10:e41304) doi:10.2196/41304

KEYWORDS

COVID-19; communication; gerontology; community living; technology; social media

Introduction

Background

The first cases of COVID-19 caused by SARS-CoV-2 were detected in late 2019 in Wuhan, Hubei province, China, and within months, it had spread to 113 countries in the world, leading to the declaration of a pandemic by the World Health Organization (WHO) on March 11, 2020 [1]. The COVID-19 pandemic spread across the globe and substantially impacted all aspects of daily life. Based on their cultural beliefs, political philosophies, available resources, and health care systems, nations responded differently. Several strategies, such as social distancing and isolation, case detection and contact tracing, general lockdown, and quarantine of exposed individuals, were effective in the prevention of disease spread while the virus was being studied and vaccines were being developed [1].

There is a growing body of scholarly research exploring the relationship between technology (eg, digital devices, the internet, digital gaming, social media, and mobile apps) [2,3] and loneliness [4-6]. To negotiate the constraints associated with the pandemic, technology use significantly increased, since many activities, including employment, education, health care, and other daily activities, moved to online spaces [6]. Additionally, technology has been used as a coping mechanism to follow news, get entertained, connect with others, shop online, and participate in exercise [7]. Unfortunately, despite the great range of coping strategies, loneliness prevailed among multiple groups in the population. For example, research conducted for a duration of 1 month by Groarke et al [8] at the beginning of the UK lockdown (March 23, 2020) found that the frequency of loneliness was significantly higher among younger respondents aged 18-24 years (41.0%) and 25-34 years (28.2%) than among adults aged ≥65 years (3.3%). Marital status impacted feelings of loneliness, with respondents who reported being separated or divorced (46.9%), or single or never married (40.1%) experiencing greater loneliness than those who were married/living with a partner (40.1%) or widowed (34.8%). Additionally, people who were living alone also reported higher loneliness compared to that among those with coresidents. As a result, finding ways to reduce isolation was a primary area of concern for researchers and policymakers during the pandemic, with technology use being in the forefront of this discourse as one of the potential solutions [9-18].

The pandemic brought to the fore the pivotal role the internet and Wi-Fi access played in the lives of individuals across the globe. Many individuals who conducted in-person (eg, work, leisure, and social connections) activities in the prepandemic society had to quickly transition online to ensure the same activities were achievable in this new world [7-14]. Globally, by understanding how technology was used by people living in different countries, we can better enhance our understanding of digital practices and the activities that are associated with technology use and digital practices. The United Nations (UN) [15] acknowledges that the pandemic was not only a global health crisis but also a disaster that impacted regions at the socioeconomic, security, and humanitarian levels. Further recognition notes how the pandemic has affected individuals, families, communities, and societies alike, with the UN [15] identifying strategies for socioeconomic responses.

Technology was used to reduce isolation and to address the negative outcomes of the COVID-19 pandemic and lockdowns [6,7]. The negative outcomes of the pandemic included the loss of employment and educational opportunities [5] and lack of access to the health care system [14] and mental health services, coupled with the uncertainty of the future and lack of knowledge about the virus. The issues of increased isolation and deteriorating mental health were identified as concerns during the COVID-19 pandemic [12]. To mitigate these negative outcomes, technology was employed as a possible solution [7]. This was particularly true in areas where access to technology and the internet was relatively universal [16-18]. For example, in a study conducted in April 2020 among 1374 US residents (54% female), increased use of digital communication was reported across platforms, including text messaging (43%), voice calls (36%), social media (35%), and video calls (30%) [19]. Interestingly, the same study also reported reduced digital communication use in 5% of participants during the pandemic, which included communication over social media (8%), voice calls (9%), email (10%), video calls (13%), and online gaming (17%) [19]. Younger people and women who were living alone and those who were concerned about their internet access reported increased use of digital communication, while older people reported reduced use of digital communication [19,20].

For many people globally, addressing social isolation experienced by themselves, friends, members of the community, or loved ones during lockdowns played a key role in their mental health [18,21]. Technology afforded people opportunities to remain digitally connected and explore new leisure experiences across virtual and digital environments [16,21,22]. Pennington found that social networking sites could allow users to "stay connected," and findings from this study ascertained that respondents who were actively engaging in posts felt less loneliness than those who were engaging with individuals on a face-to-face basis [23]. Technology use to maintain contact with family and friends is common across both rural and urban environments; however, a pre–COVID-19 study exploring technology use by adults aged 70 years or older in the United Kingdom and Canada found that participants from rural communities were more positive about the use of the internet, but the viewpoint of social media platforms was negative, and these individuals did not have a social media profile and preferred to engage in face-to-face conversations [24]. Participants in rural Canada engaged with social media platforms more than participants in rural United Kingdom, and participants in the urban areas of the United Kingdom and Canada used social media and networking sites frequently [24]. These studies suggested that the experiences with technology may differ across age, geography, and other demographic characteristics. Therefore, it is important to further understand the unique differences in the relationships among technology use, social isolation, self-reported mental health and well-being, and demographic characteristics during the COVID-19 pandemic.

Study Aims

This paper aimed to provide key insights from this exploratory descriptive study about the impact of loneliness and psychological well-being among people across different age cohorts and types of communities (eg, rural, urban, and metropolitan). Additionally, this paper will detail how technology played a role in access to community support via social media platforms from across diverse countries during the pandemic. The objectives were as follows: (1) to understand how technology played a role in access to community support for well-being; and (2) to examine the interaction among technology use, social isolation, and self-reported mental health and well-being during the COVID-19 pandemic across age, gender, home environment, and geography (including population density [rural, suburban, and urban] and country).

Methods

Overview

We report the methods and findings of an international multisite study conducted by a consortium of scholars from 13 countries to explore technology use, psychological well-being, COVID-19–specific questions (eg, access to support groups via social media sites), and loneliness among adults aged ≥18 years during the COVID-19 pandemic.

Study Design

The study protocol was developed by a consortium of scholars from Austria, France, Germany, India, Malta, Portugal, Romania, Singapore, Spain, Turkey, and the United Kingdom, and has been described elsewhere [25,26]. This protocol describes the process of backward translation, the methods and approaches to participant recruitment, the different measures used in the online surveys, and the different versions of the surveys pertaining to respective legislation in countries (eg, Singapore) [25]. Two additional sites (the United States and Canada) joined the consortium after the protocol was published and therefore were not included in the earlier publication. A convenience sample was used across all countries during the rapid rollout and deployment in 2020 and 2021 [25]. A virtual snowball sampling approach was applied across the partners' existing networks using the capabilities of the internet [27,28].

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Human Research Ethics Committee of The Open University (protocol code HREC/3551/MARSTON). The survey was rolled out on April 4, 2020. Each partner communicated with the project lead prior to deployment of their country survey, and all respective documentation was provided to the project lead, which in turn was shared with the institutional ethics committee for an update. Data collected from this phase are referred to as Wave 1 data.

Two additional sites (the United States and Canada) joined the consortium in November/December 2020. Small changes in the wordings of the surveys were made to accommodate for differences in North American and British English, in addition to adjusting for the options available in North American communities. For example, "Ordering from a local bakery" was replaced with "Ordering take-out food," "Streaming BBC iPlayer" was replaced with "Reading and streaming the news," and "key worker" was replaced with "essential worker." Additionally, response options pertaining to the question of race and ethnicity were added to follow census categories. Data collected from these 2 countries are referred to as Wave 2 data.

Informed consent was obtained from all subjects involved in the study. Each site received ethical approval: National University of Political Studies and Public Administration (SNSPA–Romania) (no protocol number; granted April 20, 2020); Open University of Catalonia (Spain); Singapore University of Social Sciences (Singapore) (no protocol number; granted April 23, 2020); Ethics Committee of the Universitat Oberta de Catalunya (Spain) (no protocol number; granted April 22, 2020); Department of Health Sciences Management and Leadership, University of Malta (Malta) (protocol number 5274_04052020; granted May 19, 2020); Department of Informatics Engineering (DEI)/Center for Informatics and Systems (CISUC) at the University of Coimbra (Portugal) (protocol number CE-057/2020_PaulaSilva; granted May 27, 2020); Department of Mass Communication and Media Studies at the Central University of Punjab (India) (protocol number CUPB/IEC/29/05/20_8; granted May 29, 2020); Nursing Science, Age and Care Research Group at the Medical University Graz (Austria); Department of Sociology at the University of Vienna, the Institute of Nursing Science at the Medical University of Graz (Austria) (protocol number 32-425 ex 19/20; June 5, 2020); the Board for the Ethical Review of

Research Projects of the Institute for Communication Science (IfK) of the Westphalian-Wilhelms University of Münster (Germany) (no protocol number; granted May 7, 2020); Canakkale Onsekiz Mart University (Turkey) (protocol code 2020/83; granted June 15, 2020); Clemson University (United States) (IRB2020-435); and University of Northern British Columbia (Canada) (protocol code E2021.0323.009.00; granted May 19, 2021).

Recruitment

Data collection for Waves 1 and 2 involved online survey invitations (deployed via Qualtrics) distributed through various professional and personal networks, mailing lists, social media platforms, snowball sampling, and the project website [28].

The Wave 1 survey (English/United Kingdom) was deployed online on April 4, 2020, and from that point onwards, consortium partners joined the project organically. The criteria for participation were as follows: (1) age of 18 years or above and (2) regular use of information and communication technology. The first wave of data was collected between April 4, 2020, and September 30, 2020, in 10 countries (Austria, France, Germany, India, Malta, Portugal, Romania, Singapore, Turkey, and the United Kingdom) and in 9 languages (Catalan, English, French, German, Hindi, Mandarin, Romanian, Spanish, and Turkish). Each survey was open for 3 months, with the English/United Kingdom survey closing on July 4, 2020. The final survey in the first wave of data closed at the end of September 2020. Wave 2 data were collected in the United States (March 29, 2021, to June 29, 2021) and Canada (June 29, 2021, to October 3, 2021).

Materials

The survey deployed can be found in Multimedia Appendix 1 and in the study protocol [25]. The survey included multiple questions organized into several sections. Section A focused on questions relating to computer use and behavior based on previous iterations of the survey conducted in previous projects [2,3,29-31] and described in the study protocol [25]. Section B focused on COVID-19–related questions and the purpose of using technology (eg, using social media to communicate, and challenges faced). Section C focused on activities of daily living during COVID-19. These items were new and were added to the survey to capture social connections/friendships, time spent, key worker responsibilities, and giving something back [28]. Section D focused on psychological well-being [32,33] and included 18 items and 6 aspects (autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance of psychological well-being). The Cronbach alpha was .844. Section E focused on eHealth/digital literacy [34] and included an 8-item measure (1-5 points on a Likert scale). Section F focused on loneliness and included the UCLA Loneliness Scale version 3. This measure involves a Likert scale (1-4 points) [35], and the Cronbach alpha was .862 across all countries. It has been used to accurately measure loneliness in both younger and older populations [36-38]. This survey has been applied for wider use across the general population [35]. Section G focused on digital software technologies. These items were new and were added to the survey to capture the use of technology to relay messages via a national emergency alert system (eg, mobile app, SMS text message, etc) [2,3,39]. Section H focused on demographic questions. These included age group (18-29 years old, 30-39 years old, 40-49 years old, 50-59 years old, or ≥60 years old), gender (male, female, or prefer to self-describe), education (primary or less than high school, high school, bachelor's degree, master's/professional degree, or PhD), marital status (having a partner, widowed/divorced, or single), number of people staying in the same household, employment status (working, retired, or out of a job), and physical space (metropolitan/city, suburban, small town, or rural area) [2,3,24,29-31,40,41].

The study protocol [25] describes clearly and succinctly how this project was established into a multisite project. Because of national, linguist, and legal differences, there were minor changes across the different versions of the deployed surveys. This was led by each project lead (site) and the principal investigator.

Data Analysis

Upon completion of data collection, all missing data points and data-related issues were identified and addressed. Respondents with missing data for the UCLA Loneliness Scale version 3 measure were removed prior to data analysis. Bivariate analyses were conducted to examine continuous variables (eg, UCLA 20-item Loneliness Scale version 3) among different groups based on their age, gender, type of community, etc, and a 2-sided 1-way ANOVA or Student *t* test was used based on the number of levels. For analysis of categorical variables, crosstab analyses followed by a Pearson chi-squared test and a likelihood ratio chi-squared test were performed. Ordinary least squares (OLS) regression analyses were conducted to examine to what extent the use of technology influenced the feeling of loneliness and to identify sociodemographic factors that influence the feeling of loneliness. An alpha level of $P<.05$ was used to indicate statistical significance. Estimated effect sizes were calculated with η. It should be noted that the age categories for Wave 1 and Wave 2 data were not the same because the data collected in Wave 2 included 6% of respondents aged between 50 and 59 years. To provide power for statistical tests of the data collected in Wave 2, we combined the data pertaining to respondents aged 50 years or over into a single category.

Results

Overview

In this section, both Wave 1 (collected in 2020) and Wave 2 results (collected in 2021) are presented. Table 1 presents a breakdown of the survey response rates.

Table 1. Survey response rates.

Site (country) and language	Date survey opened	Date survey closed	Sample (N=3244), n (%)
Austria			
German	June 5, 2020	September 5, 2020	240 (7.4)
Canada			
English	June 1, 2021	September 31, 2021	209 (6.4)
France			
French	May 12, 2020	August 12, 2020	135 (4.2)
Germany			
German	June 4, 2020	September 4, 2020	329 (10.1)
India			
English	May 31, 2020	August 31, 2020	320 (9.9)
Hindi	May 31, 2020	August 31, 2020	49 (1.5)
Malta			
English	May 19, 2020	August 19, 2020	103 (3.2)
Portugal			
Portuguese	May 29, 2020	August 29, 2020	37 (1.1)
Romania			
Romanian	April 20, 2020	July 20, 2020	447 (13.8)
Singapore			
English	May 17, 2020	August 17, 2020	82 (2.5)
Mandarin	May 13, 2020	August 13, 2020	17 (0.5)
Spain/South America			
Catalan/Spanish	May 4, 2020	August 4, 2020	382 (11.8)
Turkey			
Turkish	June 29, 2020	September 29, 2020	108 (3.3)
United Kingdom			
English	April 3, 2020	July 4, 2020	548 (16.9)
United States			
English	March 29, 2021	June 18, 2021	238 (7.3)

Lockdown Directives

Lockdown measures were implemented at different times across the different sites, starting as early as February and continuing until spring 2021 [42-61]. The measures implemented by respective national and regional governments varied considerably across the different sites [42-61], with several varying forms of directives being implemented across different states, provinces, and counties. Such measures included closure of all nonessential shops and retail outlets, introduction of education and work from home orders [42-61], enforcement of curfews (eg, 6 PM to 6 AM/9 PM to 6 AM) [49,58], enforcement of fines [49,51-57,59], enforcement of border controls [42-61], adoption of appropriate measures for people coming into the country [44-46,52,61], and requirement of documentation for proof of purpose (eg, grocery shopping/medicines, or going to work/emergency work) for leaving the home during lockdown [49,58]. In some instances, older adults (age ≥65 years) were allowed to leave their homes between 11 AM and 1 PM [59], while in other regions, roadblocks were used to monitor travel [49,58] and police were deployed onto public transport networks (eg, train services) [52-57].

Respondent Characteristics

Table 2 presents various sociodemographic variables, in addition to the scores relating to loneliness, psychological well-being, and social media app use across Waves 1 and 2. Although the goal of this research was not to compare Wave 1 and Wave 2 data, it was observed through the data collected and analyzed that respondents in Wave 2 reported greater loneliness and experienced lower levels of psychological well-being when compared with the findings from the data collected during Wave 1. Data analysis of the UCLA Loneliness Scale version 3 measure showed that respondents did experience loneliness during Wave 1 (mean 48.11, SD 6.26) and Wave 2 (mean 49.63, SD 9.40). Psychological well-being was greater among

respondents in Wave 1 (mean 69.04, SD 10.21) than among respondents in Wave 2 (mean 60.42, SD 10.73). Regarding the number of social media apps used by the respondents, most respondents across both waves used 3 to 4 social media apps, while few respondents used ≥5 social media apps.

Table 3 presents data relating to respondents who reported joining a specific online COVID-19 support group. Overall, less than 40% of the respondents reported being a member of an online community group, with a lower proportion in Wave 1 (265/1187, 22.3%) than in Wave 2 (132/337, 39.2%). From the data collected in Wave 2, the only significant association was between the use of social networking messengers (eg, Facebook Messenger, Snap Chat, etc) and the UCLA Loneliness Scale score (t_{347}=3.79; *P*<.001). The levels of loneliness were higher among respondents who reported using social networking messengers than among those who did not use social networking messengers.

On investigating the impact of loneliness based on the type of community respondents reported living in, there was no significant difference in the data for Wave 1. Moreover, there was no statistical significance or interaction effects between the type of community the respondents lived in and technology use for Wave 1 data.

However, observations were ascertained and significant differences were identified from the data collected in Wave 2 ($F_{3,344}$=3.28; *P*=.02). The levels of loneliness were higher among respondents living in a small town (n=115; mean 51.16, SD 8.27) than among those living in a suburban area (n=124; mean 47.82, SD 10.22; *P*=.03). There were no other significant findings involving the different types of communities and the levels of loneliness from the data collected in Wave 2.

Wave 2 data showed a significant main effect based on the feeling of loneliness and the number of social media apps used ($F_{3,332}$=4.67; *P*=.003). Respondents who used no social media apps reported the lowest levels of loneliness (n=34; mean 43.88, SD 9.80). Additionally, the findings ascertained significance among respondents who were using 1 or 2 social media apps (n=116; mean 50.16, SD 8.36; *P*=.002), 3 or 4 apps (n=152; mean 50.58, SD 9.30; *P*<.001), and ≥5 apps (n=46; mean 50.85, SD 10.83; *P*=.004). There were no other significant findings involving the number of social media apps used by the respondents and the feeling of loneliness from the data collected in Wave 2. Moreover, the type of community where the respondents lived was included as an independent variable to investigate any potential interaction effects between these variables. However, data analysis showed that there were no significant interaction effects ($F_{9,332}$=0.98; *P*=.45).

Wave 1 data showed that there were no differences in psychological well-being among the types of communities the respondents lived in. However, Wave 2 data showed that there was a significant main effect based on the type of community respondents lived in and their psychological well-being ($F_{3,361}$=4.86; *P*=.003) (Table 4).

In Wave 2, the levels of well-being were significantly lower among respondents who reported living in a rural area (n=35; mean 53.91, SD 14.02) than among those who reported living in a small town (n=119; mean 61.37, SD 10.89; *P*=.002) or a suburban area (n=131; mean 60.75, SD 11.13; *P*=.006). Data analysis of Wave 2 showed no other significant differences involving the type of community respondents lived in and their psychological well-being. Moreover, there were no significant differences ($F_{3,349}$=1.28; *P*=.28) on investigating the interaction of the type of community and the number of social media apps used with the psychological well-being of the respondents.

Tables 5 and 6 present OLS models based on the respondent characteristics and how technology use influences the feeling of loneliness according to the data collected in Waves 1 and 2, respectively. The OLS models include the independent variables of age, gender, education level, marital status, employment status, residence area, number of people living together in the same home environment, and psychological well-being, and the dependent variable of loneliness.

Two additional independent variables were included in the models and relate to the use of technology (number of social media apps used and joining a specific online COVID-19 support group). These specific independent variables were selected based on the research objective, which aimed to investigate the effects of technology use on the levels of loneliness experienced by the respondents while controlling for the characteristics during the COVID-19 pandemic. For each OLS model (Wave 1 and Wave 2), data reporting includes the estimated unstandardized coefficient (β) and standard error. Furthermore, we include the adjusted R-squared to describe the model fit.

For Wave 1, there was no association between technology use and loneliness scores. However, the levels of loneliness were higher among respondents who reported being single than among those who reported having a partner (*P*<.001). The levels of loneliness were higher among respondents who reported being unemployed than among those who reported being employed (*P*=.03). Moreover, the levels of loneliness were lower among respondents who reported having a PhD degree (*P*<.001), a master's degree or a professional degree (*P*<.001), a bachelor's degree (*P*<.001), or a high school level of education (*P*=.003) than among those who reported having a primary school level of education or no formal education at all. Data analyses showed that there were no differences among respondents located in European countries and the other countries.

Table 6 presents the data collected during Wave 2. The levels of loneliness were higher among respondents who reported being aged between 30 and 39 years than among those who reported being aged between 40 and 49 years (*P*=.04) or those who reported being aged ≥50 years (*P*=.01). Loneliness scores were higher among male respondents than among female respondents (*P*=.04). Furthermore, the levels of loneliness were higher among respondents who reported being unemployed or retired (*P*=.02) than among those who reported being employed (*P*=.04). Moreover, the levels of loneliness were higher among respondents who reported using one or more social media messaging apps (eg, Facebook, Snapchat, WhatsApp, etc) than

among those who reported not using any social media messaging apps (*P*=.003).

Table 7 presents data related to the psychological well-being of the respondents during Waves 1 and 2. The number of social media apps used and whether respondents joined (via a social media platform such as Facebook) a specific online COVID-19 support group were statistically significant. Psychological well-being was observed to be worse among respondents who reported living in a small town than among those who reported living in a metropolitan area or a city community (coefficient=−1.974; *P*=.004).

Psychological well-being was more likely to be worse among respondents who reported being single than among those who reported having a partner (coefficient=−1.768; *P*<.05). Psychological well-being was lower among respondents aged between 18 and 29 years than among those aged between 30 and 39 years (*P*=.003) and those aged between 40 and 49 years (*P*=.04). Additionally, psychological well-being was lower among male respondents than among female respondents (*P*=.006). Moreover, psychological well-being was lower among respondents who reported being unemployed than among those who reported being employed (*P*=.04). Data analysis also identified the type of community impacted in terms of psychological well-being, and psychological well-being was higher among respondents who reported living in a small-town community than among those who reported living in a metropolitan or city community (*P*=.004).

Table 2. Sociodemographic characteristics.

Characteristic	Wave 1 (N=1187), n (%)	Wave 2 (N=337), n (%)
Member of a support group on social media	265 (22.3)	132 (39.2)
Number of social messaging apps used		
0	43 (3.6)	31 (9.2)
1-2	427 (36.0)	113 (33.5)
3-4	454 (45.9)	147 (43.9)
≥5	172 (14.5)	46 (13.4)
Age group (years)		
18-29	314 (26.5)	120 (35.6)
30-39	284 (23.9)	92 (27.3)
40-49	303 (25.5)	64 (19.0)
50-59	161 (13.6)	23 (6.8)
≥60	125 (10.5)	38 (11.3)
Gender		
Male	340 (28.7)	91 (27.0)
Female	831 (70.0)	242 (71.8)
Nonbinary	8 (0.7)	2 (0.5)
Choose not to answer	8 (0.7)	2 (0.7)
Education level		
Primary or less than high school	58 (4.9)	11 (3.3)
High school	177 (14.9)	23 (6.8)
College diploma/some college or university	N/A[a]	107 (31.8)
Bachelor’s degree/professional degree	321 (27.0)	85 (25.2)
Master’s degree	416 (35.0)	62 (18.4)
PhD	215 (18.1)	49 (14.5)
Marital status		
Having a partner/married	628 (52.9)	194 (57.6)
Divorced/separated	82 (6.9)	12 (3.6)
Widowed	34 (37.3)	7 (2.1)
Single	443 (2.9)	121 (35.9)
Prefer not to say	0 (0)	3 (0.9)
Employment status		
Employed	844 (71.1)	264 (78.3)
Retired	58 (4.9)	30 (8.9)
Not employed (out of a job or due to other reasons)	285 (24.0)	43 (12.8)
Type of community (residence)		
Metropolitan/city	608 (51.2)	74 (22.0)
Suburban	233 (19.6)	117 (34.7)
Small town	188 (15.8)	113 (33.5)
Rural area	158 (13.3)	33 (9.8)
Number of people living in the home environment		
1	182 (15.3)	41 (12.2)
2	416 (35.1)	137 (40.7)

Characteristic	Wave 1 (N=1187), n (%)	Wave 2 (N=337), n (%)
3	222 (18.7)	58 (17.2)
4	238 (20.1)	49 (14.5)
≥5	129 (10.9)	52 (15.4)
Region		
Europe	821 (69.2)	N/A
North America	124 (10.5)	337 (100.0)
Asia, Middle East, or South America	242 (20.4)	N/A

[a]N/A: not applicable.

Table 3. Impact of technology use on loneliness scores.

Variable	Wave 1		Wave 2	
	Loneliness score, mean (SD)	*P* value	Loneliness score, mean (SD)	*P* value
Member of an online community support group		.49		.29
Yes	43.88 (5.65)		48.87 (8.86)	
No	44.18 (6.42)		49.96 (9.96)	
Use of the internet to stay connected with friends, family, or peers		.50		.25
Yes	48.46 (6.97)		49.78 (6.91)	
No	48.27 (7.30)		48.21 (8.57)	
Use of social media platforms		.21		.18
Yes	48.47 (6.97)		49.66 (9.35)	
No	48.02 (7.44)		46.20 (11.7)	
Use of social networking messengers		.23		<.001
Yes	48.41 (6.98)		50.15 (9.28)	
No	49.47 (7.23)		43.89 (9.36)	

Table 4. Impact of the type of community on psychological well-being in Wave 2.

Community comparison	Mean difference	95% CI	*P* value
Metro/city vs rural	4.73	−0.96 to 10.43	.14
Metro/city vs small town	−2.71	−6.78 to 1.34	.31
Metro/city vs suburban	−2.10	−10.43 to 0.96	.53
Small town vs rural	7.45	2.05 to 12.85	.002
Suburban vs rural	6.83	1.49 to 12.18	.006
Small town vs suburban	0.62	−2.93 to 4.18	.97

Table 5. Wave 1 ordinary least squares regression of sociodemographic characteristics and use of technology regarding loneliness scores.

Variable	Coefficient[a] (SE)
Age group (reference: 18-29 years)	
30-39 years	0.170 (0.61)
40-49 years	0.722 (0.62)
50-59 years	0.459 (0.71)
≥60 years	−0.378 (0.87)
Gender (reference: female)	
Male or nonbinary/refused to answer	−0.045 (0.39)
Education level (reference: primary or less than high school)	
High school	−2.756 (0.94)[b]
Bachelor's degree	−3.622 (0.90)[c]
Master's/professional degree	−3.981 (0.88)[c]
PhD	−4.073 (0.95)[c]
Marital status (reference: having a partner)	
Divorced/separated/widowed	0.568 (0.67)
Single	2.441 (0.48)[c]
Employment status (reference: working)	
Retired	−0.883 (1.01)
Not working (out of a job or due to other reasons)	1.045 (0.49)[d]
Residence (reference: metropolitan/city)	
Rural	−0.025 (0.58)
Small town	0.959 (0.53)
Suburban	0.564 (0.48)
Number of people living in the home environment (reference: 1)	
2	0.094 (0.61)
3	−0.674 (0.67)
4	−0.368 (0.68)
≥5	−0.223 (0.79)
Number of social media apps used (reference: 0)	
1-2	−0.549 (0.98)
3-4	−0.852 (0.98)
≥5	−0.799 (1.06)
Joining a specific online COVID-19 community support group on social media (reference: no)	
Yes	−0.44 (0.43)

[a]Adjusted R^2=0.059.

[b]P<.01.

[c]P<.001.

[d]P<.05.

Table 6. Wave 2 ordinary least squares regression of sociodemographic characteristics and use of technology regarding loneliness scores.

Variable	Coefficient[a] (SE)
Age group (reference: 18-29 years)	
30-39 years	1.13 (1.52)
40-49 years	−2.89 (1.83)[b]
≥50 years	−3.67 (2.15)[b]
Gender (reference: female)	
Male	2.10 (1.24)[b]
Prefer not to say	0.57 (7.39)
Nonbinary, gender fluid	4.53 (5.60)[b]
Education level (reference: primary or less than high school)	
High school	−0.04 (3.55)
Bachelor's degree	3.80 (3.13)
College diploma/some college	1.64 (3.09)
Master's/professional degree	2.24 (3.22)
PhD	4.72 (3.27)
Marital status (reference: having a partner)	
Divorced/separated	1.59 (2.87)
Widowed	−0.44 (4.03)
Single	2.07 (1.53)
Prefer not to say	3.61 (5.99)
Employment status (reference: working)	
Retired	4.42 (2.37)[b]
Not working (out of a job or other reasons)	3.90 (1.65)[b]
Residence (reference: metropolitan/city)	
Rural	−1.18 (2.07)
Small town	0.64 (1.49)
Suburban	−2.90 (1.47)
Number of people living in the home environment (reference: 1)	
2	−1.84 (1.89)
3	−1.52 (2.08)
4	−1.03 (2.26)
≥5	−0.62 (2.29)
Number of social media apps used (reference: 0)	
1-2	5.95 (1.98)[c]
3-4	5.96 (1.97)[c]
≥5	4.97 (2.30)[b]
Joining a specific online COVID-19 community support group on social media (reference: no)	
Yes	−1.02 (1.17)

[a]Adjusted R^2=0.081.

[b]P<.05.

[c]P<.01.

Table 7. Ordinary least squares regression of sociodemographic characteristics and use of technology regarding psychological well-being.

Variable	Wave 1, coefficient[a] (SE)	Wave 2, coefficient[b] (SE)
Age group (reference: 18-29 years)		
30-39 years	−0.72 (1.00)	4.85 (1.62)[c]
40-49 years	−0.02 (1.03)	4.06 (1.96)[d]
50-59 years	−1.63 (1.17)	—[e]
≥60 years	−0.29 (1.44)	—
≥50 years	—	4.04 (2.30)
Gender (reference: female)		
Male or nonbinary/refused to answer	0.26 (0.65)	−3.67 (1.31)[c]
Prefer not to say	—	1.67 (8.03)
Nonbinary, gender fluid	—	−5.75 (7.47)
Education level (reference: primary or less than high school)		
High school	−1.77 (1.55)	−6.17 (3.69)
Bachelor's degree	−2.77 (1.48)	−1.75 (3.28)
College diploma/some college	—	−2.99 (3.20)
Master's/professional degree	−1.38 (1.46)	−0.66 (3.36)
PhD	−2.20 (1.57)	−2.28 (3.43)
Marital status (reference: having a partner)		
Divorced/separated/widowed	−0.28 (1.12)	−0.44 (2.91)
Widowed	—	5.68 (3.99)
Single	−1.77 (0.80)[d]	1.88 (1.62)
Prefer not to say	—	2.91 (6.51)
Employment status (reference: working)		
Retired	0.75 (1.68)	−1.02 (2.51)
Not working (out of a job or due to other reasons)	−0.78 (0.82)	−3.48 (1.75)[d]
Residence (reference: metropolitan/city)		
Rural	1.05 (0.96)	−3.19 (2.20)
Small town	−1.97 (0.88)[d]	4.60 (1.57)[c]
Suburban	−0.35 (0.80)	2.82 (1.53)
Number of people living in the home environment (reference: 1)		
2	−0.29 (1.01)	3.89 (1.95)[d]
3	1.51 (1.10)	6.55 (2.16)[c]
4	−1.92 (1.13)	3.90 (2.34)
≥5	−2.29 (1.31)	5.88 (2.39)[d]
Number of social media apps used (reference: 0)		
1-2	0.53 (1.64)	2.56 (2.14)
3-4	0.65 (1.64)	1.60 (2.12)
≥5	2.2 (1.76)	3.93 (2.45)
Joining a specific online COVID-19 community support group on social media (reference: no)		
Yes	−0.39 (0.72)	−7.31 (1.23)[f]

[a]Adjusted R^2=0.021.
[b]Adjusted R^2=0.190.
[c]P<.01.
[d]P<.05.
[e]Category not present.
[f]P<.001.

Discussion

Principal Findings

This paper explored the relationship between technology use and loneliness during the COVID-19 pandemic. We observed some associations between social network messaging but only during the second wave of data collection in 2021 from respondents in North America. Loneliness scores were higher among respondents who were using social network messaging apps than among respondents who were not using such apps. We did not observe such associations during the initial first wave of data collection in 2020 across 13 other countries. Additional findings showed that gender and age of the respondents influenced loneliness scores in both waves of data collection. In the second wave, feelings of loneliness were higher among males than females and were higher among respondents aged 30-39 years than among those in older age groups.

Comparison With Prior Work

Our findings align with the findings of other studies that social media communication and internet usage increased loneliness during the pandemic [17,18,62-64]. Individuals who were lonely tended to use social media and the internet more, which has been associated with poorer mental health and increased prevalence of anxiety and depression. Such findings call for actions and standards from schools, workplaces, and governments regarding the use and misuse of social media and the internet. A "one-size-fits-all" approach may not benefit everyone. Exploring and identifying different options should be considered to stay (remotely) connected without a negative impact on mental health.

Similar to the findings of prior studies, we observed that being a member of community support groups or being a part of a group activity had a positive impact on loneliness [65]. Additionally, the UN [15] has outlined challenges and specificities in a bid to recover regions and impacted areas identified during the pandemic. From a socioeconomic perspective, the UN response includes the following 5-point framework [15]: (1) *ensuring all essential health services are still available and protecting health systems*; (2) *helping people cope with adversity, through social protection and basic services*; (3) *protecting jobs and supporting small and medium-sized enterprises and informal sector workers through economic response and recovery programs*; (4) *guiding the necessary surge in fiscal and financial stimulus to make macroeconomic policies work for the most vulnerable and strengthening multilateral and regional responses*; and (5) *promoting social cohesion and investing in community-led resilience and response systems*. These 5 streams are connected by a strong environmental sustainability and gender equality imperative to build back better [15].

These findings further support the observation that it is not necessarily the use of the internet or social media that influences loneliness but the involvement in online communities and activities that can influence loneliness. Building on the existing reports of the UN, WHO, and Pan American Health Organization surrounding digital (eHealth) transformation [66-69] and appropriate strategies for preparing and responding to influenza pandemics, future research may investigate the direction of this relationship and the impact of social cohesion with a view to improving community and societal resilience. Yet, we identified that single respondents and those with lesser formal education were lonelier than their counterparts across both waves of data collection. Additionally, we observed that males, younger respondents (aged 18-29 years), and unemployed respondents were comparatively lonelier. These findings align with the findings of prior studies that investigated the factors associated with loneliness during the COVID-19 pandemic [8,44,70-72], although there are exceptions pertaining to the observations associated with gender. Although the UN [15] acknowledged how industries used digital transformation throughout 2020 to enable employees to continue working from home, there are still areas that were affected, resulting in an increase in unemployment and the number of hours lost. With regard to unemployment, the UN noted that it has led to a greater impact on the health and well-being of individuals and families.

Data from the first wave of collection showed that the levels of psychological well-being were lower among respondents living in small towns than among those living in metropolitan areas. However, the data collected in the second wave showed that the levels of psychological well-being were lower among respondents living in rural areas than among those living in small towns or suburban settings. Moreover, the UN [15] has mentioned how the pandemic has revealed inequalities, resulting in many challenges for service provision and frontline staff in delivering health care specifically in urban areas. According to the UN [15], these challenges are varied and include health care access, inadequate housing, poor infrastructure (eg, transport, water, and sanitation), and employment precarity. However, the UN has mentioned that cities or metropolitan areas are perceived as "hubs of resilience and human ingenuity and this crisis has shown how city dwellers can adapt overnight to new ways of working and functioning while demonstrating extraordinary solidarity and support for one another." Further, in the second wave data set, the levels of well-being were lower among respondents living in metropolitan areas or cities than among those living in suburban areas, and those living in small towns were lonelier than those living in suburban areas.

Collectively, these findings align with the findings of some previous studies that identified how small towns and rural areas may provide fewer options than cities for in-person activities, social engagements, health care access opportunities, and other

programs that influence mental health [73-79]. Still, other studies have observed how psychological distress is higher among people living in urban areas than among those living in rural areas. It appears that the data presented in this paper align with the narrative described by the UN [15], and this situation may be explained not by residing in urban areas per se but by the lack of access to outside spaces and environmental amenities (ie, green spaces) that leads to psychological distress [80,81]. The data related to respondents residing in cities and metropolitan areas indicate a greater sense of community and resilience. Moreover, analyses of Canadian data have identified age and gender differences across varying community types, use of social media, and loneliness [82].

Contribution and Implications

This international multicenter study was launched as a rapid response to the WHO declaring a pandemic [1] and presents a snapshot of the impact technology has on people based on the type of community they live in, their age, and their marital status during the pandemic. The consortium's goal was to provide scholars, policymakers, educationalists, and historians (in the future) an opportunity to garner a greater understanding of societal behavior and technology use during a specific timeframe. Specifically, we aimed to enhance societal understanding of the role technology plays within and across our societies when it comes to addressing environmental aging, loneliness, and isolation, which may facilitate appropriate planning for future scenarios and crises.

Limitations

One limitation of this work is the difficulty in presenting a cross-national perspective, given the global spread of the COVID-19 pandemic and the locations of the consortium members. The rates of survey completion varied considerably across study sites. Given the nature of survey deployment, consortium partners used their own existing mailing lists, networks, and social media platforms, and with this, the English version of the survey may have reached people who were not necessarily located in the United Kingdom but abroad instead. The design of this project only provides a snapshot of experiences shared by people who have access to technology and internet services. Future research should consider collecting more representative data.

Conclusions

Our multisite international study showed a contrary trend whereby respondents from 2 countries in 2021, who installed more social media apps on their mobile devices, experienced greater feelings of loneliness. However, these trends did not extend equally across all countries where data were collected in 2020. Additionally, we observed how some sociodemographic factors, specifically age, gender, marital status, and type of community, were associated with loneliness and psychological well-being during the pandemic. Access and use of digital technologies, the internet, and social media have increased since the beginning of the COVID-19 pandemic within developed and developing countries. Although some increases can be attributed to trends in working and studying from home, the increased use of social media platforms and messengers has also been tied to desires to stay connected with friends, families, and peers to avoid loneliness. Future studies and discourse should start to consider the role and access of transgenerational technology [83] to explore and understand the sociodemographic factors when implementing programs and activities to reduce loneliness and improve well-being across the rural and urban spectrum.

Acknowledgments

We would like to thank all participants who completed the surveys across the different sites during 2020 and 2021. This research received no external funding, and the article publishing charge was funded by a fee waiver.

Authors' Contributions

Conceptualization: HRM and SE; methodology: HRM; formal analysis: VGP and PCK; investigation: HRM; resources: HRM and SE; data curation: HRM, IS, MHEB, PCK, LI, SCB, RK, PS, VGP, FG, HOC, HA, BBK, SF, CR, and GS; writing-original draft preparation: HRM, IS, VGP, PCK, SF, and CR; writing-review and editing: HRM, SE, SCB, RK, PAS, SF, MHEB, IS, and CR; supervision: HRM, SE, IS, and MHEB; project administration: HRM, IS, LI, and CR. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online survey in Wave 1.
[PDF File (Adobe PDF File), 238 KB - mental_v10i1e41304_app1.pdf]

References

1. Khanna RC, Cicinelli MV, Gilbert SS, Honavar SG, Murthy GSV. COVID-19 pandemic: Lessons learned and future directions. Indian J Ophthalmol 2020;68(5):703-710 [FREE Full text] [doi: 10.4103/ijo.ijo_843_20]

2. Freeman S, Marston HR, Olynick J, Musselwhite C, Kulczycki C, Genoe R, et al. Intergenerational Effects on the Impacts of Technology Use in Later Life: Insights from an International, Multi-Site Study. Int J Environ Res Public Health 2020 Aug 07;17(16):5711 [FREE Full text] [doi: 10.3390/ijerph17165711] [Medline: 32784651]
3. Genoe R, Kulczycki C, Marston H, Freeman S, Musselwhite C, Rutherford H. E-Leisure and Older Adults: Findings from an International Exploratory Study. Therapeutic Recreation Journal 2018;52(1):1-18 [FREE Full text] [doi: 10.18666/TRJ-2018-V52-I1-8417]
4. Haddon L, Loos E, Mante-Meijer E. Generational Use of New Media. Abingdon, UK: Routledge; 2016.
5. Shah SGS, Nogueras D, van Woerden HC, Kiparoglou V. The COVID-19 Pandemic: A Pandemic of Lockdown Loneliness and the Role of Digital Technology. J Med Internet Res 2020 Nov 05;22(11):e22287 [FREE Full text] [doi: 10.2196/22287] [Medline: 33108313]
6. Burholt V, Windle G, Gott M, Morgan DJ. Technology-Mediated Communication in Familial Relationships: Moderated-Mediation Models of Isolation and Loneliness. Gerontologist 2020;60(7):1202-1212 [FREE Full text] [doi: 10.1093/geront/gnaa040] [Medline: 32369112]
7. Vargo D, Zhu L, Benwell B, Yan Z. Digital technology use during COVID-19 pandemic: A rapid review. Human Behav and Emerg Tech 2020;3(1):13-24 [FREE Full text] [doi: 10.1002/hbe2.242]
8. Groarke JM, Berry E, Graham-Wisener L, McKenna-Plumley PE, McGlinchey E, Armour C. Loneliness in the UK during the COVID-19 pandemic: Cross-sectional results from the COVID-19 Psychological Wellbeing Study. PLoS One 2020 Sep 24;15(9):e0239698 [FREE Full text] [doi: 10.1371/journal.pone.0239698] [Medline: 32970764]
9. Garfin DR. Technology as a coping tool during the coronavirus disease 2019 (COVID-19) pandemic: Implications and recommendations. Stress Health 2020;36(4):555-559 [FREE Full text] [doi: 10.1002/smi.2975] [Medline: 32762116]
10. Gauthier A, Lagarde C, Mourey F, Manckoundia P. Use of Digital Tools, Social Isolation, and Lockdown in People 80 Years and Older Living at Home. Int J Environ Res Public Health 2022;19(5):2908 [FREE Full text] [doi: 10.3390/ijerph19052908] [Medline: 35270600]
11. Sen K, Prybutok G, Prybutok V. The use of digital technology for social wellbeing reduces social isolation in older adults: A systematic review. SSM Popul Health 2022;17:101020 [FREE Full text] [doi: 10.1016/j.ssmph.2021.101020] [Medline: 35024424]
12. Williams SN, Armitage CJ, Tampe T, Dienes K. Public perceptions and experiences of social distancing and social isolation during the COVID-19 pandemic: a UK-based focus group study. BMJ Open 2020 Jul 20;10(7):e039334 [FREE Full text] [doi: 10.1136/bmjopen-2020-039334] [Medline: 32690752]
13. Ganesan B, Al-Jumaily A, Fong KNK, Prasad P, Meena SK, Tong RKY. Impact of Coronavirus Disease 2019 (COVID-19) Outbreak Quarantine, Isolation, and Lockdown Policies on Mental Health and Suicide. Front Psychiatry 2021 Apr 16;12:565190 [FREE Full text] [doi: 10.3389/fpsyt.2021.565190] [Medline: 33935817]
14. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr 27;26:459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
15. United Nations Comprehensive Response to COVID-19: Saving Lives, Protecting Societies, Recovering Better. United Nations. 2020. URL: https://unsdg.un.org/resources/united-nations-comprehensive-response-covid-19-saving-lives-protecting-societies-0 [accessed 2023-02-17]
16. Fisk M, Livingstone A, Pit SW. Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]
17. Marston HR, Kowert R. What role can videogames play in the COVID-19 pandemic? [version 2; peer review: 2 approved]. Emerald Open Res 2020;2:34 [FREE Full text] [doi: 10.35241/emeraldopenres.13727.2]
18. Marston HR, Morgan DJ. International Psychogeriatric Association. URL: https://oro.open.ac.uk/70814/ [accessed 2023-02-17]
19. Schlomann A, Seifert A, Zank S, Woopen C, Rietz C. Use of Information and Communication Technology (ICT) Devices Among the Oldest-Old: Loneliness, Anomie, and Autonomy. Innov Aging 2020;4(2):igz050 [FREE Full text] [doi: 10.1093/geroni/igz050] [Medline: 31911953]
20. Nguyen MH, Gruber J, Fuchs J, Marler W, Hunsaker A, Hargittai E. Changes in Digital Communication During the COVID-19 Global Pandemic: Implications for Digital Inequality and Future Research. Soc Media Soc 2020 Jul 09;6(3):2056305120948255 [FREE Full text] [doi: 10.1177/2056305120948255] [Medline: 34192039]
21. Hwang T, Rabheru K, Peisah C, Reichman W, Ikeda M. Loneliness and social isolation during the COVID-19 pandemic. Int Psychogeriatr 2020;32(10):1217-1220 [FREE Full text] [doi: 10.1017/S1041610220000988] [Medline: 32450943]
22. Son JS, Nimrod G, West ST, Janke MC, Liechty T, Naar JJ. Promoting Older Adults' Physical Activity and Social Well-Being during COVID-19. Leisure Sciences 2021;43(1-2):287-294 [FREE Full text] [doi: 10.1080/01490400.2020.1774015]
23. Pennington N. Communication outside of the home through social media during COVID-19. Comput Hum Behav Rep 2021;4:100118 [FREE Full text] [doi: 10.1016/j.chbr.2021.100118] [Medline: 34568638]
24. Marston HR, Genoe R, Freeman S, Kulczycki C, Musselwhite C. Older Adults' Perceptions of ICT: Main Findings from the Technology In Later Life (TILL) Study. Healthcare (Basel) 2019;7(3):86 [FREE Full text] [doi: 10.3390/healthcare7030086] [Medline: 31277387]

25. Marston HR, Ivan L, Fernández-Ardèvol M, Rosales Climent A, Gómez-León M, Blanche- TD, et al. COVID-19: Technology, Social Connections, Loneliness, and Leisure Activities: An International Study Protocol. Front Sociol 2020;5:574811 [FREE Full text] [doi: 10.3389/fsoc.2020.574811] [Medline: 33869500]
26. COVID-19: Technology, social connections, loneliness and leisure activities. The Open University. 2020. URL: https://www.open.ac.uk/health-wellbeing/covid-19/technology-social-connections-loneliness-leisure-activities [accessed 2023-02-17]
27. Ayhan H. Non-probability Sampling Survey Methods. In: Lovric M, editor. International Encyclopedia of Statistical Science. Berlin, Heidelberg: Springer; 2011:979-982.
28. Benfield JA, Szlemko WJ. Internet-Based Data Collection: Promises and Realities. Journal of Research Practice 2006;2(2):Article D1 [FREE Full text]
29. Marston HR. Millennials and ICT—Findings from the Technology 4 Young Adults (T4YA) Project: An Exploratory Study. Societies 2019;9(4):80 [FREE Full text] [doi: 10.3390/soc9040080]
30. Marston HR. Older adults as 21st century game designers. Comput Game J 2012 May 15;1(1):90-102 [FREE Full text] [doi: 10.1007/bf03392330]
31. Marston HR, Kroll M, Fink D, de Rosario H, Gschwind YJ. Technology use, adoption and behavior in older adults: Results from the iStoppFalls project. Educational Gerontology 2016 Jan 08;42(6):371-387 [FREE Full text] [doi: 10.1080/03601277.2015.1125178]
32. Ryff CD, Keyes CL. The structure of psychological well-being revisited. Journal of Personality and Social Psychology 1995;69(4):719-727 [FREE Full text] [doi: 10.1037/0022-3514.69.4.719]
33. Ryff CD, Singer B. The Contours of Positive Human Health. Psychological Inquiry 1998;9(1):1-28 [FREE Full text] [doi: 10.1207/s15327965pli0901_1]
34. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
35. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996;66(1):20-40 [FREE Full text] [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
36. Lee E, Depp C, Palmer B, Glorioso D, Daly R, Liu J, et al. High prevalence and adverse health effects of loneliness in community-dwelling adults across the lifespan: role of wisdom as a protective factor. Int. Psychogeriatr 2018;31(10):1447-1462 [FREE Full text] [doi: 10.1017/s1041610218002120]
37. Jeste DV, Di Somma S, Lee EE, Nguyen TT, Scalcione M, Biaggi A, et al. Study of loneliness and wisdom in 482 middle-aged and oldest-old adults: a comparison between people in Cilento, Italy and San Diego, USA. Aging Ment Health 2021;25(11):2149-2159 [FREE Full text] [doi: 10.1080/13607863.2020.1821170] [Medline: 33000647]
38. Tomstad S, Dale B, Sundsli K, Saevareid HI, Söderhamn U. Who often feels lonely? A cross-sectional study about loneliness and its related factors among older home-dwelling people. Int J Older People Nurs 2017;12(4):e12162 [FREE Full text] [doi: 10.1111/opn.12162] [Medline: 28752653]
39. Sheerman L, Marston HR, Musselwhite C, Morgan D. COVID-19 and the secret virtual assistants: the social weapons for a state of emergency [version 1; peer review: 2 approved, 1 not approved]. Emerald Open Res 2020;2:19 [FREE Full text] [doi: 10.35241/emeraldopenres.13571.1]
40. Education in the United States. Wikipedia. URL: https://en.wikipedia.org/wiki/Education_in_the_United_States [accessed 2022-06-05]
41. List of ethnic groups. UK Government. URL: https://www.ethnicity-facts-figures.service.gov.uk/ethnic-groups [accessed 2022-06-05]
42. Pollak VM, Kowarz N, Partheymüller J. Chronology of the Corona Crisis in Austria - Part 1: Background, the way to the lockdown, the acute phase and economic consequences. University of Vienna. 2020. URL: https://viecer.univie.ac.at/en/projects-and-cooperations/austrian-corona-panel-project/corona-blog/corona-blog-beitraege/blog51/ [accessed 2023-02-17]
43. Pollak M, Kowarz N, Partheymüller J. Chronology of the Corona Crisis in Austria - Part 3: A calm summer and the beginning of the second wave. University of Vienna. 2020. URL: https://viecer.univie.ac.at/en/projects-and-cooperations/austrian-corona-panel-project/corona-blog/corona-blog-beitraege/blog79/ [accessed 2023-02-17]
44. COVID-19 in Canada: A One-year Update on Social and Economic Impacts. Statistics Canada. 2021. URL: https://www150.statcan.gc.ca/n1/pub/11-631-x/11-631-x2021001-eng.htm [accessed 2022-12-13]
45. Urrutia D, Manetti E, Williamson M, Lequy E. Overview of Canada's Answer to the COVID-19 Pandemic's First Wave (January-April 2020). Int J Environ Res Public Health 2021;18(13):7131 [FREE Full text] [doi: 10.3390/ijerph18137131] [Medline: 34281075]
46. Long JA, Malekzadeh M, Klar B, Martin G. Do regionally targeted lockdowns alter movement to non-lockdown regions? Evidence from Ontario, Canada. Health Place 2023;79:102668 [FREE Full text] [doi: 10.1016/j.healthplace.2021.102668] [Medline: 34548221]
47. Bosen R, Thuran J. Chronology: How COVID has spread in Germany. DW. 2021. URL: https://www.dw.com/en/covid-how-germany-battles-the-pandemic-a-chronology/a-58026877 [accessed 2023-02-17]
48. COVID-19 pandemic in Portugal. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Portugal [accessed 2022-12-13]

49. COVID-19 pandemic in Spain. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Spain [accessed 2022-12-13]
50. COVID-19 pandemic in Turkey. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Turkey [accessed 2022-12-13]
51. Timeline of UK coronavirus lockdowns, March 2020 to March 2021. Institute for Government. URL: https://www.instituteforgovernment.org.uk/sites/default/files/timeline-lockdown-web.pdf [accessed 2023-02-13]
52. COVID-19 lockdown in the United Kingdom. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_lockdown_in_the_United_Kingdom [accessed 2022-12-13]
53. Timeline for Coronavirus (COVID-19) in Scotland. SPICe Spotlight. URL: https://spice-spotlight.scot/2022/11/25/timeline-of-coronavirus-covid-19-in-scotland/ [accessed 2022-12-11]
54. COVID-19 pandemic in Scotland. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Scotland [accessed 2022-12-13]
55. Coronavirus timeline: Welsh and UK governments' response. Welsh Parliament. URL: https://research.senedd.wales/research-articles/coronavirus-timeline-welsh-and-uk-governments-response/ [accessed 2022-12-11]
56. COVID-19 pandemic in Wales. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Wales [accessed 2022-12-13]
57. COVID-19 pandemic in Northern Ireland. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Northern_Ireland [accessed 2022-12-13]
58. COVID-19 pandemic in France. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_France [accessed 2022-12-13]
59. COVID-19 pandemic in Romania. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Romania [accessed 2022-12-13]
60. COVID-19 pandemic in Germany. Wikipedia. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Germany [accessed 2022-12-13]
61. Marston HR, Shore L, White P. How does a (Smart) Age-Friendly Ecosystem Look in a Post-Pandemic Society? Int J Environ Res Public Health 2020;17(21):8276 [FREE Full text] [doi: 10.3390/ijerph17218276] [Medline: 33182413]
62. Geirdal A, Ruffolo M, Leung J, Thygesen H, Price D, Bonsaksen T, et al. Mental health, quality of life, wellbeing, loneliness and use of social media in a time of social distancing during the COVID-19 outbreak. A cross-country comparative study. J Ment Health 2021;30(2):148-155 [FREE Full text] [doi: 10.1080/09638237.2021.1875413] [Medline: 33689546]
63. Boursier V, Gioia F, Musetti A, Schimmenti A. Facing Loneliness and Anxiety During the COVID-19 Isolation: The Role of Excessive Social Media Use in a Sample of Italian Adults. Front Psychiatry 2020;11:586222 [FREE Full text] [doi: 10.3389/fpsyt.2020.586222] [Medline: 33363484]
64. Bonsaksen T, Ruffolo M, Leung J, Price D, Thygesen H, Schoultz M, et al. Loneliness and Its Association With Social Media Use During the COVID-19 Outbreak. Social Media + Society 2021;7(3):205630512110338 [FREE Full text] [doi: 10.1177/20563051211033821]
65. Lisitsa E, Benjamin KS, Chun SK, Skalisky J, Hammond LE, Mezulis AH. Loneliness Among Young Adults During COVID-19 Pandemic: The Mediational Roles of Social Media Use and Social Support Seeking. Journal of Social and Clinical Psychology 2020;39(8):708-726 [FREE Full text] [doi: 10.1521/jscp.2020.39.8.708]
66. Developing a National eHealth Strategy. Pan American Health Organization. URL: https://iris.paho.org/handle/10665.2/55661 [accessed 2022-12-12]
67. WHO compendium of innovative health technologies for low-resource settings: 2016-2017: medical devices, eHealth/mHealth, medical simulation devices, personal protective equipment, assistive products, other technologies. World Health Organization. URL: https://apps.who.int/iris/handle/10665/274893 [accessed 2023-02-13]
68. Pandemic influenza preparedness and response : a WHO guidance document. World Health Organization. URL: https://apps.who.int/iris/handle/10665/44123 [accessed 2023-02-13]
69. Progress report on utilizing eHealth solutions to improve national health systems in the African Region: information document. World Health Organization. URL: https://apps.who.int/iris/handle/10665/334099 [accessed 2023-02-13]
70. Marmarosh CL, Forsyth DR, Strauss B, Burlingame GM. The psychology of the COVID-19 pandemic: A group-level perspective. Group Dynamics: Theory, Research, and Practice 2020;24(3):122-138 [FREE Full text] [doi: 10.1037/gdn0000142]
71. Hoffart A, Johnson SU, Ebrahimi OV. Loneliness and Social Distancing During the COVID-19 Pandemic: Risk Factors and Associations With Psychopathology. Front Psychiatry 2020;11:589127 [FREE Full text] [doi: 10.3389/fpsyt.2020.589127] [Medline: 33329136]
72. Hansen T, Nilsen TS, Yu B, Knapstad M, Skogen JC, Vedaa Ø, et al. Locked and lonely? A longitudinal assessment of loneliness before and during the COVID-19 pandemic in Norway. Scand J Public Health 2021;49(7):766-773 [FREE Full text] [doi: 10.1177/1403494821993711] [Medline: 33645336]
73. O'Connor RC, Wetherall K, Cleare S, McClelland H, Melson AJ, Niedzwiedz CL, et al. Br J Psychiatry 2021;218(6):326-333 [FREE Full text] [doi: 10.1192/bjp.2020.212] [Medline: 33081860]

74. Bu F, Steptoe A, Fancourt D. Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. Public Health 2020 Sep;186:31-34 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.036] [Medline: 32768621]
75. Mueller JT, McConnell K, Burow PB, Pofahl K, Merdjanoff AA, Farrell J. Impacts of the COVID-19 pandemic on rural America. Proc Natl Acad Sci U S A 2021;118(1):2019378118 [FREE Full text] [doi: 10.1073/pnas.2019378118] [Medline: 33328335]
76. Henning-Smith C. The Unique Impact of COVID-19 on Older Adults in Rural Areas. J Aging Soc Policy 2020;32(4-5):396-402 [FREE Full text] [doi: 10.1080/08959420.2020.1770036] [Medline: 32475255]
77. Wang Y, Kala MP, Jafar TH. Factors associated with psychological distress during the coronavirus disease 2019 (COVID-19) pandemic on the predominantly general population: A systematic review and meta-analysis. PLoS One 2020;15(12):e0244630 [FREE Full text] [doi: 10.1371/journal.pone.0244630] [Medline: 33370404]
78. Isaac V, Cheng T, Townsin L, Assareh H, Li A, McLachlan CS. Associations of the Initial COVID-19 Lockdown on Self-Reported Happiness and Worry about Developing Loneliness: A Cross-Sectional Analysis of Rural, Regional, and Urban Australian Communities. Int J Environ Res Public Health 2021;18(18):9501 [FREE Full text] [doi: 10.3390/ijerph18189501] [Medline: 34574425]
79. Hubbard G, den Daas C, Johnston M, Murchie P, Thompson CW, Dixon D. Are Rurality, Area Deprivation, Access to Outside Space, and Green Space Associated with Mental Health during the COVID-19 Pandemic? A Cross Sectional Study (CHARIS-E). Int J Environ Res Public Health 2021;18(8):3869 [FREE Full text] [doi: 10.3390/ijerph18083869] [Medline: 33917067]
80. Grima N, Corcoran W, Hill-James C, Langton B, Sommer H, Fisher B. The importance of urban natural areas and urban ecosystem services during the COVID-19 pandemic. PLoS One 2020;15(12):e0243344 [FREE Full text] [doi: 10.1371/journal.pone.0243344] [Medline: 33332364]
81. Noël C, Rodriguez-Loureiro L, Vanroelen C, Gadeyne S. Perceived Health Impact and Usage of Public Green Spaces in Brussels' Metropolitan Area During the COVID-19 Epidemic. Front. Sustain. Cities 2021;3:668443 [FREE Full text] [doi: 10.3389/frsc.2021.668443]
82. Chidiac M, Ross C, Marston HR, Freeman S. Age and Gender Perspectives on Social Media and Technology Practices during the COVID-19 Pandemic. Int J Environ Res Public Health 2022;19(21):13969 [FREE Full text] [doi: 10.3390/ijerph192113969] [Medline: 36360853]
83. Marston HR, Shore L, Stoops L, Turner RS. Transgenerational Technology and Interactions for the 21st Century: Perspectives and Narratives. Bingley, United Kingdom: Emerald Publishing; 2022.

Abbreviations

OLS: ordinary least squares
UN: United Nations
WHO: World Health Organization

Edited by J Torous; submitted 21.07.22; peer-reviewed by Y Quintana, J Torous; comments to author 09.12.22; revised version received 25.01.23; accepted 26.01.23; published 06.03.23.

Please cite as:
Marston HR, Ko PC, Girishan Prabhu V, Freeman S, Ross C, Sharaievska I, Browning MHEM, Earle S, Ivan L, Kanozia R, Öztürk Çalıkoğlu H, Arslan H, Bilir-Koca B, Alexandra Silva P, Buttigieg SC, Großschädl F, Schüttengruber G
Digital Practices by Citizens During the COVID-19 Pandemic: Findings From an International Multisite Study
JMIR Ment Health 2023;10:e41304
URL: https://mental.jmir.org/2023/1/e41304
doi:10.2196/41304
PMID:36877558

Original Paper

Evaluation of an App-Delivered Psychological Flexibility Skill Training Intervention for Medical Student Burnout and Well-being: Randomized Controlled Trial

Elizabeth Ditton[1,2,3], BA (Psych) Hons, MA Psych (Clin); Brendon Knott[4], BSc Psych (Hons), MA Psych (Clin); Nicolette Hodyl[1,2,5], BSc Hons, GradCert Biostats, PhD; Graeme Horton[3], BS, MB, MEnvStud, PhD; Christopher Oldmeadow[2,3], BMaths (Hons), PhD; Frederick Rohan Walker[1,2,3*], BA Psych (Hons), PhD; Michael Nilsson[1,2,3,6*], MD, PhD

[1]Centre for Rehab Innovations, University of Newcastle, Callaghan, Australia

[2]Hunter Medical Research Institute, New Lambton Heights, Australia

[3]College of Health, Medicine and Wellbeing, University of Newcastle, Callaghan, Australia

[4]Contextual Interventions, Newcastle, Australia

[5]New South Wales Regional Health Partners, Newcastle, Australia

[6]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Elizabeth Ditton, BA (Psych) Hons, MA Psych (Clin)
Centre for Rehab Innovations
University of Newcastle
University Drive
Callaghan, 2308
Australia
Phone: 61 2 404 ext 20738
Email: elizabeth.ditton@uon.edu.au

Abstract

Background: Physician burnout is a common problem, with onset frequently occurring during undergraduate education. Early intervention strategies that train medical students in psychological flexibility skills could support well-being and mitigate burnout risks associated with unmodifiable career stressors. There is a need for randomized controlled trials to assess effectiveness. As psychological flexibility varies contextually and among individuals, tailoring interventions may improve outcomes. Smartphone apps can facilitate individualization and accessibility, and the evaluation of this approach is an identified research priority.

Objective: This study aimed to evaluate the effectiveness of a stand-alone app–delivered Acceptance and Commitment Training intervention for improving medical students' self-reported burnout, well-being, psychological flexibility, and psychological distress outcomes. We aimed to explore whether an individualized app would demonstrate benefits over a nonindividualized version.

Methods: This parallel randomized controlled trial was conducted with a sample of medical students from 2 Australian universities (N=143). Participants were randomly allocated to 1 of 3 intervention arms (individualized, nonindividualized, and waitlist) using a 1:1:1 allocation ratio. Individualized and nonindividualized participants were blinded to group allocation. The 5-week intervention included an introductory module (stage 1) and on-demand access to short skill training activities (stage 2), which students accessed at their own pace. Stage 2 was either nonindividualized or individualized to meet students' identified psychological flexibility training needs.

Results: The mean differences in change from baseline between the intervention groups and the waitlist group were not statistically significant for burnout outcomes: exhaustion (primary; individualized: −0.52, 95% CI −3.70 to 2.65, *P*=.75; nonindividualized: 1.60, 95% CI −1.84 to 5.03, *P*=.37), cynicism (individualized: −1.26, 95% CI −4.46 to 1.94, *P*=.44; nonindividualized: 1.00, 95% CI −2.45 to 4.46, *P*=.57), and academic efficacy (individualized: 0.94, 95% CI −0.90 to 2.79, *P*=.32; nonindividualized: 2.02, 95% CI 0.02-4.03, *P*=.05). Following the intervention, the individualized group demonstrated improved psychological flexibility (0.50, 95% CI 0.12-0.89; *P*=.01), reduced inflexibility (0.48, 95% CI −0.92 to −0.04; *P*=.04), and reduced

stress (−6.89, 95% CI −12.01 to 5.99; *P*=.01), and the nonindividualized group demonstrated improved well-being (6.46, 95% CI 0.49-12.42; *P*=.04) and stress (−6.36, 95% CI −11.90 to −0.83; *P*=.03) compared with waitlist participants. Between-group differences for the individualized and nonindividualized arms were not statistically significant. High attrition (75/143, 52.4%) was observed.

Conclusions: This trial provides early support for the potential benefits of Acceptance and Commitment Training for medical student well-being and psychological outcomes and demonstrates that psychological flexibility and inflexibility can be trained using a smartphone app. Although postintervention burnout outcomes were not statistically significant, improvements in secondary outcomes could indicate early risk mitigation. Replication studies with larger samples and longer-term follow-up are required, and future research should focus on improving implementation frameworks to increase engagement and optimize individualization methods.

Trial Registration: Australian New Zealand Clinical Trials Registry 12621000911897; https://tinyurl.com/2p92cwrw

International Registered Report Identifier (IRRID): RR2-10.2196/32992

(JMIR Ment Health 2023;10:e42566) doi:10.2196/42566

KEYWORDS

burnout; psychological; burnout interventions; well-being; medicine; medical student; digital intervention; app-delivered intervention; individualized intervention; randomized controlled trial; RCT; randomized; Acceptance and Commitment Training; stress; mobile health; mHealth; mobile phone

Introduction

Burnout and the Medical Profession

The medical profession is facing a burnout crisis [1]. Prevalence estimates among physicians range between 37% [2] and 80% [3], and the enduring challenges associated with the COVID-19 pandemic have exacerbated this existing problem [4,5]. Burnout is a psychological outcome of exposure to work-related stress characterized by varying degrees of exhaustion (feelings of overload and depleted emotional energy), cynicism (detachment, withdrawal behaviors, and diminished idealism), and inefficacy (the perception that one's own work performance lacks quality and value) [6,7]. Although previously believed to emerge during the later stages of a medical career [8], a high global prevalence among medical students (44.2% [9]) is indicative of earlier onset. Increasingly, burnout is arising in response to pervasive imbalances between medical education stressors and a student's coping resources, and this can persist throughout an individual's professional life [10,11]. Individuals affected by burnout are at risk of a range of associated outcomes that extend beyond the work context, including poor physical [6,12] and psychological health (including diminished well-being [13], depression [14,15], anxiety [16], and suicide [14,17]). Furthermore, burnout impedes medical service delivery and productivity through its adverse impact on work engagement (eg, reduced participation) [18] and performance (eg, increased medical errors and diminished quality of patient care [2,15]).

Developing effective strategies to address the problem of physician burnout and its associated outcomes was an identified research priority before the pandemic [1,19], with a particular emphasis on early intervention strategies that could facilitate prevention [1,20-22]. However, progress toward this agenda was disrupted by the emergence of COVID-19, which required the medical profession to rapidly prioritize and respond to the ensuing public health crisis, often under conditions of uncertainty and insufficient resources [4]. The persistent stress of the pandemic has contributed to an increase in rates of physician burnout [23] and declining mental health among medical students, to the extent that many report reconsidering their decision to pursue a medical career [24]. There are growing concerns that the potential endurance of these adverse mental health impacts beyond the pandemic could further diminish the resilience of the medical workforce and health care systems, adding to the urgency of identifying and deploying effective interventions [4].

Adaptive Psychological Skill Training for Medical Students

Organizational interventions that modify external stressors within the work or study environment (eg, inadequate resources, excessive workloads, and time pressures) have demonstrated benefits [25-27] and are essential to burnout prevention strategies [28]. However, the pandemic has highlighted that the modifiability of some of these factors may be limited during extended periods of crisis when health care resources are stretched beyond capacity. Medical students in a recent study reported that pandemic-related stressors contributed to elevated burnout and diminished mental health, some of which were less modifiable under the circumstances (eg, web-based learning fatigue, restricted opportunities for clinical experience, and mandatory isolation) than others (eg, quality of web-based learning) [24,29]. Furthermore, physicians and medical students encounter unmodifiable demands inherent to their training and work, resulting in unavoidable contact with certain risk factors during the normal course of their careers [8] (eg, academic pressures, exposure to death and dying [11,30], and role responsibility [15]).

Psychological and behavioral responses to demands and stressors play a role in burnout development among medical students, whose risk of burnout is almost doubled when maladaptive coping patterns are adopted [30-33]. There is recent evidence suggesting that, in addition to its augmentation of external stressors, the pandemic has adversely affected medical students' psychological resources for coping with these challenges [24]. Maladaptive coping patterns are unlikely to spontaneously

improve over time without intervention [34], which may contribute to the increasing risk of burnout that medical students face as they progress through training and into their careers [8,30,35]. Individual-level interventions have the potential to buffer against burnout and other psychological ill health outcomes by training *modifiable* cognitive, emotional, and behavioral skills that can facilitate adaptive responses to unmodifiable contextual stressors [8,34,36,37]. There is growing support for the implementation of such interventions during undergraduate medical training, which is recognized as a critical stage of learning and career preparation [21,38,39]. Physicians experiencing burnout have proposed that learning self-care should be prioritized equally with clinical skills during medical education [40]. By assisting students in developing adaptive coping repertoires, individual-level interventions can pre-emptively prepare them to respond to the inevitable stressors of medical education and future practice in ways that support their psychological health and well-being [21,38,39]. This early intervention strategy could offer longer-term benefits for burnout prevention within the medical profession [8,30,31,37,38].

Intervention Model: Psychological Flexibility

Despite the considerable increase in the number of studies investigating such interventions in recent years, systematic reviews have highlighted the need for more rigorous studies identifying which adaptive psychological skill sets can be trained to produce optimal improvements in burnout and well-being outcomes among medical students [21,28]. Psychological flexibility is a set of adaptive cognitive and behavioral skills that is a promising intervention target [36,41-43]. The psychological flexibility model encompasses 6 modifiable flexibility (and corresponding *inflexibility*) processes: present-moment awareness (nonawareness of present moment), experiential acceptance (experiential avoidance), cognitive defusion (cognitive fusion), self-as-context (self-as-content), contact with values (lack of contact with values), and committed action toward values (inaction) [44,45]. When faced with challenging situations or uncomfortable internal experiences, individuals who are high in psychological flexibility tend to respond in ways that are effective in the moment and supportive of their broader well-being [46,47], including bringing conscious awareness and openness to the conditions of their present-moment experiences and purposefully committing to effective actions that align with personally held values [44]. Conversely, psychological inflexibility manifests as behavioral rigidity in response to internal (eg, thoughts, emotions, and physical sensations) experiences, which can have an adverse impact on an individual's capacity to adapt and function in psychologically healthy ways [44]. An individual can learn to improve their psychological flexibility skills using a range of intervention approaches, including Acceptance and Commitment Training (ACT) [48,49]. Rather than focusing on directly modifying psychological symptoms, ACT aims to increase psychological flexibility and decrease psychological inflexibility processes in the service of expanding adaptive, values-based behavioral repertoires [44]. This can alter the way in which an individual responds to stressful experiences, producing secondary benefits across a wide range of psychological and organizational variables [49-51].

Within the broader literature, psychological flexibility is associated with well-being [52,53] and has been shown to protect against burnout and adverse mental health outcomes in stressful situations [42,47,52] (including depression and anxiety [51]). Interventions that train individuals to develop and strengthen their psychological flexibility skills are effective in reducing work-related stress and burnout severity [36,41,54-57] and improving well-being [36,41,43]. Mediation studies demonstrate that ACT interventions exert their beneficial impacts on burnout and well-being by improving an individual's psychological flexibility [36], suggesting that this skill set is a mechanism of change for these outcomes. In a longitudinal ACT intervention study, improvements in psychological flexibility mediated reductions in exhaustion [42], and this prevented the later development of cynicism [42]. This is an important finding with respect to the potential burnout prevention benefits of psychological flexibility training among medical students, for whom exhaustion is the most prevalent factor and is considered a foundational manifestation of burnout [2,9].

There is early evidence suggesting that psychological flexibility skills may function as important personal resources in the medical profession [31]. Low psychological flexibility is associated with burnout risk among medical students, physicians [58], and resident physicians [59]. Medical students who report low psychological flexibility also demonstrate diminished satisfaction with life and greater personal distress when seeing others in harm, which may increase burnout risk during their careers [31]. Furthermore, burnout risk is higher among medical students who engage in experiential avoidance [11,32] or non–values-based actions [60]. Conversely, higher psychological flexibility was recently found to predict lower burnout among medical students and physicians during the pandemic [58]. There is minimal research examining the benefits of psychological flexibility skill training interventions for medical students. A recent study found that distressed medical students who completed ACT training as a requirement of their undergraduate medical curriculum demonstrated improved burnout outcomes, but this study did not include a control group [61]. A small semiexperimental study demonstrated improvements in well-being and psychological distress among female medical students in Iran following completion of an ACT intervention [62]. Further rigorous efficacy studies are needed to evaluate whether training medical students in psychological flexibility skills improves burnout and psychological well-being outcomes.

App-Based Intervention Delivery

Accessibility

A recent systematic review of individual resource-building interventions for medical students highlighted the need for research evaluating the effectiveness of nontraditional delivery methods such as smartphone apps [21]. As medical students use smartphones frequently [63], stand-alone app–delivered interventions have the potential for cost-effective scalability [39,64] and can offset known accessibility barriers by providing anonymous and private access to medical students concerned about mental health stigma [11,32,64]. Stand-alone apps offer accessibility in times when face-to-face delivery is not an option,

such as during pandemic-related lockdowns [65], and can deliver brief training components at convenient times for those with busy schedules [64,66]. Psychological skill generalization may also be enhanced because of the accessibility of training opportunities in everyday life situations [67]. However, feasibility trials of app-based psychological interventions for medical students also indicate that maintaining engagement in this accessible medium can be challenging [63,68].

Individualization

App technologies have been identified as important to advance key research priorities within the psychological flexibility literature because of their potential to facilitate methodologies that accommodate individual heterogeneity and enhance intervention precision [69]. Although the benefits of ACT have been assessed in >1000 randomized controlled trials (RCTs) [50], traditional approaches evaluating the effectiveness of generic interventions by comparing aggregated group data are limiting as they overlook important information regarding individual differences in training needs and outcomes [69-71]. Generalized deployment of interventions found to be effective at the group level is likely to result in some individuals not receiving the type or amount of training they need and others receiving more than necessary [57,70]. With respect to psychological flexibility processes, research suggests that individuals might require training in different skills at different times and in different situations [45]. Although high psychological flexibility is often associated with low psychological inflexibility (and vice versa), several distinct and more complex profiles have recently been observed [46], indicating that flexibility and inflexibility processes do not necessarily exist at opposite ends of a spectrum but may vary temporally and contextually within individuals [45,46,72]. Thus, there is a growing focus on individualized approaches in which intervention decisions are driven by an individual participant's identified needs and skill deficits in varying moments and contexts [69,70]. Apps have the potential to facilitate both the real-time ecological momentary assessment (EMA) [69] of these needs in a participant's everyday life and the targeted delivery of just-in-time adaptive intervention components that align with these needs [73]. Importantly, as an individualized approach adjusts for heterogeneity at the intervention level, an RCT study design can be used to assess its treatment utility over a nonindividualized approach [70].

As ACT is a theory-driven intervention with practical training components that have clear functional links to the corresponding core process within the psychological flexibility model [44,48,74], it is well suited to individualized delivery via an app [75]. Levin et al [75] demonstrated a simple EMA strategy for identifying a participant's present-moment psychological flexibility training needs, involving a brief subjective assessment of which process an individual felt they were experiencing the most difficulty with each time they accessed an app-delivered ACT intervention. This information was used to individualize the training by delivering a practical ACT skill activity that aligned with the identified psychological flexibility process at that moment. Using a 3-arm RCT, the researchers demonstrated that, compared with a nonindividualized ACT app and a nontreatment group, university students who engaged with the individualized app demonstrated statistically significantly greater improvements in psychological distress and well-being. Empirical evaluation of individualized ACT interventions is in its early stages, and more research is needed to examine effectiveness and advance the development of individualization methodologies [69,75,76]. To date, no studies have examined the potential burnout and well-being benefits of training medical students in psychological flexibility skills using individualized or nonindividualized ACT apps.

Study Aims

The aim of this RCT was to evaluate the effectiveness of an app-delivered ACT intervention for medical students with respect to burnout (exhaustion [primary outcome], cynicism, and academic efficacy), well-being, and psychological flexibility and inflexibility outcomes. We hypothesized that medical students who engaged with either an individualized or nonindividualized version of the ACT app would demonstrate greater postintervention improvements in outcomes than those in a waitlist group and that intervention effects would be greater for the individualized group than for the nonindividualized group. We aimed to examine whether any observed postintervention improvements in burnout or well-being would be mediated by improvements in psychological flexibility and inflexibility. Furthermore, we aimed to explore whether engaging in either version of the ACT app would improve other relevant secondary psychological outcomes (stress, depression, and anxiety).

Methods

Ethics Approval

Ethics approval for this study was granted by the University of Newcastle Human Research Ethics Committee on January 21, 2021 (approval ID: H-2020-0311), and ratified by the University of New England Human Research Ethics Committee on February 11, 2021.

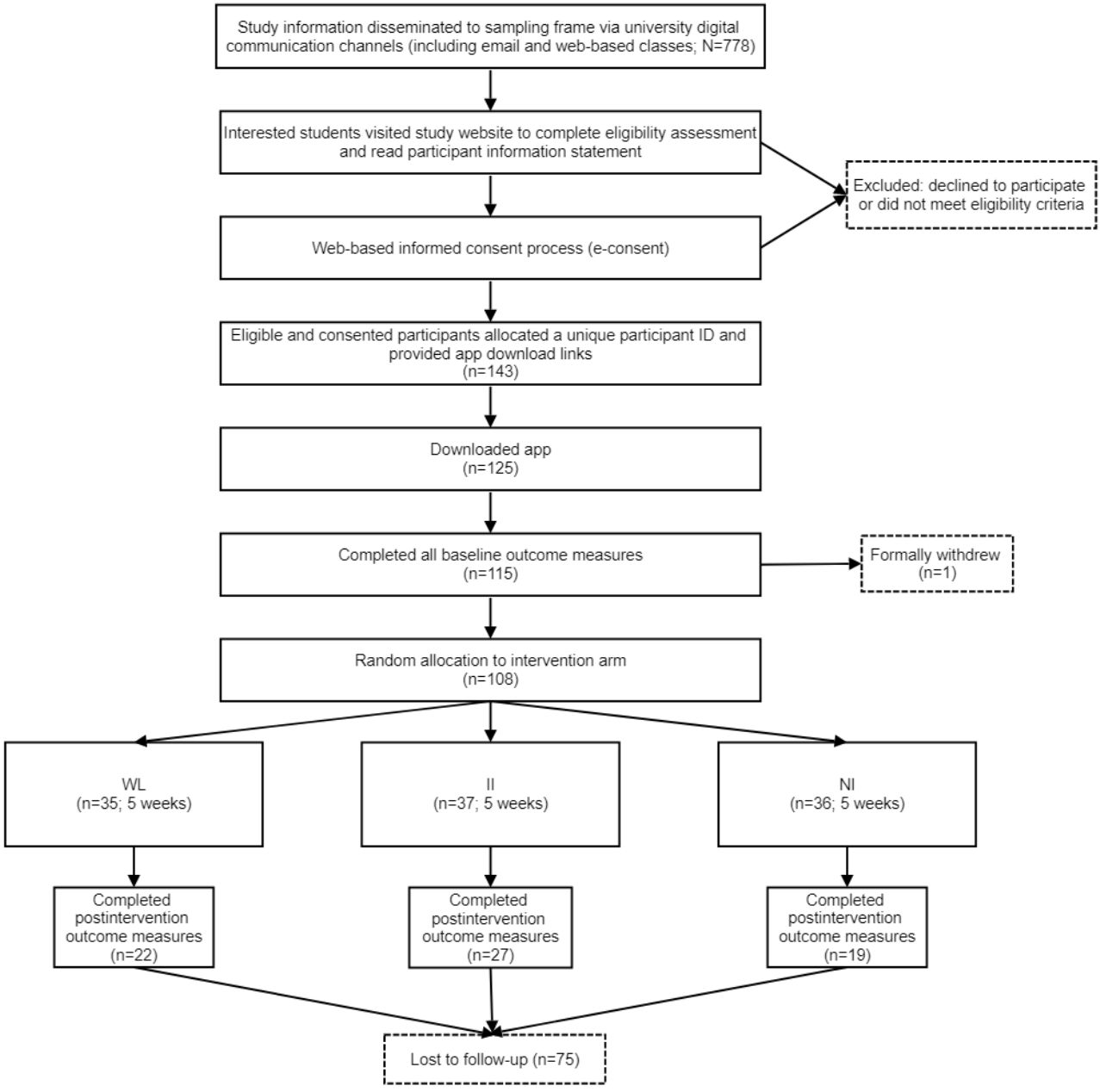
Trial Design

This study was a 3-arm, parallel RCT of a 2-stage psychological flexibility skill training app for medical student burnout and well-being. Randomization was performed within the app using a 1:1:1 allocation ratio and a simple randomization procedure where each student had a 1 in 3 chance of allocation to each intervention arm (individualized intervention, nonindividualized intervention, and waiting list). The inclusion of the nonindividualized group was an important element of this research design as it provided a generic ACT intervention control condition against which to evaluate the potential relative benefits (or "treatment utility" [70]) of the individualized approach adopted [75]. The functional limitations of the app meant that it was not possible to stratify randomization by participant baseline characteristics. Allocation was blinded for participants assigned to the individualized and nonindividualized intervention groups. Blinding was not possible for the waitlist group.

Recruitment and Study Setting

Recruitment was conducted for 6 weeks during August 2021 and September 2021 (refer to Figure 1 for the participant flow diagram). This unintentionally coincided with a mandated lockdown because of a regional COVID-19 outbreak. The sampling frame was students enrolled in first, second, fourth, and fifth years of the Joint Medical Program (JMP) at the University of Newcastle or the University of New England, Australia (N=778). During the first and second years of the JMP, students predominantly engage in in-class academic learning, whereas the fourth and fifth years involve a stronger focus on applied clinical training. Students enrolled in the third year of the JMP were not invited to participate in this study as they had previously taken part in a feasibility trial of the app.

Figure 1. Participant flow diagram. II: individualized intervention; NI: nonindividualized intervention; WL: waiting list.



Students were invited to participate in a trial to assess the efficacy of a psychological flexibility skill training app for well-being and burnout prevention. An invitation email was sent to students' university accounts from a JMP administrative account approximately every 2 weeks during the recruitment period, and a member of the research team also delivered a verbal invitation at the end of web-based classes. Students were provided with a URL and QR code that gave them access to the enrollment website (located on a secure web-based survey and database platform, REDCap [Research Electronic Data Capture; Vanderbilt University] [77,78], hosted at Hunter Medical Research Institute) during the recruitment period.

Students who accessed the enrollment website were assessed for eligibility and were eligible for inclusion if they had regular access to a reliable internet connection and an electronic device compatible with app use (smartphone or tablet). There were no ineligibility criteria. Participation was voluntary, confidential,

and independent of students' academic programs. Students were required to read a participant information statement before deciding whether to enroll in the study. This included information about the purpose of the study, time required to complete outcome measures (approximately 10 minutes at each time point), intervention and study flow, privacy and data storage, and withdrawal processes. Students who elected to participate registered their consent through an e-consent process. Following completion of a brief self-report demographic assessment (baseline participant characteristics), enrolled students were provided with links to download the app via the Play Store (Android) and App Store (Apple). The app was password-protected to ensure that only study participants had access to the outcome measures (closed surveys) and intervention.

Data Privacy

To ensure the privacy and anonymity of participant data, 2 password-protected databases were used to store personal and demographic information (in an identifiable "Participant Information Database") separately from the outcome data (stored in an anonymous "Study Database"). Outcome data were deidentified, with participant ID used as a linkage key. Only the lead author had access to the participant information database. Members of the research team did not have access to identifying participant information. To protect privacy while using the app, students input their unique participant ID as their username when registering and accessing their account.

Procedures

Overview

The study period for each participant commenced when they downloaded the app and registered their account. All further assessment, intervention, and data collection procedures were conducted within the app, which students accessed independently during the study period. The intervention and assessment components were resumable, allowing students to engage with the app at their own pace. Participants first completed baseline psychological outcome measures followed by a brief (10 minutes) onboarding session, which outlined how to use the app and introduced psychological flexibility as an adaptive behavioral skill set that could help students do what is important to them while effectively navigating life's challenges. Students were then randomized into 1 of the 3 intervention arms, after which the app was programmed to deliver the intervention pathway associated with their group allocation. The individualized and nonindividualized groups were given access to the psychological flexibility training intervention via the app for 5 weeks. For students assigned to the waiting list, access to app content was disabled for 5 weeks. After the study period, all 3 groups were asked to complete the postintervention outcome measures. Upon completion, waitlist participants were given access to the individualized version of the app, and intervention group participants were given continued access to their allocated version. All students who completed the postintervention outcome measures were given an Aus $30 (US $20.70) digital gift voucher to acknowledge the time commitment involved in participation.

The 5-week intervention duration was set in consultation with the participating medical schools. This provided students with the opportunity to participate in the study during the semester while reducing participation burden by ensuring that the timing of outcome measures did not correspond with critical assessment periods. This duration was considered appropriate as Levin et al [75] observed intervention effects for a similar 4-week program, and feasibility trialing of the current app demonstrated that medical students engaged most frequently during the early weeks of access [68].

Intervention

Overview

The intervention was an ACT-based psychological flexibility [44] training program delivered via a smartphone app ("BiSi: Build it. Sustain it.") created for this study. A more detailed protocol for this study is available elsewhere [79]. BiSi was developed by clinical psychologists (ED and BK) with extensive experience in the psychological flexibility model and involved the adaptation of existing ACT concepts and training activities to suit the app-based context and target participant group. The intervention was delivered in 2 stages.

Stage 1: Learn the Concepts (Introductory Module)

Stage 1 involved the delivery of an introductory module ("Learn the Concepts"), which was identical for the individualized and nonindividualized groups. The purpose of the module was to familiarize students with the psychological flexibility model and its potential benefits and provide a conceptual framework for understanding the stage 2 experiential skill training activities. Students could complete the module over multiple sittings (total completion time of approximately 60 minutes depending on individual pace). The module comprised 7 sections (<10 min each) presented in a fixed order. Students were required to complete each section before progressing to the next. Section 1 provided psychoeducation about burnout (focusing on destigmatizing burnout-related experiences and how to recognize the signs) and well-being (including the importance of choosing actions that involve the consideration of personal well-being). Sections 2 to 7 provided education about each psychological flexibility and inflexibility process and outlined how psychological flexibility skills could be implemented to support well-being and protect against burnout. Although the module was primarily conceptual, experiential components encouraged self-reflection and provided opportunities to practice each skill set (eg, personal values identification, mindfulness, experiential acceptance, thought defusion, perspective shifting, and values-based goal setting). Written and audio versions of the psychoeducational content were provided, along with accompanying images. Some skill activities were presented in either written or audio format depending on which translated best to the app-based delivery mode. Completion of the stage 1 introductory module unlocked students' access to the stage 2 on-demand skill-training dashboard.

Stage 2: Learn the Behaviors (On-Demand Skill Training)

Stage 2 provided access to a library of short (3 to 8 minutes) experiential psychological flexibility skill activities that students could practice at any time (on demand). Each of the 6

psychological flexibility processes had its own dashboard where all ACT training activities for that skill set were displayed on a list (20 activities per process; 120 in total). Although most activities were presented in audio format with accompanying images, some included written components and items that required the user to input a response.

During stage 2, students from both intervention groups were presented with a main dashboard (home screen) each time they opened the app. To access the psychological flexibility skill activities, students selected "Practice a Skill" on the home screen and were presented with a single-item EMA screening question ("Which of the following are you having the most difficulty with today?") to identify which psychological flexibility skill set might be most relevant to their training needs on that occasion (adapted from Levin et al [75]). Students selected 1 of 6 response options, each of which corresponded to one of the 6 psychological flexibility processes: "Struggling with your feelings" (acceptance), "Unable to do what matters to you" (committed action), "Stuck in your thoughts" (defusion), "Stuck in autopilot or struggling to stay in the present moment" (present-moment awareness), "Disconnected from a sense of meaning or purpose" (values), and "Stuck in stories about who you are or who you should be" (self-as-context).

The EMA screening question was used to tailor the intervention to the training needs of students in the individualized group. Upon completion of the screening question, participants in the individualized group were presented with the dashboard corresponding to the psychological flexibility skill set they had identified as having the most difficulty with on that occasion (eg, if a student reported feeling *disconnected from a sense of meaning or purpose*, the app displayed the values dashboard). This allowed students in the individualized group to practice a skill targeted to the area of identified need each time they accessed the app. For the nonindividualized group, students' responses to the screening question had no impact on the skill set in which they received training. After responding to the screening question, students in the nonindividualized group were presented with 1 of the 6 psychological flexibility dashboards at random (simple randomization; each process had an equal chance of selection). Students were informed that there were 2 intervention groups and that the only difference was the way in which skill activities were selected for them by the app.

Once students were given access to a psychological flexibility dashboard, they could practice any activity from that skill set, either by choosing from the list or by allowing the app to select for them. The latter involved a simple randomization process where each activity had an equal chance of selection (1 in 20). Upon completion of an activity, students had the option to complete another within the same psychological flexibility skill set. If they selected "yes," these steps were repeated until the participant elected to discontinue.

The version of BiSi implemented in this study incorporated medical student feedback from a small feasibility trial of the app [68] with the aim of enhancing intervention relevance and engagement. This included clear explanations of what to expect during each stage of the intervention, providing progress indicators for all app components, providing earlier opportunities for personally relevant experiential learning and self-reflection (during stage 1), and delivering content in both written and audio formats where practical (during stage 1). We also introduced positive reinforcement components that were delivered after students completed certain activities or a certain number of activities (eg, achievement badges and experience points). Students were asked to complete a minimum of 4 stage 2 skill activities but were also encouraged to optimize skill learning by practicing more regularly. Students were sent reminders to use the app at 6 PM each day. This frequency and time were based on feedback from the feasibility trial. Although reminders were intended to be delivered using push notifications as per medical students' reported preferences, this function did not operate as intended during the study, and reminders were sent by email instead.

Outcome Measures

Psychological Outcome Measures

Overview

Self-report psychological outcome measures were administered at 2 time points: baseline (collected before randomization) and postintervention measurement (collected 5 weeks after randomization). The outcome measures were presented in the order of the following sections. Participants were required to provide a response to each item before progressing to the next. Owing to the limitations of the app, participants could not alter their responses once submitted.

Burnout

Exhaustion was selected as the primary outcome for this study as it is the most prevalent factor among medical students [9] and physicians [2] and early improvements in exhaustion following an ACT intervention may prevent the future development of other burnout factors (cynicism) [42]. Cynicism and academic efficacy were assessed as secondary burnout outcomes. The Maslach Burnout Inventory [80] is a valid [81], gold-standard [82] measure of the 3-factor burnout model. Of the versions available, the General Survey for Students (MBI-GS [S]) [80,81] was the most appropriate for the medical student cohort. The 16-item self-report questionnaire assesses the degree to which students are experiencing each factor using a 7-point Likert scale ranging from 0 ("never") to 6 ("every day"). Higher total scores for exhaustion and cynicism and lower total scores for academic efficacy are indicative of higher frequencies of burnout-related experiences. The reliability of the MBI-GS (S) has been demonstrated among medical students [66,83]. In this study, internal consistency was excellent for exhaustion (Cronbach α=.90) and good for cynicism (Cronbach α=.83) and academic efficacy (Cronbach α=.83). Items were presented in consecutive order, with 1 item displayed per app screen (18 screens, including introductory text).

Well-being

Well-being was measured as a secondary outcome using the Mental Health Continuum–Short Form [84], which assesses self-reported hedonic (ie, feeling good) and eudaimonic (ie, functioning well) aspects of well-being [85]. Participants rate the frequency of 14 well-being experiences during the previous month using a 6-point Likert scale ranging from 0 ("never") to

5 (“every day”). Total well-being is estimated by summing all scale items. Higher scores reflect higher overall well-being. The scale has demonstrated validity [84,85], reliability [85,86], and sensitivity to change in web-based intervention studies [87]. Internal consistency was excellent in our sample (Cronbach α=.92). Items were presented in consecutive order, with 1 item displayed per app screen (15 screens, including introductory text).

Psychological Flexibility and Inflexibility

The Multidimensional Psychological Flexibility Inventory–Short Form [45] was administered to evaluate whether the intervention improved medical students’ psychological flexibility and reduced their psychological inflexibility (secondary outcomes). This 24-item self-report questionnaire assesses the frequency of psychological flexibility and psychological inflexibility experiences during the previous 2 weeks using a 6-point Likert scale ranging from 1 (“never true”) to 6 (“always true”). The scale provides separate average global composite scores for psychological flexibility and inflexibility as recent research suggests that they are related but “conceptually distinct” processes that may have disparate relationships with well-being and psychological distress outcomes and may respond differently to interventions [45,46,72]. The scale has demonstrated validity [72,88] and reliability [88,89] and is responsive to changes over time [45]. This study demonstrated good internal consistency for psychological flexibility (Cronbach α=.87) and inflexibility (Cronbach α=.87). Items were presented in consecutive order, with 1 item displayed per app screen (24 screens).

Psychological Distress (Depression, Anxiety, and Stress)

The Depression, Anxiety, and Stress scale–21 [90] is a valid and reliable [90,91] 21-item self-report questionnaire assessed using a 4-point Likert scale. The measure provides subscale scores that estimate the severity of depression, anxiety, and stress symptoms [90]. We included this measure to explore whether the ACT intervention improved these secondary outcomes and evaluate whether these psychological distress factors affected engagement in the intervention, as has been observed in previous digital intervention studies [92]. Internal consistency was excellent for depression (Cronbach α=.90) and good for anxiety (Cronbach α=.81) and stress (Cronbach α=.87) in this study. Items were presented in consecutive order, with 7 items displayed per app screen (3 screens).

Intervention Engagement Outcomes

Behavioral engagement data and subjective intervention feedback were collected from participants throughout the study.

Study Attrition

Attrition was defined as formal withdrawal or loss to follow-up at any stage during the study without completing the postintervention outcome measures.

Intervention Adherence

Adherence to the individualized and nonindividualized intervention arms was defined as the completion of all stage 1 components and engagement in at least 4 skill activities during stage 2. This level of adherence provided students with the opportunity to learn about and practice each of the psychological flexibility processes (stage 1) and practice the skills a few times in their everyday lives (stage 2).

Intervention Feedback

Students were invited to submit feedback on their experience of using the app via a form presented halfway through stage 1. The feedback form was also accessible via the main dashboard. During stage 2, participants rated whether they liked each skill activity they practiced using a single-item binary measure (*thumbs up* [like] or *thumbs down* [dislike] icon) [93]. Participants were invited to report concerns or harms experienced during the study using contact links provided within the app and via email communication (eg, daily reminders).

Data Analysis

Power Analysis

R (statistical computing package; R Foundation for Statistical Computing) was used to calculate the standard power for a 2-tailed independent-sample *t* test for between-group differences not accounting for repeated measures. This was considered a conservative choice as reliable within-person correlation information needed for power based on a mixed model was not available. This analysis indicated that a sample of 117 participants would provide sufficient power (80%) to detect a clinically meaningful effect size (SD 0.65) between either intervention arm and the control arm in the primary outcome (exhaustion), measured using the MBI-GS (S), with a type-I error rate of 5%. We aimed to recruit up to 153 participants to provide a 30% margin for attrition.

Participant Psychological Characteristics at Baseline

Average participant psychological characteristics at baseline were compared with reference samples (including other medical student or general population samples from previously published studies) using single-sample *t* tests.

Intervention Effects

This study adopted an intention-to-treat analysis, which included data collected from all participants randomized into a study group. Each participant’s data were analyzed based on the study group to which they were randomized irrespective of their degree of intervention engagement or whether they met the adherence criteria. This approach maintains randomization benefits and minimizes bias when assessing intervention efficacy [94].

Differences between the intervention arms were assessed using linear mixed regression models for primary (exhaustion) and secondary (cynicism, academic efficacy, well-being, psychological flexibility and inflexibility, depression, anxiety, and stress) outcomes. A separate model was estimated for each outcome variable. The models included fixed categorical effects for time (baseline as the referent), intervention group, the interaction between intervention and time, and self-reported baseline participant characteristics that were imbalanced after randomization (ie, gender and whether students were studying medicine as their first career [“first career”]). The model included a random participant-specific intercept to account for the repeated measures for each participant. The adjusted difference between the intervention groups in mean change from

baseline to postintervention measurement is presented, as well as the within-group change from baseline, 95% CIs, and 2-tailed *P* values. Model assumptions were assessed by inspecting residual plots. For outcomes that violated modeling assumptions (ie, demonstrated nonnormality of residuals or nonconstant variance), a robust linear mixed-effects model was estimated, with the same fixed and random effects as the previous models.

To control for elevated type-I errors arising from having 2 primary contrasts of interest (each intervention vs control), we followed a hierarchical testing procedure for the primary outcome (exhaustion) where the less intensive intervention (nonindividualized) would only be declared to be significantly different from the control (at a 5% significance threshold) if the more intensive intervention (individualized) was statistically significantly different from the control at a 5% significance level. All other analyses were exploratory. We note that the study was not powered to directly compare the individualized and nonindividualized intervention arms.

Mediation Analyses

We planned to conduct mediation analyses to assess whether changes in process outcomes (psychological flexibility and inflexibility) between baseline and postintervention measurement mediated changes in psychological outcomes. However, the observed intervention effects did not support the implementation of these analyses for reasons outlined in the Results section.

Study Attrition, Intervention Adherence, and Engagement

Baseline participant demographic and psychological characteristics were compared between (1) those who were lost to follow-up at any point during the study versus those who were not and (2) those who met intervention adherence criteria versus those who did not using chi-square tests for categorical variables (or the Fisher exact test where cell size was <5) and independent-sample Student *t* tests for continuous variables (Welch *t* tests used where the assumption of equal variance was violated). The average rates of engagement during stage 2 were compared between the individualized and nonindividualized groups using independent-sample *t* tests.

Results

Participant Demographics

A total of 143 medical students were enrolled in this study. Table 1 shows demographic characteristics of the total enrolled sample and each intervention group. The enrolled participants were aged between 18 and 51 years (mean 24.0, SD 5.48 years), and more than half (88/143, 61.5%) were female. Most participants were nonindigenous (133/143, 93%) and domestic students (135/143, 94.4%) and were training in medicine as their first career (108/143, 75.5%). The average time spent in the workforce was 5.74 (SD 5.71) years. A substantial proportion of the enrolled students (122/143, 85.3%) reported having previously experienced burnout, and 21% (30/143) were engaging in psychological treatment at the time of the study. Students rated the quality of their health, diet, and self-care using a Likert scale ranging from 1 (“very poor”) to 5 (“excellent”). Mean ratings for quality of health (mean 3.72, SD 0.89) and diet (mean 3.61, SD 0.73) fell between “average” (3) and “good” (4), whereas mean ratings for self-care (mean 3.19, SD 0.82) were closer to an “average” rating.

Table 1. Participant demographic characteristics at baseline (by study group allocation and total; N=143).

	II[a] (n=37)	NI[b] (n=36[c])	WL[d] (n=35)	Not allocated (n=35)	Total enrolled
Age (years), mean (SD; range)	23.7 (5.60; 18-46)	25.4 (5.78; 19-42)	22.1 (3.01; 18-31)	24.7 (6.55; 19-51)	24.0 (5.48; 18-51)
Gender, n (%)					
Female	26 (70.3)	17 (47.2)	23 (65.7)	22 (62.9)	88 (61.5)
Male	11 (29.7)	18 (50)	9 (25.7)	13 (37.1)	51 (35.7)
Nonbinary	0 (0)	0 (0)	3 (8.6)	0 (0)	3 (2.1)
Enrollment, n (%)					
Domestic	35 (94.6)	34 (94.4)	32 (91.4)	34 (97.1)	135 (94.4)
International	2 (5.4)	1 (2.8)	3 (8.6)	1 (2.9)	7 (4.9)
Year of study, n (%)					
1	9 (24.3)	6 (16.7)	11 (31.4)	18 (51.4)	44 (30.8)
2	15 (40.5)	14 (38.9)	7 (20)	7 (20)	43 (30.1)
4	7 (18.9)	7 (19.4)	8 (22.9)	4 (11.4)	26 (18.2)
5	6 (16.2)	8 (22.2)	9 (25.7)	6 (17.1)	29 (20.3)
Indigenous, n (%)	2 (5.4)	4 (11.1)	0 (0)	4 (11.4)	10 (7)
First career, n (%)	26 (70.3)	25 (69.4)	32 (91.4)	25 (71.4)	108 (75.5)
Years in the workforce, mean (SD)	5.19 (6.0)	6.54 (5.95)	3.91 (3.12)	7.34 (6.69)	5.74 (5.71)
Previous burnout, n (%)	34 (91.9)	31 (86.1)	27 (77.1)	30 (85.7)	122 (85.3)
Current therapy, n (%)	6 (16.2)	7 (19.4)	11 (31.4)	6 (17.1)	30 (21)
Health rating, mean (SD)	3.62 (0.92)	4.00 (0.80)	3.40 (0.98)	3.83 (0.86)	3.72 (0.89)
Diet rating, mean (SD)	3.43 (0.87)	3.97 (0.57)	3.43 (0.71)	3.54 (0.82)	3.61 (0.73)
Self-care rating, mean (SD)	3.22 (0.81)	3.29 (0.66)	3.17 (0.71)	3.06 (0.84)	3.19 (0.82)

[a]II: individualized intervention.

[b]NI: nonindividualized intervention.

[c]Demographic data missing for 1 participant in this group.

[d]WL: waiting list.

Participant Psychological Characteristics at Baseline

Outcome scores were calculated for all participants who completed the measures at baseline and only those who were randomized to an intervention arm. Removal of the participants lost to follow-up before random allocation did not greatly alter the mean scores for any psychological outcome. Therefore, we compared the average psychological characteristics of only the participants who proceeded to randomization and were included in the efficacy analyses (108/143, 75.5%) with reference samples (other medical students or the general population) from previously published studies (Multimedia Appendix 1 [66,88,90,95]).

On average, medical students presented a mixed profile of burnout scores, demonstrating exhaustion levels (mean 16.29, SD 7.20) comparable with those of a sample of medical students not experiencing burnout (mean 14.96, SD 5.71; t_{269}=1.69; P=.09), academic efficacy scores (mean 24.69, SD 6.47) comparable with those of medical students who perceived themselves to be high in burnout (mean 24.81, SD 5.35; t_{216}=0.15; P=.88), and cynicism scores (mean 10.87, SD 6.92) significantly lower than those of the burnout sample (mean 14.44, SD 5.59; t_{216}=4.19; P<.001) but higher than those of the nonburnout sample (mean 7.59, SD 5.16; t_{269}=4.46; P<.001) [66]. Total well-being scores (mean 43.03, SD 13.31) were significantly lower than those of a large multi-institutional sample of medical students in the United States (mean 47.0, SD 12.67; t_{2796}=3.12; P=.002) [95]. Psychological inflexibility scores (mean 3.21, SD 0.91) were significantly higher than those of a general population sample (mean 2.73, SD 0.90; t_{2769}=5.43; P<.001) [88]. Average depression (mean 14.89, SD 10.25), anxiety (mean 8.50, SD 8.09), and stress (mean 11.87, SD 10.34) scores for the medical student sample in this study were significantly higher than the general population averages (*depression:* mean 6.34, SD 6.97, t_{3020}=12.27, and P<.001; *anxiety:* mean 4.7, SD 4.91, t_{3020}=7.67, and P<.001; *stress:* mean 10.11, SD 7.91, t_{3020}=2.24, and P=.03) [90].

Intervention Effects

Overview

Table 2 shows the results for the burnout; well-being; psychological flexibility and inflexibility; and depression, anxiety, and stress outcomes.

Table 2. Data summaries of outcomes at each time point (baseline and postintervention measurement) for each treatment arm (mean and SD), mixed model–based estimates of within-group change (95% CI), and estimated between- and within-group change (95% CI and *P* value).

Outcome, subcategory, and intervention group	Baseline (n=108), mean (SD)	Postintervention measurement (n=68), mean (SD)	Baseline–postintervention measurement estimated change (95% CI)	*P* value
Burnout				
Exhaustion (primary)				
WL[a]	17.91 (7.02)	17.05 (6.81)	−1.61 (−3.96 to 0.745)	N/A[b]
II[c]	16.65 (7.22)	14.48 (6.85)	−2.13 (−4.27 to 0.008)	N/A
NI[d]	14.33 (7.09)	14.79 (5.60)	−0.01 (−2.51 to 2.49)	N/A
II-WL	N/A	N/A	−0.52 (−3.70 to 2.65)	.75
NI-WL	N/A	N/A	1.60 (−1.84 to 5.03)	.37
II-NI	N/A	N/A	−2.12 (−5.41 to 1.17)	.21
Cynicism				
WL	11.5 (7.35)	12.0 (7.62)	−0.59 (−1.94 to 0.77)	N/A
II	11.7 (7.01)	9.74 (5.70)	−1.76 (−3.91 to 0.39)	N/A
NI	9.36 (6.32)	9.95 (5.54)	0.50 (−2.01 to 3.02)	N/A
II-WL	N/A	N/A	−1.26 (−4.46 to 1.94)	.44
NI-WL	N/A	N/A	1.00 (−2.45 to 4.46)	.57
II-NI	N/A	N/A	−2.26 (−5.57 to 1.05)	.18
Academic efficacy				
WL	23.2 (6.29)	23.9 (6.27)	1.29 (0.51 to 2.08)	N/A
II	25.9 (6.42)	27.5 (5.29)	1.25 (0.01 to 2.49)	N/A
NI	24.9 (6.60)	26.4 (6.86)	2.33 (0.86 to 3.80)	N/A
II-WL	N/A	N/A	0.94 (−0.90 to 2.79)	.32
NI-WL	N/A	N/A	2.02 (0.02 to 4.03)	.05
II-NI	N/A	N/A	−1.08 (−3.00 to 0.84)	.27
Well-being				
Total well-being				
WL	39.5 (13.3)	40.4 (13.1)	3.83 (1.49 to 6.18)	N/A
II	44.2 (13.7)	48.1 (11.3)	3.51 (−0.20 to 7.22)	N/A
NI	45.2 (12.5)	52.1 (8.8)	7.22 (2.87 to 11.57)	N/A
II-WL	N/A	N/A	2.74 (−2.77 to 8.26)	.33
NI-WL	N/A	N/A	6.46 (0.49 to 12.42)	.04
II-NI	N/A	N/A	−3.71 (−9.43 to 2.00)	.21
Psychological flexibility				
Flexibility				
WL	3.65 (0.73)	3.69 (0.79)	0.26 (0.09 to 0.42)	N/A
II	3.74 (0.87)	4.22 (0.83)	0.48 (0.22 to 0.74)	N/A
NI	3.78 (0.73)	4.10 (0.72)	0.32 (0.02 to 0.62)	N/A
II-WL	N/A	N/A	0.50 (0.12 to 0.89)	.01
NI-WL	N/A	N/A	0.35 (−0.07 to 0.76)	.11
II-NI	N/A	N/A	0.16 (−0.24 to 0.56)	.44
Inflexibility				
WL	3.32 (0.91)	3.34 (0.84)	−0.28 (−0.47 to −0.10)	N/A

Outcome, subcategory, and intervention group	Baseline (n=108), mean (SD)	Postintervention measurement (n=68), mean (SD)	Baseline–postintervention measurement estimated change (95% CI)	*P* value
II	3.27 (1.04)	2.77 (0.69)	−0.53 (−0.82 to −0.23)	N/A
NI	3.03 (0.76)	2.83 (0.75)	−0.28 (−0.62 to 0.07)	N/A
II-WL	N/A	N/A	−0.48 (−0.92 to −0.04)	.04
NI-WL	N/A	N/A	−0.23 (−0.70 to 0.24)	.34
II-NI	N/A	N/A	−0.25 (−0.70 to 0.21)	.29
Psychological distress				
Depression[e]				
WL	11.8 (9.49)	11.7 (8.92)	−0.54 (−3.28 to 2.19)	N/A
II	12.8 (12.1)	7.7 (7.25)	−4.40 (−6.93 to −1.86)	N/A
NI	11.1 (9.4)	6.42 (5.56)	−3.95 (−6.92 to −0.97)	N/A
II-WL	N/A	N/A	−3.85 (−7.58 to −0.12)	.046
NI-WL	N/A	N/A	−3.40 (−7.44 to 0.64)	.10
II-NI	N/A	N/A	−0.45 (−4.36 to 3.46)	.82
Anxiety[e]				
WL	8.63 (7.46)	7.13 (6.29)	−1.20 (−3.19 to 0.79)	N/A
II	9.41 (9.3)	6.44 (7.28)	−3.10 (−4.94 to −1.26)	N/A
NI	7.61 (7.5)	4.00 (3.83)	−3.45 (−5.62 to −1.29)	N/A
II-WL	N/A	N/A	−1.90 (−4.61 to 0.81)	.17
NI-WL	N/A	N/A	−2.25 (−5.19 to 0.69)	.14
II-NI	N/A	N/A	0.36 (−2.49 to 3.20)	.81
Stress[e]				
WL	13.9 (8.99)	15.8 (8.8)	2.24 (−1.51 to 5.99)	N/A
II	16.1 (11.30)	11.6 (8.16)	−4.65 (−8.14 to −1.15)	N/A
NI	14.5 (10.40)	10.0 (7.42)	−4.12 (−8.17 to −0.06)	N/A
II-WL	N/A	N/A	−6.89 (−12.01 to 5.99)	.01
NI-WL	N/A	N/A	−6.36 (−11.90 to −0.83)	.03
II-NI	N/A	N/A	−0.53 (−5.88 to 4.82)	.85

[a]WL: waiting list.

[b]N/A: not applicable.

[c]II: individualized intervention.

[d]NI: nonindividualized intervention.

[e]Robust linear mixed-effects models estimated because of violation of linear mixed regression modeling assumptions (ie, nonnormality of residuals or nonconstant variance).

Intervention (Individualized and Nonindividualized) Versus Waitlist Group Comparisons

Burnout

There were no statistically significant differences between the intervention arms and the waitlist control group for burnout outcomes (exhaustion [primary outcome]: −0.52, 95% CI −3.70 to 2.65, and *P*=.75 for individualized and 1.60, 95% CI −1.84 to 5.03, and *P*=.34 for nonindividualized; cynicism: −1.26, 95% CI −4.46 to 1.94, and *P*=.44 for individualized and 1.00, 95% CI −2.45 to 4.46, and *P*=.57 for nonindividualized; academic efficacy: −0.90, 95% CI −0.90 to 2.79, and *P*=.32 for individualized and 2.02, 95% CI 0.02-4.03, and *P*=.05 for nonindividualized).

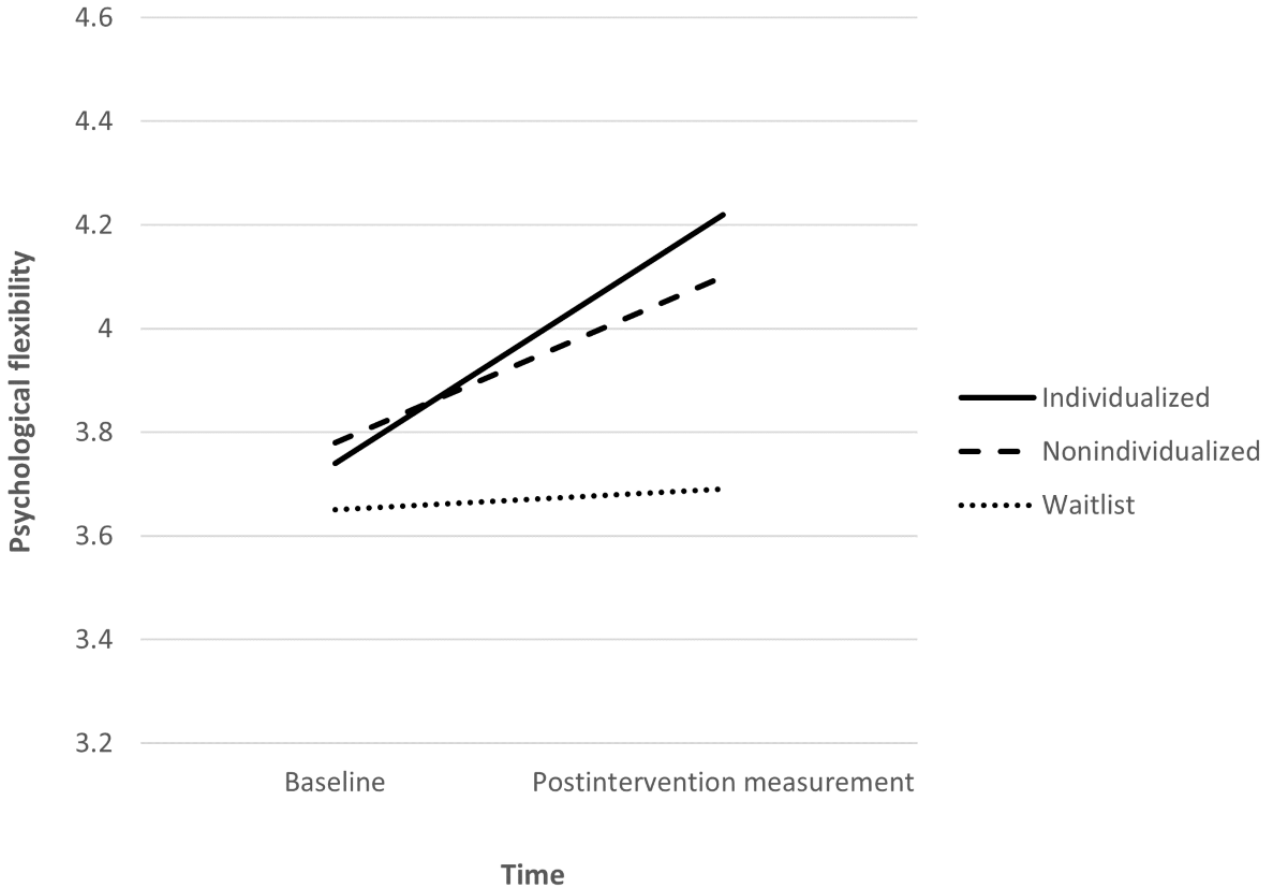
Well-being

The estimated change in total well-being between baseline and postintervention measurement was significantly greater for the nonindividualized group (0.52) than for the waitlist group (0.27; *P*=.04), indicating that medical students in the nonindividualized group experienced improved well-being following intervention engagement compared with students who received no intervention.

Psychological Flexibility and Inflexibility

The estimated increase in psychological flexibility between baseline and postintervention measurement was 0.50 points greater for the individualized group than for the waitlist group (0.26), and this difference was significant (P=.01). Similarly, the estimated decrease in psychological inflexibility between baseline and postintervention measurement was significantly greater for the individualized group (−0.53) than for the waitlist group (−0.28; P=.04). These findings indicate that medical students in the individualized group experienced improvements in both psychological flexibility (Figure 2) and psychological inflexibility following engagement in the intervention compared with students who received no intervention.

Figure 2. Mean psychological flexibility scores by intervention group at baseline and postintervention measurement.



Depression

The estimated change in depression between baseline and postintervention measurement was significantly greater for the individualized group (−4.40) than for the waitlist group (−0.54; P=.046). This indicates that depressive symptoms reduced significantly more for students who engaged in the individualized version of the app than for those who received no intervention.

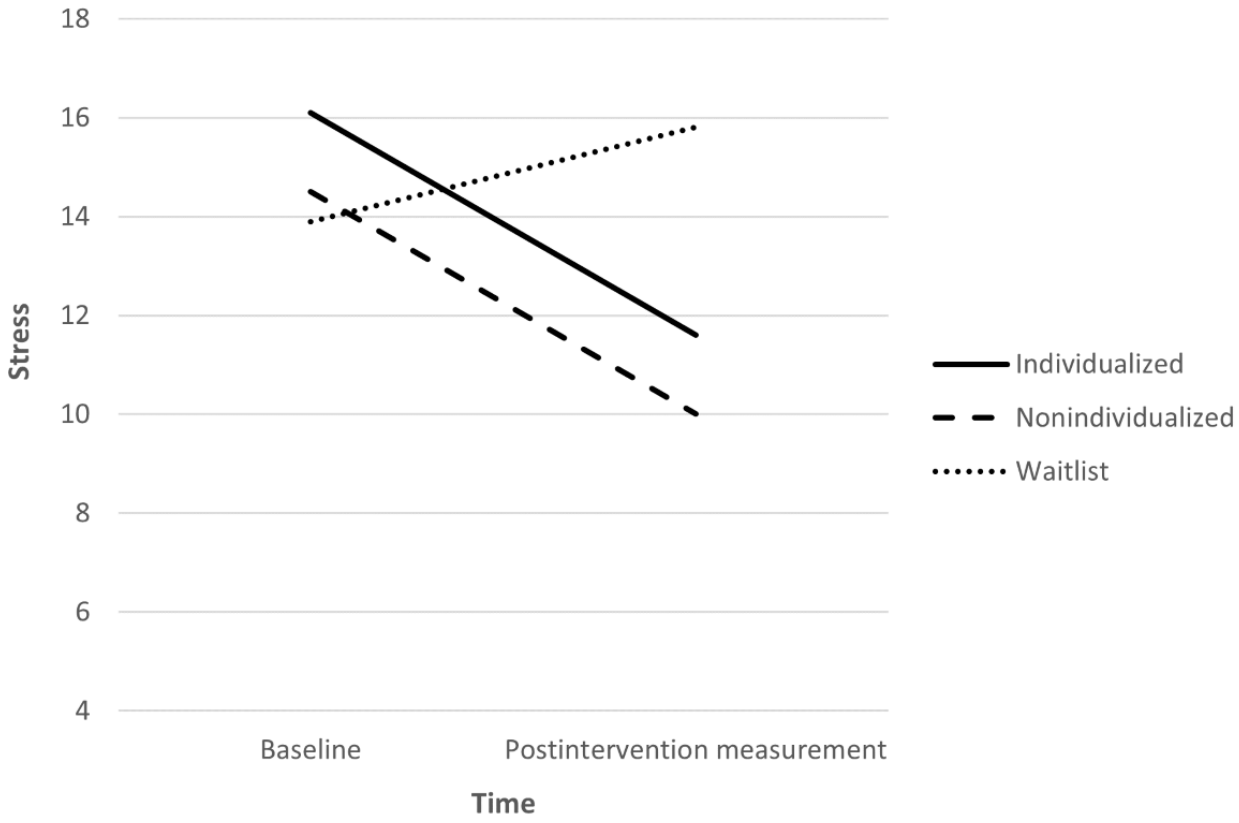
Anxiety

There were no statistically significant differences between the intervention arms and the control group for anxiety.

Stress

There was a significant difference in the estimated change in stress between baseline and postintervention measurement for both the individualized (P=.01; Figure 3) and nonindividualized (P=.03) groups compared with the waitlist group (individualized=−4.65; nonindividualized=−4.12; waiting list=2.24), indicating that stress reduced significantly more for participants in both intervention groups than for participants in the waitlist group.

Figure 3. Mean stress scores by intervention group at baseline and postintervention measurement.



Individualized Versus Nonindividualized Group Comparisons

There were no statistically significant differences between the individualized and nonindividualized arms in any of the psychological outcomes assessed.

Mediation Analyses

Although the study was statistically powered to conduct psychological flexibility mediation analyses for the primary outcome (exhaustion), given the failure to observe a meaningful intervention effect for this outcome, formal mediation analyses of exhaustion and the secondary psychological outcomes would have been too exploratory to have any confidence in the results.

Study Attrition, Intervention Adherence, and Engagement

A high rate of attrition (75/143, 52.4%) was observed during this study (Multimedia Appendix 2). Almost half (35/75, 47%) of those lost to follow-up discontinued before commencing the intervention and were not randomized to a study group. Only 1 participant formally withdrew, citing perceived personal irrelevance of the baseline outcome measures as the reason. Students lost to follow-up did not differ significantly from those who completed the study with respect to baseline demographic or psychological characteristics (Multimedia Appendix 3).

Of the participants allocated to either the individualized or nonindividualized group, 63% (46/73) completed stage 1, and 29% (21/73) met the predefined adherence criteria (ie, completion of at least 4 stage 2 activities; Multimedia Appendix 4). Intervention adherence rates did not differ by group allocation (N=73, χ^2_1=3.0; *P*=.08). The only baseline participant characteristics associated with adherence rates were *age*—students who adhered to the study were older on average (mean 27.5, SD 7.9 years) than those who did not (mean 23.3, SD 4.0 years; t_{70}=−2.31; *P*=.03)—and *medicine as a first career*—significantly lower adherence rates among students studying medicine as a first career (10/51, 20%) than among those with previous career experience in another field (11/21, 52%; N=72, χ^2_1=7.7; *P*=.005; Multimedia Appendix 5).

Participants completed a total of 255 skill activities during stage 2, with those in the individualized group completing an average of 6 (SD 5.7) activities and those in the nonindividualized group completing an average of 8 (SD 12.4) activities. This difference was not significant (t_{36}=−0.53; *P*=.60). Of note, 3 participants accounted for 38.8% (99/255) of all skill activities completed during stage 2. With these participants removed, the average number of activities completed was 5 (SD 3.8) for the individualized group and 3 (SD 2.1) for the nonindividualized group. This between-group difference was not significant (t_{33}=1.69; *P*=.10).

Intervention Harms and Feedback

No harms were reported during the study. Participants who used the in-app feedback form (15/143, 10.5%) reported finding the intervention content interesting and helpful in relation to their psychological health and well-being. Some reported that, despite finding the app helpful, time was a key barrier to engagement. Other feedback included requests for additional usability

functions, such as dark mode and written versions of activities that were only presented in audio format. Students in the intervention groups liked 87.1% (222/255) of the skill activities completed during stage 2 and disliked 12.9% (33/255).

Discussion

Intervention Outcomes

This is the first RCT to evaluate individualized and nonindividualized versions of a smartphone app–delivered psychological flexibility skill training (ACT) intervention for medical students. Although we did not demonstrate immediate postintervention improvements in burnout outcomes among medical students who used the app, the observed improvements in psychological flexibility, well-being, and stress outcomes provide promising support for this intervention approach.

On the basis of the existing literature, we anticipated that training medical students in psychological flexibility skills using an app-delivered ACT intervention would lead to improvements in burnout. Primary (exhaustion) and secondary (cynicism and academic efficacy) burnout outcomes for the sample in this study were not significantly affected by engagement in either version of the ACT app at the end of the 5-week intervention access period. We selected exhaustion as the primary burnout outcome because of its high reported prevalence among medical students [9] and the possible longer-term burnout prevention benefits associated with this factor's early responsivity to psychological flexibility interventions [42]. However, baseline exhaustion and cynicism scores for this sample were suggestive of low levels of burnout compared with reference medical student samples [66], which may have produced a floor effect with respect to the potential for improving these outcomes. Previous studies assessing the benefits of web-based ACT interventions were less likely to demonstrate intervention effects when the samples under investigation demonstrated nonclinical baseline levels of the psychological outcome under investigation [96].

Interestingly, medical students in this sample demonstrated high baseline levels of stress compared with a general population sample [90], and our exploratory secondary analyses showed significant postintervention improvements in this outcome for both the individualized and nonindividualized groups. As stress is a risk factor for burnout among medical students [63], it is possible that stress reduction might have been an early indicator of ACT intervention efficacy and that medical students may require longer to implement psychological flexibility skills in adaptive ways in their everyday lives before experiencing benefits for the more chronic and distal state of burnout. This is supported by the results of a recent RCT with a heterogeneous employee sample, which demonstrated that reduced stress was an immediate response to an 8-week web-based ACT intervention, whereas improvements in burnout became more evident 1 year later [36]. Similarly, systematic review findings indicate that mitigation of early stress experiences could be an important first step for ACT interventions with respect to improving burnout or preventing its development [97]. Future investigation of this theory regarding the current ACT app will require the prospective evaluation of stress as a primary outcome and replication of these study findings, as well as longer-term follow-up of burnout outcomes. This could provide greater clarity regarding the potential importance of psychological flexibility processes in facilitating adaptive responses to stress in medical students and the possible impact on burnout prevention.

As psychological flexibility is associated with well-being [52,53], we expected that engaging in an ACT intervention would improve this outcome among medical students. Medical students in the nonindividualized group experienced significant postintervention improvements in well-being compared with the waitlist group. To our knowledge, this is the first RCT to demonstrate the efficacy of an ACT intervention for medical student well-being and the first to show that well-being benefits can be achieved among medical students using an app-based mode of delivery. These findings suggest that providing medical students with access to ACT skill training interventions could support personal thriving and positive psychological health, which is a key objective of medical educators [38]. Importantly, previous research has shown that improvements in well-being can offset adverse consequences of burnout (eg, unprofessional behaviors and suicidal ideation) [95,98]. Furthermore, although traditional ACT interventions can require considerable time commitment [99-101], the observed well-being outcomes followed a relatively brief app-based intervention, which is an important consideration for this (and any other) time-poor cohort [64,66]. Future research is needed to replicate these secondary findings and explore whether continued intervention engagement is required to maintain well-being benefits in the long term.

It is unclear why only the nonindividualized group experienced significant improvements in well-being given that both versions of the ACT app delivered similar training content and that the individualized intervention was expected to produce stronger outcomes. This discrepancy is not attributable to engagement as there were no significant between-group differences in intervention adherence. Although we were unable to evaluate mediation effects for well-being outcomes among nonindividualized participants, we note that psychological flexibility and inflexibility did not improve significantly in this group. It is possible that unassessed mediating factors may have contributed to the observed inconsistencies [102], and future studies are required to evaluate this further to better understand the processes underlying the well-being benefits of this ACT-based app.

Following the intervention, medical students who engaged in the individualized version of the app demonstrated increased psychological flexibility and reduced psychological inflexibility. This is the first study to show that medical students can be trained to improve psychological flexibility and inflexibility using an individualized ACT intervention delivered via an app. Although there was no corresponding improvement in burnout and well-being outcomes for this intervention group, there is an extensive literature base demonstrating that this adaptive skill set is broadly beneficial to individual health and functioning [50] (including anxiety, depression, physical health outcomes, satisfaction with life, and work performance [47,72,103]). In fact, psychological flexibility has recently been described as "the cornerstone of psychological health and resiliency" [104].

Demonstrating that this skill set can be trained using an accessible digital delivery methodology and a low time commitment could be an important step toward developing interventions that can improve medical students' psychological health and functioning more broadly. Our exploratory findings hinted at the potential wider benefits of improving medical students' psychological flexibility, showing that depressive symptoms among those in the individualized group also improved following engagement in the intervention. Longer-term studies are needed to evaluate whether improving medical students' psychological flexibility and inflexibility is beneficial to distal burnout outcomes or whether these processes might have greater relevance to other important psychological outcomes (eg, well-being, stress, and depression).

Although medical students who engaged in either the individualized or nonindividualized intervention demonstrated improvements in at least one psychological outcome compared with those in the waitlist group, unlike Levin et al [75], we did not observe statistically significant differences between the 2 intervention groups for any of the outcomes assessed. However, in addition to being underpowered to detect these effects, it is possible that methodological factors obfuscated our ability to evaluate the true potential treatment utility of individualized over nonindividualized psychological flexibility skill training. Although our individualization method was similar to that of Levin et al [75], students from both groups in this study were required to complete the same introductory module before progressing to their respective individualized or nonindividualized skill training. Our intention-to-treat analysis approach meant that students were evaluated based on the group they were allocated to rather than the degree to which they actually engaged in the intervention. Owing to the high level of nonadherence, both groups contained participants who either did not reach the point where the intervention methodologies diverged or did not engage sufficiently in the skill training stage for individualization to have a meaningful impact. Given this considerable overlap between the intervention groups and the likelihood that our study did not adequately differentiate between individualized and nonindividualized conditions, these nonsignificant findings should be interpreted cautiously. The refinement of individualized ACT interventions is important for progressing key goals within the psychological flexibility literature, including improving precision and effectiveness for all individuals who access training [69,70]. Furthermore, ensuring the efficiency of psychological interventions is particularly important for medical students because of the impact of time constraints on engagement [64,66]. Comparing differences in outcomes between an EMA-driven individualized app and a nonindividualized condition offers a number of advantages to future research, including the facilitation of empirical evaluation of whether individualized ACT interventions do strengthen outcomes, providing a study design framework that facilitates continued optimization of individualization methods and furthering our understanding of how specific intervention components drive changes in adaptive psychological processes [69,70]. Future app-based ACT research adopting this study design should ensure that individualized and nonindividualized interventions are appropriately distinct and that sample sizes are large enough to detect potential between-group differences.

Adherence

Adherence to the app was an observed challenge in this early-phase study. Only two-thirds of the intervention group participants (46/73, 63%) completed the introductory module, fewer than one-third (21/73, 29%) met the adherence criteria, and overall study attrition was high. Low adherence has the potential to undermine the feasibility and scalability of this app-delivered intervention approach, and understanding the contributing factors is essential for effective development and implementation [63]. Unlike previous digital intervention studies [92], the baseline psychological characteristics of the medical student sample in this study did not significantly affect attrition or adherence to the ACT app. Older students and those who had previously worked in a career other than medicine were more likely to adhere, and it is possible that these students perceived the intervention to be more relevant or necessary because of previous life or work experiences. "Like" data from stage 2 suggest that medical students who remained engaged in this study were satisfied with the content of the ACT intervention, which may have motivated continued participation among these students. As immediate feedback for stage 1 activities was not collected, it is unclear whether early discontinuation might have been affected by satisfaction with the initial intervention content. However, the few students who used the available feedback forms during stage 1 reported favorable experiences regarding content but cited lack of time and usability factors as engagement and adherence barriers.

Although previous research suggests that medical student engagement in digital interventions may be improved by the incorporation of face-to-face training components [63], the stand-alone app–based mode of delivery was adopted because of uncertainty associated with COVID-19 restrictions at the time of development. Given the likelihood of ongoing accessibility barriers for psychological interventions among medical students [32] and the potential reach of this approach, future research should incorporate a formal evaluation of the user experiences contributing to engagement and disengagement [63] to develop stronger and more engaging implementation frameworks for stand-alone apps.

Strengths and Limitations

This study had several strengths, including its randomized controlled design; assessment of a theory-driven intervention specifically developed for medical students; incorporation of medical student end-user feedback before implementation; evaluation among students from more than one university; and its app-based delivery, which facilitated access and implementation during a COVID-19 lockdown. Although we were underpowered to detect differences between the individualized and nonindividualized groups, the use of technologies that can adapt interventions to individual needs and the adoption of research methodologies that assess the treatment utility of these interventions are important to the psychological flexibility literature and should be explored further [69].

A strength of this study was the separate evaluation of psychological flexibility and inflexibility [45], which are often measured as opposite ends of a single-factor construct [105]. Although many individuals who are high in psychological flexibility demonstrate correspondingly low inflexibility, some individuals may show high levels of both simultaneously, whereas others show mixed patterns across all core processes [46]. Given that psychological flexibility and inflexibility appear to be conceptually distinct processes that may have disparate relationships with well-being and psychological distress outcomes and may respond differently to interventions [45,46,72], evaluating them separately provides a stronger understanding of intervention methodologies that can address this heterogeneity [70,71]. This demonstration of postintervention improvements in both psychological flexibility and inflexibility following engagement in the individualized ACT app provides promising support for this approach with respect to the simultaneous strengthening of adaptive behavioral repertoires and undermining of maladaptive repertoires among medical students.

Although we made the best efforts to calculate and recruit a sufficient sample size, the absence of any previous within-person correlation data, together with the high attrition rate, represents a limitation of this study, meaning that we may have been underpowered to detect changes in the primary burnout outcome (exhaustion). Furthermore, although the observed intervention effects for secondary outcomes are promising, the small sample size and number of between-group comparisons mean that there is a risk of type-I and type-II errors. Further research is needed to determine whether the findings of this early-phase study can be replicated in larger samples of medical students.

Intervention outcomes should also be interpreted with caution because of the potential for overestimation of effect sizes when a waitlist comparison group is used in behavior change research [106]. The waitlist group was a necessary ethical component of this study as stakeholders from the participating university medical schools wanted to ensure that all interested students were offered access to the intervention app as a well-being resource so that no students would be disadvantaged by allocation to the control condition. However, previous research suggests that waitlist participants' expectations of receiving support at a later time could hinder the self-driven behavior change that they might otherwise have made during the same period [106]. We note that students allocated to the waiting list in this study tended not to engage with the app once they were granted access. Future replication studies might adopt an active control condition in which participants are encouraged to check in to the app regularly during the study period (eg, to complete the psychological flexibility EMA screening question without exposure to the intervention content) [75,107]. This could reduce the likelihood of disrupting any natural improvements that control group participants might make during the study period as well as facilitating subsequent access to the intervention by promoting continued engagement from the point of students' initial interest.

Although our intervention effect analyses adjusted for participant demographic characteristics that were identified as imbalanced after randomization (gender and studying medicine as a first career), there were also slight imbalances in other factors (current therapy and year of study) that may have affected outcomes. Depending on the therapeutic approach, participation in psychological treatment while using the app could potentially support or contradict the goals of the ACT intervention. Similarly, the app might affect psychological outcomes differently during the earlier academic stages of medical education compared with the applied clinical stages. Although stratification or matched-pair sampling could have mitigated group imbalances, the functional limitations of the current app did not facilitate these approaches. Ideally, improvements in the technology could support this methodology in the future. However, using the current randomization method, future research with larger samples should resolve such imbalances and provide a clearer picture of the impact of the intervention on outcomes in heterogeneous medical student samples. Alternatively, eligibility criteria could be modified to select more homogeneous samples based on participant demographic characteristics that are identified as having the potential to affect outcomes.

The generalizability of these findings may be limited given the predominantly nonindigenous Australian sample with a slight skew toward female participants. However, research suggests that female medical students may be at greater risk of burnout [34,66], and thus, intervention studies may be particularly relevant to female participants.

Implications and Future Directions

The aim of this RCT was to evaluate the effectiveness of individualized and nonindividualized versions of an app-delivered ACT intervention with the broader goal of contributing data that could inform future early intervention strategies for reducing burnout and improving well-being among medical students. Although we did not observe significant postintervention benefits for primary (exhaustion) or secondary (cynicism and academic efficacy) burnout outcomes, analyses of other relevant secondary outcomes provided promising preliminary support for the ACT intervention approach for medical students. We demonstrated that an ACT intervention may improve factors associated with future burnout development, including stress and well-being, as well as other important psychological health outcomes (ie, depression). These findings support previous literature highlighting the potential benefits of training medical students in psychological flexibility skills [31,61,62] and the value of further investigation of this approach. Replication and longitudinal studies are needed to further clarify the role that psychological flexibility and inflexibility interventions (eg, ACT) may play in improving well-being and mitigating the adverse impacts of stress within the medical profession, including burnout. Although we did not observe differences in efficacy between the individualized and nonindividualized versions of the app, the individualized intervention was beneficial for a greater number of outcomes. Further research addressing the methodological limitations observed in this study could progress key psychological flexibility literature goals related to individual heterogeneity and intervention precision.

Our findings further contribute to the nascent field of app-delivered psychological interventions for medical students [21] by demonstrating that a smartphone app could be used to improve important psychological health and well-being outcomes. This is also the first study to demonstrate that delivering an ACT intervention using an app could improve medical students' psychological flexibility and inflexibility, which is an important finding given the broad relevance of these processes to psychological health and effective functioning [50]. App-delivered approaches provide medical students with an intervention that they can access in brief segments at a time and place that suits them, using a technology medium with which they are comfortable [63] and that allows them to maintain privacy [11,32,64]. They also offer stakeholders a cost-effective scalable intervention option that could reach a large number of students [39,64] while maintaining relevance to individual needs [75]. Although further research is needed to improve engagement and establish optimal individualization methods, this early implementation study provides promising support for the potential benefits of an ACT-based app in strengthening medical students' adaptive psychological skills and improving psychological health and well-being outcomes.

Successful burnout prevention and well-being interventions likely require a combination of organizational and individual resource-building strategies [1]. Although the current app was delivered within an organizational context and supported by the universities in which it was implemented, intentional efforts were made to distance the intervention from the students' university programs to ensure that they did not feel compelled to participate. However, increased organizational efforts to support participation in a psychological skill development app may have an important impact on normalizing and prioritizing commitment to personal well-being during the early and formative stages of a medical career [26,108]. Given the preliminary support for the benefits of this ACT intervention, future implementation studies might examine whether embedding the app into broader university-based burnout and well-being initiatives strengthens adherence and outcomes. We note that, despite the engagement challenges observed in this study, the students who participated experienced benefits across a range of key psychological health outcomes. Thus, although not necessarily a solution for all medical students, a psychological flexibility (ACT) skill training app could form part of a suite of well-being options provided by educational institutions.

Acknowledgments

This research project was funded by the Centre for Rehab Innovations and the Priority Research Centre for Stroke and Brain Injury at the University of Newcastle. The authors wish to acknowledge the staff and students from the Joint Medical Program (University of Newcastle and University of New England) for participating in the study and Cogniss for assistance in developing the app. Thanks to Lucy Bailey for providing instrumental support regarding administrative aspects of the study.

Authors' Contributions

ED wrote the manuscript, designed the study protocol, created the app, ran the intervention trial, and conducted data analyses. ED and BK wrote the intervention content. NH provided consultation on protocol design, power analyses, and app usability. GH led stakeholder engagement and provided consultation on the translation of aspects of the protocol and intervention to the cohort context. CO provided consultation and guidance regarding data analyses and interpretation. All authors reviewed and contributed to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline psychological characteristics of study sample (Mean, SD) and comparisons with published reference samples (single-sample *t* tests and *P* values).
[DOCX File , 20 KB - mental_v10i1e42566_app1.docx]

Multimedia Appendix 2
Study attrition rates for enrolled participants (N=143).
[DOCX File , 14 KB - mental_v10i1e42566_app2.docx]

Multimedia Appendix 3
Demographic and psychological characteristics of participants lost to follow-up (LTF) and participants who completed the study; and between-group comparisons (chi-square or independent samples *t* tests and *P* values).
[DOCX File , 23 KB - mental_v10i1e42566_app3.docx]

Multimedia Appendix 4
Intervention completion and adherence rates by study group (N=73).

[DOCX File , 14 KB - mental_v10i1e42566_app4.docx]

Multimedia Appendix 5
Demographic and psychological characteristics of participants who met adherence criteria and participants who did not; and between-group comparisons (chi-square or independent samples *t* tests and *P* values).
[DOCX File , 25 KB - mental_v10i1e42566_app5.docx]

Multimedia Appendix 6
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1510 KB - mental_v10i1e42566_app6.pdf]

References

1. West CP, Dyrbye LN, Erwin PJ, Shanafelt TD. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. Lancet 2016 Nov 05;388(10057):2272-2281. [doi: 10.1016/S0140-6736(16)31279-X] [Medline: 27692469]
2. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Arch Intern Med 2012 Oct 08;172(18):1377-1385. [doi: 10.1001/archinternmed.2012.3199] [Medline: 22911330]
3. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, et al. Prevalence of burnout among physicians: a systematic review. JAMA 2018 Sep 18;320(11):1131-1150 [FREE Full text] [doi: 10.1001/jama.2018.12777] [Medline: 30326495]
4. Bradford L, Glaser G. Addressing physician burnout and ensuring high-quality care of the physician workforce. Obstet Gynecol 2021 Jan 01;137(1):3-11. [doi: 10.1097/AOG.0000000000004197] [Medline: 33278277]
5. Amanullah S, Ramesh Shankar R. The impact of COVID-19 on physician burnout globally: a review. Healthcare (Basel) 2020 Oct 22;8(4):421 [FREE Full text] [doi: 10.3390/healthcare8040421] [Medline: 33105757]
6. Maslach C, Leiter MP. Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry 2016 Jun;15(2):103-111 [FREE Full text] [doi: 10.1002/wps.20311] [Medline: 27265691]
7. Leiter MP, Maslach C. Latent burnout profiles: a new approach to understanding the burnout experience. Burn Res 2016 Dec;3(4):89-100. [doi: 10.1016/j.burn.2016.09.001]
8. Ishak W, Nikravesh R, Lederer S, Perry R, Ogunyemi D, Bernstein C. Burnout in medical students: a systematic review. Clin Teach 2013 Aug;10(4):242-245. [doi: 10.1111/tct.12014] [Medline: 23834570]
9. Frajerman A, Morvan Y, Krebs MO, Gorwood P, Chaumette B. Burnout in medical students before residency: a systematic review and meta-analysis. Eur Psychiatry 2019 Jan;55:36-42. [doi: 10.1016/j.eurpsy.2018.08.006] [Medline: 30384110]
10. Dyrbye L, Shanafelt T. A narrative review on burnout experienced by medical students and residents. Med Educ 2016 Jan;50(1):132-149. [doi: 10.1111/medu.12927] [Medline: 26695473]
11. Fares J, Al Tabosh H, Saadeddin Z, El Mouhayyar C, Aridi H. Stress, burnout and coping strategies in preclinical medical students. N Am J Med Sci 2016 Feb;8(2):75-81 [FREE Full text] [doi: 10.4103/1947-2714.177299] [Medline: 27042604]
12. Salvagioni DA, Melanda FN, Mesas AE, González AD, Gabani FL, Andrade SM. Physical, psychological and occupational consequences of job burnout: a systematic review of prospective studies. PLoS One 2017 Oct 4;12(10):e0185781 [FREE Full text] [doi: 10.1371/journal.pone.0185781] [Medline: 28977041]
13. Melnikow J, Padovani A, Miller M. Frontline physician burnout during the COVID-19 pandemic: national survey findings. BMC Health Serv Res 2022 Mar 19;22(1):365 [FREE Full text] [doi: 10.1186/s12913-022-07728-6] [Medline: 35303889]
14. Kuhn CM, Flanagan EM. Self-care as a professional imperative: physician burnout, depression, and suicide. Can J Anaesth 2017 Feb;64(2):158-168. [doi: 10.1007/s12630-016-0781-0] [Medline: 27910035]
15. Kumar S. Burnout and doctors: prevalence, prevention and intervention. Healthcare (Basel) 2016 Jun 30;4(3):37 [FREE Full text] [doi: 10.3390/healthcare4030037] [Medline: 27417625]
16. Shoimer I, Patten S, Mydlarski PR. Burnout in dermatology residents: a Canadian perspective. Br J Dermatol 2018 Jan;178(1):270-271. [doi: 10.1111/bjd.15549] [Medline: 28374946]
17. Dyrbye LN, Thomas MR, Massie FS, Power DV, Eacker A, Harper W, et al. Burnout and suicidal ideation among U.S. medical students. Ann Intern Med 2008 Sep 02;149(5):334-341. [doi: 10.7326/0003-4819-149-5-200809020-00008] [Medline: 18765703]
18. Shanafelt TD, Mungo M, Schmitgen J, Storz KA, Reeves D, Hayes SN, et al. Longitudinal study evaluating the association between physician burnout and changes in professional work effort. Mayo Clin Proc 2016 Apr;91(4):422-431. [doi: 10.1016/j.mayocp.2016.02.001] [Medline: 27046522]
19. Leo CG, Sabina S, Tumolo MR, Bodini A, Ponzini G, Sabato E, et al. Burnout among healthcare workers in the COVID 19 era: a review of the existing literature. Front Public Health 2021 Oct 29;9:750529 [FREE Full text] [doi: 10.3389/fpubh.2021.750529] [Medline: 34778184]

20. de Oliva Costa EF, Santos Andrade S, de Abreu Santos AT, de Melo EV, de Andrade TM. Burnout Syndrome and associated factors among medical students: a cross-sectional study. Clinics (Sao Paulo) 2012;67(6):573-580 [FREE Full text] [doi: 10.6061/clinics/2012(06)05] [Medline: 22760894]
21. Seo C, Corrado M, Fournier K, Bailey T, Haykal KA. Addressing the physician burnout epidemic with resilience curricula in medical education: a systematic review. BMC Med Educ 2021 Feb 01;21(1):80 [FREE Full text] [doi: 10.1186/s12909-021-02495-0] [Medline: 33526019]
22. Amanullah S, McNally K, Zelin J, Cole J, Cernovsky Z. Are burnout prevention programs for hospital physicians needed? Asian J Psychiatr 2017 Apr;26:66-69. [doi: 10.1016/j.ajp.2017.01.009] [Medline: 28483095]
23. Gajjar J, Pullen N, Li Y, Weir S, Wright JG. Impact of the COVID-19 pandemic upon self-reported physician burnout in Ontario, Canada: evidence from a repeated cross-sectional survey. BMJ Open 2022 Sep 21;12(9):e060138 [FREE Full text] [doi: 10.1136/bmjopen-2021-060138] [Medline: 36130759]
24. Chakladar J, Diomino A, Li WT, Tsai JC, Krishnan AR, Zou AE, et al. Medical student's perception of the COVID-19 pandemic effect on their education and well-being: a cross-sectional survey in the United States. BMC Med Educ 2022 Mar 05;22(1):149 [FREE Full text] [doi: 10.1186/s12909-022-03197-x] [Medline: 35248030]
25. De Simone S, Vargas M, Servillo G. Organizational strategies to reduce physician burnout: a systematic review and meta-analysis. Aging Clin Exp Res 2021 Apr;33(4):883-894. [doi: 10.1007/s40520-019-01368-3] [Medline: 31598914]
26. Mian A, Kim D, Chen D, Ward WL. Medical student and resident burnout: a review of causes, effects, and prevention. J Fam Med Dis Prev 2018 Dec 31;4(4):094. [doi: 10.23937/2469-5793/1510094]
27. Williams D, Tricomi G, Gupta J, Janise A. Efficacy of burnout interventions in the medical education pipeline. Acad Psychiatry 2015 Feb;39(1):47-54. [doi: 10.1007/s40596-014-0197-5] [Medline: 25034955]
28. Walsh AL, Lehmann S, Zabinski J, Truskey M, Purvis T, Gould NF, et al. Interventions to prevent and reduce burnout among undergraduate and graduate medical education trainees: a systematic review. Acad Psychiatry 2019 Aug;43(4):386-395. [doi: 10.1007/s40596-019-01023-z] [Medline: 30710229]
29. Zis P, Artemiadis A, Bargiotas P, Nteveros A, Hadjigeorgiou GM. Medical studies during the COVID-19 pandemic: the impact of digital learning on medical students' burnout and mental health. Int J Environ Res Public Health 2021 Jan 05;18(1):349 [FREE Full text] [doi: 10.3390/ijerph18010349] [Medline: 33466459]
30. Bugaj TJ, Cranz A, Junne F, Erschens R, Herzog W, Nikendei C. Psychosocial burden in medical students and specific prevention strategies. Ment Health Prev 2016 Mar;4(1):24-30. [doi: 10.1016/j.mhp.2015.12.003]
31. Palladino CL, Ange B, Richardson DS, Casillas R, Decker M, Gillies RA, et al. Measuring psychological flexibility in medical students and residents: a psychometric analysis. Med Educ Online 2013 Aug 13;18:20932 [FREE Full text] [doi: 10.3402/meo.v18i0.20932] [Medline: 23948496]
32. Thompson G, McBride RB, Hosford CC, Halaas G. Resilience among medical students: the role of coping style and social support. Teach Learn Med 2016;28(2):174-182. [doi: 10.1080/10401334.2016.1146611] [Medline: 27064719]
33. Arif NM, Roslan NS, Ismail SB, Nayak RD, Jamian MR, Mohamad Ali Roshidi AS, et al. Prevalence and associated factors of psychological distress and burnout among medical students: findings from two campuses. Int J Environ Res Public Health 2021 Aug 10;18(16):8446 [FREE Full text] [doi: 10.3390/ijerph18168446] [Medline: 34444193]
34. Voltmer E, Rosta J, Aasland OG, Spahn C. Study-related health and behavior patterns of medical students: a longitudinal study. Med Teach 2010;32(10):e422-e428. [doi: 10.3109/0142159X.2010.496008] [Medline: 20854148]
35. Thun-Hohenstein L, Höbinger-Ablasser C, Geyerhofer S, Lampert K, Schreuer M, Fritz C. Burnout in medical students. Neuropsychiatr 2021 Mar;35(1):17-27 [FREE Full text] [doi: 10.1007/s40211-020-00359-5] [Medline: 32880881]
36. Puolakanaho A, Tolvanen A, Kinnunen SM, Lappalainen R. A psychological flexibility -based intervention for burnout:a randomized controlled trial. J Contextual Behav Sci 2020 Jan;15:52-67. [doi: 10.1016/j.jcbs.2019.11.007]
37. Dunn LB, Iglewicz A, Moutier C. A conceptual model of medical student well-being: promoting resilience and preventing burnout. Acad Psychiatry 2008;32(1):44-53. [doi: 10.1176/appi.ap.32.1.44] [Medline: 18270280]
38. Kemp S, Hu W, Bishop J, Forrest K, Hudson JN, Wilson I, et al. Medical student wellbeing - a consensus statement from Australia and New Zealand. BMC Med Educ 2019 Mar 04;19(1):69 [FREE Full text] [doi: 10.1186/s12909-019-1505-2] [Medline: 30832630]
39. Harvey SB, Epstein RM, Glozier N, Petrie K, Strudwick J, Gayed A, et al. Mental illness and suicide among physicians. Lancet 2021 Sep 04;398(10303):920-930 [FREE Full text] [doi: 10.1016/S0140-6736(21)01596-8] [Medline: 34481571]
40. Lenoir AL, De Troyer C, Demoulin C, Gillain I, Bayot M. Challenges in treating physician burnout: the psychologist's perspective. Press Méd Open 2021 Jun;2:100006. [doi: 10.1016/j.lpmope.2021.100006]
41. Wersebe H, Lieb R, Meyer AH, Hofer P, Gloster AT. The link between stress, well-being, and psychological flexibility during an Acceptance and Commitment Therapy self-help intervention. Int J Clin Health Psychol 2018;18(1):60-68 [FREE Full text] [doi: 10.1016/j.ijchp.2017.09.002] [Medline: 30487911]
42. Lloyd J, Bond FW, Flaxman PE. The value of psychological flexibility: examining psychological mechanisms underpinning a cognitive behavioural therapy intervention for burnout. Work Stress 2013 Apr;27(2):181-199. [doi: 10.1080/02678373.2013.782157]

43. Kinnunen SM, Puolakanaho A, Mäkikangas A, Tolvanen A, Lappalainen R. Does a mindfulness-, acceptance-, and value-based intervention for burnout have long-term effects on different levels of subjective well-being? Int J Stress Manag 2020 Feb;27(1):82-87. [doi: 10.1037/str0000132]
44. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: The Process and Practice of Mindful Change. 2nd edition. New York, NY, USA: Guilford Press; 2012.
45. Rolffs JL, Rogge RD, Wilson KG. Disentangling components of flexibility via the hexaflex model: development and validation of the multidimensional psychological flexibility inventory (MPFI). Assessment 2018 Jun;25(4):458-482. [doi: 10.1177/1073191116645905] [Medline: 27152011]
46. Stabbe OK, Rolffs JL, Rogge RD. Flexibly and/or inflexibly embracing life: identifying fundamental approaches to life with latent profile analyses on the dimensions of the Hexaflex model. J Contextual Behav Sci 2019 Apr;12:106-118. [doi: 10.1016/j.jcbs.2019.03.003]
47. Onwezen MC, van Veldhoven MJ, Biron M. The role of psychological flexibility in the demands–exhaustion–performance relationship. Eur J Work Organ Psychol 2014;23(2):163-176. [doi: 10.1080/1359432x.2012.742242]
48. Moran DJ. ACT for leadership: using acceptance and commitment training to develop crisis-resilient change managers. Int J Behav Consult Ther 2010 Jan;6(4):341-355. [doi: 10.1037/h0100915]
49. Moran DJ. Acceptance and Commitment Training in the workplace. Curr Opin Psychol 2015 Apr;2:26-31. [doi: 10.1016/j.copsyc.2014.12.031]
50. Hayes S. ACT Randomized Controlled Trials (1986 to Present). Association for Contextual Behavioral Science. 2022. URL: https://contextualscience.org/act_randomized_controlled_trials_1986_to_present: [accessed 2023-01-20]
51. Landi G, Pakenham KI, Crocetti E, Tossani E, Grandi S. The trajectories of anxiety and depression during the COVID-19 pandemic and the protective role of psychological flexibility: a four-wave longitudinal study. J Affect Disord 2022 Jun 15;307:69-78 [FREE Full text] [doi: 10.1016/j.jad.2022.03.067] [Medline: 35378147]
52. Gloster AT, Meyer AH, Lieb R. Psychological flexibility as a malleable public health target: evidence from a representative sample. J Contextual Behav Sci 2017 Apr;6(2):166-171. [doi: 10.1016/j.jcbs.2017.02.003]
53. Puolakanaho A, Tolvanen A, Kinnunen SM, Lappalainen R. Burnout-related ill-being at work: associations between mindfulness and acceptance skills, worksite factors, and experienced well-being in life. J Contextual Behav Sci 2018 Oct;10:92-102. [doi: 10.1016/j.jcbs.2018.09.003]
54. Bond FW, Bunce D. Mediators of change in emotion-focused and problem-focused worksite stress management interventions. J Occup Health Psychol 2000 Jan;5(1):156-163. [doi: 10.1037//1076-8998.5.1.156] [Medline: 10658893]
55. Flaxman PE, Bond FW. A randomised worksite comparison of acceptance and commitment therapy and stress inoculation training. Behav Res Ther 2010 Aug;48(8):816-820. [doi: 10.1016/j.brat.2010.05.004] [Medline: 20627269]
56. Frögéli E, Djordjevic A, Rudman A, Livheim F, Gustavsson P. A randomized controlled pilot trial of acceptance and commitment training (ACT) for preventing stress-related ill health among future nurses. Anxiety Stress Coping 2016;29(2):202-218. [doi: 10.1080/10615806.2015.1025765] [Medline: 25759942]
57. Kinnunen SM, Puolakanaho A, Tolvanen A, Mäkikangas A, Lappalainen R. Does mindfulness-, acceptance-, and value-based intervention alleviate burnout?—A person-centered approach. Int J Stress Manag 2019 Feb;26(1):89-101. [doi: 10.1037/str0000095]
58. Johns G, Waddington L, Samuel V. Prevalence and predictors of mental health outcomes in UK doctors and final year medical students during the COVID-19 pandemic. J Affect Disord 2022 Aug 15;311:267-275 [FREE Full text] [doi: 10.1016/j.jad.2022.05.024] [Medline: 35569608]
59. Solms L, van Vianen AE, Theeboom T, Koen J, de Pagter AP, de Hoog M. Keep the fire burning: a survey study on the role of personal resources for work engagement and burnout in medical residents and specialists in the Netherlands. BMJ Open 2019 Nov 05;9(11):e031053 [FREE Full text] [doi: 10.1136/bmjopen-2019-031053] [Medline: 31694848]
60. Kroska EB, Calarge C, O'Hara MW, Deumic E, Dindo L. Burnout and depression in medical students: relations with avoidance and disengagement. J Contextual Behav Sci 2017 Oct;6(4):404-408. [doi: 10.1016/j.jcbs.2017.08.003]
61. Szarko AJ, Houmanfar RA, Smith GS, Jacobs NN, Smith BM, Assemi K, et al. Impact of Acceptance and Commitment Training on psychological flexibility and burnout in medical education. J Contextual Behav Sci 2022 Jan;23:190-199. [doi: 10.1016/j.jcbs.2022.02.004]
62. Shahsavari Googhari Z, Hafezi F, Asgari P, Heidari A. The effectiveness of mindfulness-based cognitive therapy and acceptance and commitment therapy on medical science students' subjective well-being, psychological distress, and emotion regulation. J Shahrekord Univ Med Sci 2022 Jan 26;24(1):35-41. [doi: 10.34172/jsums.2022.07]
63. Yogeswaran V, El Morr C. Effectiveness of online mindfulness interventions on medical students' mental health: a systematic review. BMC Public Health 2021 Dec 18;21(1):2293 [FREE Full text] [doi: 10.1186/s12889-021-12341-z] [Medline: 34920715]
64. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: a randomized controlled clinical trial. Behav Res Ther 2016 Mar;78:30-42. [doi: 10.1016/j.brat.2016.01.001] [Medline: 26848517]

65. Fischer R, Bortolini T, Karl JA, Zilberberg M, Robinson K, Rabelo A, et al. Rapid review and meta-meta-analysis of self-guided interventions to address anxiety, depression, and stress during COVID-19 social distancing. Front Psychol 2020 Oct 28;11:563876 [FREE Full text] [doi: 10.3389/fpsyg.2020.563876] [Medline: 33192837]
66. Obregon M, Luo J, Shelton J, Blevins T, MacDowell M. Assessment of burnout in medical students using the Maslach Burnout Inventory-Student Survey: a cross-sectional data analysis. BMC Med Educ 2020 Oct 21;20(1):376 [FREE Full text] [doi: 10.1186/s12909-020-02274-3] [Medline: 33087080]
67. Levin ME, Haeger J, Pierce B, Cruz RA. Evaluating an adjunctive mobile app to enhance psychological flexibility in acceptance and commitment therapy. Behav Modif 2017 Nov;41(6):846-867. [doi: 10.1177/0145445517719661] [Medline: 28689449]
68. Ditton E, Knott B, Hodyl N, Horton G, Walker FR, Nilsson M. Medical student experiences of engaging in a psychological flexibility skill training app for burnout and well-being: pilot feasibility study. JMIR Form Res 2023 Jan 10;7:e43263 [FREE Full text] [doi: 10.2196/43263] [Medline: 36626191]
69. Hayes SC, Merwin RM, McHugh L, Sandoz EK, A-Tjak JG, Ruiz FJ, et al. Report of the ACBS Task Force on the strategies and tactics of contextual behavioral science research. J Contextual Behav Sci 2021 Apr;20:172-183. [doi: 10.1016/j.jcbs.2021.03.007]
70. Hayes SC, Hofmann SG, Stanton CE, Carpenter JK, Sanford BT, Curtiss JE, et al. The role of the individual in the coming era of process-based therapy. Behav Res Ther 2019 Jun;117:40-53. [doi: 10.1016/j.brat.2018.10.005] [Medline: 30348451]
71. Hofmann SG, Hayes SC. The future of intervention science: process-based therapy. Clin Psychol Sci 2019 Jan;7(1):37-50 [FREE Full text] [doi: 10.1177/2167702618772296] [Medline: 30713811]
72. Rogge RD, Daks JS, Dubler BA, Saint KJ. It's all about the process: examining the convergent validity, conceptual coverage, unique predictive validity, and clinical utility of ACT process measures. J Contextual Behav Sci 2019 Oct;14:90-102. [doi: 10.1016/j.jcbs.2019.10.001]
73. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
74. Villatte JL, Vilardaga R, Villatte M, Plumb Vilardaga JC, Atkins DC, Hayes SC. Acceptance and Commitment Therapy modules: differential impact on treatment processes and outcomes. Behav Res Ther 2016 Feb;77:52-61 [FREE Full text] [doi: 10.1016/j.brat.2015.12.001] [Medline: 26716932]
75. Levin ME, Haeger J, Cruz RA. Tailoring acceptance and commitment therapy skill coaching in the moment through smartphones: results from a randomized controlled trial. Mindfulness (N Y) 2019;10(4):689-699. [doi: 10.1007/s12671-018-1004-2]
76. Gates KM, Hellberg SN. Commentary: person-specific, multivariate, and dynamic analytic approaches to actualize ACBS task force recommendations for contextual behavioral science. J Contextual Behav Sci 2022 Jul;25:35-43. [doi: 10.1016/j.jcbs.2022.06.002]
77. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
78. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
79. Ditton E, Knott B, Hodyl N, Horton G, Walker FR, Nilsson M. Assessing the efficacy of an individualized psychological flexibility skills training intervention app for medical student burnout and well-being: protocol for a randomized controlled trial. JMIR Res Protoc 2022 Feb 04;11(2):e32992 [FREE Full text] [doi: 10.2196/32992] [Medline: 35119378]
80. Maslach C, Jackson SE, Leiter MP. Maslach burnout inventory: third edition. In: Zalaquett CP, Wood RJ, editors. Evaluating Stress: A Book of Resources. Lanham, MD, USA: Scarecrow Education; 1997:191-218.
81. Schaufeli WB, Martínez IM, Pinto AA, Salanova M, Bakker AB. Burnout and engagement in university students: a cross-national study. J Cross Cult Psychol 2016 Jul 27;33(5):464-481. [doi: 10.1177/0022022102033005003]
82. Williamson K, Lank PM, Cheema N, Hartman N, Lovell EO, Emergency Medicine Education Research Alliance (EMERA). Comparing the Maslach burnout inventory to other well-being instruments in emergency medicine residents. J Grad Med Educ 2018 Oct;10(5):532-536 [FREE Full text] [doi: 10.4300/JGME-D-18-00155.1] [Medline: 30386478]
83. Galán F, Sanmartín A, Polo J, Giner L. Burnout risk in medical students in Spain using the Maslach Burnout Inventory-Student Survey. Int Arch Occup Environ Health 2011 Apr;84(4):453-459. [doi: 10.1007/s00420-011-0623-x] [Medline: 21373879]
84. Keyes CL. Mental illness and/or mental health? Investigating axioms of the complete state model of health. J Consult Clin Psychol 2005 Jun;73(3):539-548 [FREE Full text] [doi: 10.1037/0022-006X.73.3.539] [Medline: 15982151]
85. Lamers SM, Westerhof GJ, Bohlmeijer ET, ten Klooster PM, Keyes CL. Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF). J Clin Psychol 2011 Jan;67(1):99-110. [doi: 10.1002/jclp.20741] [Medline: 20973032]

86. Keyes CL. Brief description of the mental health continuum short form (MHC-SF). Emory College of Arts and Science. Atlanta, GA, USA: Emory College of Arts and Science; 2009. URL: https://peplab.web.unc.edu/wp-content/uploads/sites/18901/2018/11/MHC-SFoverview.pdf [accessed 2022-01-08]
87. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. Behav Modif 2017 Jan 01;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]
88. Grégoire S, Gagnon J, Lachance L, Shankland R, Dionne F, Kotsou I, et al. Validation of the english and french versions of the multidimensional psychological flexibility inventory short form (MPFI-24). J Contextual Behav Sci 2020 Oct;18:99-110. [doi: 10.1016/j.jcbs.2020.06.004]
89. Seidler D, Stone B, Clark BE, Koran J, Drake CE. Evaluating the factor structure of the Multidimensional Psychological Flexibility Inventory: an independent replication and extension. J Contextual Behav Sci 2020 Jul;17:23-31. [doi: 10.1016/j.jcbs.2020.04.007]
90. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. 2nd edition. Sydney, Australia: Psychology Foundation of Australia; 1995.
91. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]
92. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
93. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J Med Internet Res 2021 Mar 23;23(3):e24850 [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]
94. McCoy CE. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017 Oct;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
95. Dyrbye LN, Harper W, Moutier C, Durning SJ, Power DV, Massie FS, et al. A multi-institutional study exploring the impact of positive mental health on medical students' professionalism in an era of high burnout. Acad Med 2012 Aug;87(8):1024-1031. [doi: 10.1097/ACM.0b013e31825cfa35] [Medline: 22722352]
96. Thompson EM, Destree L, Albertella L, Fontenelle LF. Internet-based acceptance and commitment therapy: a transdiagnostic systematic review and meta-analysis for mental health outcomes. Behav Ther 2021 Mar;52(2):492-507. [doi: 10.1016/j.beth.2020.07.002] [Medline: 33622516]
97. Prudenzi A, Graham CD, Clancy F, Hill D, O'Driscoll R, Day F, et al. Group-based acceptance and commitment therapy interventions for improving general distress and work-related distress in healthcare professionals: a systematic review and meta-analysis. J Affect Disord 2021 Dec 01;295:192-202 [FREE Full text] [doi: 10.1016/j.jad.2021.07.084] [Medline: 34479127]
98. Dyrbye LN, Thomas MR, Power DV, Durning S, Moutier C, Massie Jr FS, et al. Burnout and serious thoughts of dropping out of medical school: a multi-institutional study. Acad Med 2010 Jan;85(1):94-102. [doi: 10.1097/ACM.0b013e3181c46aad] [Medline: 20042833]
99. Dindo L. One-day acceptance and commitment training workshops in medical populations. Curr Opin Psychol 2015 Apr 01;2:38-42 [FREE Full text] [doi: 10.1016/j.copsyc.2015.01.018] [Medline: 25793217]
100. Kroska EB, Roche AI, O'Hara MW. How much is enough in brief Acceptance and Commitment Therapy? A randomized trial. J Contextual Behav Sci 2020 Jan;15:235-244. [doi: 10.1016/j.jcbs.2020.01.009]
101. Towey-Swift KD, Lauvrud C, Whittington R. Acceptance and commitment therapy (ACT) for professional staff burnout: a systematic review and narrative synthesis of controlled trials. J Ment Health (forthcoming) 2022 Feb 02:1-13. [doi: 10.1080/09638237.2021.2022628] [Medline: 35105253]
102. Kalani SD, Azadfallah P, Oreyzi H, Adibi P. Interventions for physician burnout: a systematic review of systematic reviews. Int J Prev Med 2018 Sep 17;9:81 [FREE Full text] [doi: 10.4103/ijpvm.IJPVM_255_18] [Medline: 30283613]
103. Biglan A, Hayes SC, Pistorello J. Acceptance and commitment: implications for prevention science. Prev Sci 2008 Sep;9(3):139-152 [FREE Full text] [doi: 10.1007/s11121-008-0099-4] [Medline: 18690535]
104. Pakenham KI, Landi G, Boccolini G, Furlani A, Grandi S, Tossani E. The moderating roles of psychological flexibility and inflexibility on the mental health impacts of COVID-19 pandemic and lockdown in Italy. J Contextual Behav Sci 2020 Jul;17:109-118 [FREE Full text] [doi: 10.1016/j.jcbs.2020.07.003] [Medline: 32834969]
105. Hayes SC, Strosahl K, Wilson KG, Bissett RT, Pistorello J, Toarmino D, et al. Measuring experiential avoidance: a preliminary test of a working model. Psychol Rec 2004;54(4):553-578. [doi: 10.1007/bf03395492]
106. Cunningham JA, Kypri K, McCambridge J. Exploratory randomized controlled trial evaluating the impact of a waiting list control design. BMC Med Res Methodol 2013 Dec 06;13:150 [FREE Full text] [doi: 10.1186/1471-2288-13-150] [Medline: 24314204]
107. Kinser PA, Robins JL. Control group design: enhancing rigor in research of mind-body therapies for depression. Evid Based Complement Alternat Med 2013;2013:140467 [FREE Full text] [doi: 10.1155/2013/140467] [Medline: 23662111]

108. Carrau D, Janis JE. Physician burnout: solutions for individuals and organizations. Plast Reconstr Surg Glob Open 2021 Feb;9(2):e3418 [FREE Full text] [doi: 10.1097/GOX.0000000000003418] [Medline: 33680666]

Abbreviations

ACT: Acceptance and Commitment Training
EMA: ecological momentary assessment
JMP: Joint Medical Program
MBI-GS (S): Maslach Burnout Inventory General Survey for Students
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by J Torous, G Eysenbach; submitted 09.09.22; peer-reviewed by E Thomas, M Cederberg; comments to author 19.10.22; revised version received 06.12.22; accepted 24.12.22; published 06.02.23.

Please cite as:
Ditton E, Knott B, Hodyl N, Horton G, Oldmeadow C, Walker FR, Nilsson M
Evaluation of an App-Delivered Psychological Flexibility Skill Training Intervention for Medical Student Burnout and Well-being: Randomized Controlled Trial
JMIR Ment Health 2023;10:e42566
URL: https://mental.jmir.org/2023/1/e42566
doi:10.2196/42566
PMID:36745486

Original Paper

Effects of Social Media Use on Connectivity and Emotions During Pandemic-Induced School Closures: Qualitative Interview Study Among Adolescents

Elisa Liang[1], BA; Emily R Kutok[2], BA, BSc; Rochelle K Rosen[1,3,4,5], PhD; Taylor A Burke[6,7], PhD; Megan L Ranney[1,4,8,9], MD, MPH

[1]Brown-Lifespan Center for Digital Health, Providence, RI, United States
[2]Lynch School of Education and Human Development, Boston College, Chestnut Hill, MA, United States
[3]Department of Psychiatry and Human Behavior, Alpert Medical School of Brown University, Providence, RI, United States
[4]Department of Behavioral and Social Science, Brown University School of Public Health, Providence, RI, United States
[5]Center for Behavioral and Preventive Medicine, The Miriam Hospital, Providence, RI, United States
[6]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States
[7]Department of Psychiatry, Harvard Medical School, Boston, MA, United States
[8]Department of Emergency Medicine, Alpert Medical School of Brown University, Providence, RI, United States
[9]Department of Emergency Medicine, Rhode Island Hospital, Providence, RI, United States

Corresponding Author:
Megan L Ranney, MD, MPH
Brown-Lifespan Center for Digital Health
139 Point St.
Providence, RI, 02903
United States
Phone: 1 401 444 2557
Email: megan_ranney@brown.edu

Abstract

Background: The COVID-19 pandemic provided a unique opportunity to examine social media and technology use during a time in which technology served as adolescents' primary form of socialization. The literature is mixed regarding how increased screen time during this period affected adolescent mental health and well-being. The mechanisms by which screen time use affected adolescent psychosocial outcomes are also unknown.

Objective: We aimed to deepen our understanding of how social media and technology use, social connectivity, and emotional well-being intersected during pandemic-related school closures.

Methods: English-speaking adolescents aged 13 to 17 years were recruited on Instagram for a brief screening survey; 39 participants were purposefully selected to complete a semistructured interview regarding their social media and technology use during the pandemic. Interview summaries were abstracted from recordings, and deductive codes were created for the primary question stems. These codes were subsequently reviewed for the main themes.

Results: The main themes were as follows: adolescent social media and technology use during school closures usually allowed for more and easier social connectivity, but the amount and relative ease of connectivity differed according to purpose and type of use. Emotions, particularly those of stress and happiness, were connected to whether adolescents actively or passively engaged with social media and technology.

Conclusions: Our results suggest a nuanced relationship among social media and technology use, adolescent social support, and emotional well-being, including during the pandemic. Specifically, how adolescents use or engage with web-based platforms greatly influences their ability to connect with others and their feelings of stress and happiness. In the context of the COVID-19 pandemic and as technology in general remains at the core of the adolescent experience, future research should continue to examine how adolescents navigate and use web-based spaces in beneficial and harmful ways. This will inform education and interventions that foster healthy social media and technological habits.

(JMIR Ment Health 2023;10:e37711) doi:10.2196/37711

KEYWORDS

social media; adolescents; COVID-19; emotions; connectivity

Introduction

Background

Social media has become a ubiquitous part of adolescent lives. In a recent survey, 95% of adolescents in the United States reported the use of at least 1 social media platform and 97% of adolescents reported being on the web daily [1]. The percentage of adolescents who say they use the internet almost constantly has doubled in the past few years from 24% in 2014 to 2015 to 46% in 2022. This increasingly prevalent internet-based environment not only provides convenient ways for adolescents to connect and network with peers but also influences adolescents' mental and emotional health. A national survey assessing youth risk behavior indicated a 40% increase in adolescents who reported persistent feelings of sadness or hopelessness from 2009 to 2019 [2]. As the rates of social media use and adolescent mental illness have escalated over the past decade, so have empirical investigations on the positive and negative effects of social media use on adolescent mental health and well-being. Social media may improve adolescent well-being by increased social support and feelings of connectedness [3,4]. It may also worsen adolescent well-being via increased depressive and anxiety symptoms [3,5]. Existing quantitative and qualitative research on adolescent social media use and well-being is limited in that it targets specific platforms, forms of social media use, or outcome variables rather than exploring a more holistic narrative of adolescents' use of social media and technology. For example, numerous studies have examined the effects of passive Facebook use on well-being [6,7] and others have investigated how Facebook and Instagram were used to cope specifically with feelings of loneliness [8,9]; however, research exploring how complex web-based platforms may simultaneously provide positive and negative experiences is sparse.

A more nuanced examination of the sometimes-conflicting effects of social media on adolescent mental and emotional health is particularly important in light of the recent and ongoing COVID-19 pandemic. Increases in social media and technology use, increases in depressive and anxiety symptoms, and decreases in in-person support and interactions have been observed during pandemic-related school closures [10,11]. However, as in prepandemic studies, the relationship between social media use and mental well-being may be complicated. For example, 1 recent study found that higher total screen time use by adolescents during the pandemic was associated with poorer mental health and greater perceived stress, whereas lower screen time was associated with more social support and coping behaviors [12]. Other studies, in contrast, have found that young adults' use of social media to express emotions was related to positive mental well-being during the pandemic and that organizing digital gatherings and reporting stronger web-based support were related to greater well-being and feelings of positive affect [13,14]. Our own work suggests that the prepandemic mental health effects of social media may be moderated by adolescents' purpose for social media use such that adolescents who use social media for coping purposes may experience a protective effect from screen use compared with the negative correlation between increased screen use and well-being found in adolescents who infrequently use social media for coping purposes [10]. Other factors that may influence the relationship between adolescent social media use and psychosocial outcomes during the pandemic are social support and upward social comparison [15].

In others' work, different media-related coping strategies were associated with different indicators of well-being in young adults during the pandemic. Young adults who used social media to joke about or reframe pandemic-related situations reported more positive affect and higher mental health scores than those who used social media for escapist and avoidant coping (defined as using social media to seek distraction from frustration, stress, or anxiety owing to the pandemic) [16]. In another study, humorous coping via social media (ie, watching funny videos alone or with others) was found to be associated with greater happiness among adolescents during the lockdown [17]. By definition, humorous coping may also qualify as avoidant coping; if so, these results contrast with the previous study and suggest that some forms of avoidant media coping may have beneficial effects. In support of this, some conceptualizations of escapist media used as a form of emotion-focused avoidance coping suggest that it may have positive short-term effects on well-being by temporarily reducing stress and anxiety [18]. More research is needed to better define the purposes for which adolescents use social media during the pandemic and its related effects on emotions and well-being to further distinguish and define potential adaptive and maladaptive social media coping strategies.

Qualitative research on social media use during school closures—and specifically the types of social media use that affect mental health—could help address this gap in knowledge and generate new hypotheses about both impact and potential interventions to enhance adolescents' mental health in the context of the COVID-19 pandemic. A qualitative approach allows the generation of novel themes that can provide a framework for future inquiries [19]. Qualitative data may also provide insight into the nuances of how and why specific types of social media and technology use affect adolescent emotional well-being. Finally, qualitative evaluation of individual adolescents' perspectives on the relationship between social media use and their emotions may drive future investigations of specific aspects of social media related to negative and positive affect.

Objectives

This analysis was a component of a larger project examining adolescents' mental health in the United States and changing technology use during the COVID-19 pandemic [10]. The goal of this analysis was to explore how adolescents' interactions with social media and technology changed during school closures and how different types of social media use and purposes for social media use (eg, connecting with friends,

scrolling through posts, etc) influenced perceived well-being. In this study, "school closures" refers to the time when students were not able to attend school in-person. We incorporated quantitative self-report measures on adolescents' psychosocial outcomes (ie, depressive and anxiety symptoms and perceived social support) to contextualize our qualitative findings.

Methods

Recruitment

From September 23 to December 16, 2020, we used targeted Instagram advertisements to recruit adolescents (13-17 years of age, English speaking, and residing in the United States) to complete a brief screening survey administered using REDCap (Research Electronic Data Capture; Vanderbilt University) [20]; this screening survey confirmed basic demographics and assessed social media use and general well-being over the past 7 days [10]. All participants who completed the screening survey were asked to complete a web-based assent form and 6 comprehension questions that reviewed study protocols. The selection process for a subsequent qualitative interview began with the subsample of all adolescents who completed the assent process and was then refined in an effort to maintain purposeful representation of demographic subgroups (ie, representation across regions of the United States, race, gender, and age). Selected adolescents were phoned to confirm contact information, and parental consent was obtained. All participants interviewed completed a self-report survey using REDCap as well as a semistructured interview.

Ethics Approval

This study was approved by the Rhode Island Hospital Institutional Review Board (#883341).

Measures

Semistructured Interviews

A trained research assistant with a background in psychology conducted semistructured interviews with all consenting participants via a video call–based platform (Zoom). Participants were asked open-ended questions regarding general technology and social media use, how COVID-19 changed their technology or social media use, and the role of technology in their emotions (Multimedia Appendix 1). The interviews lasted between 15 and 51 minutes (average length 31, SD 8.85 minutes) and were digitally recorded. Qualitative debriefs (a brief review of the interview process, quality, and key data) were written after each interview to ensure quality and consistency. Interviews were continued until saturation of content was reached based on a review of the interviewer's qualitative debriefs and discussion at team meetings.

Demographics

To measure age, race, ethnicity, and socioeconomic status, select self-report items from the National Study for Adolescent Health were included in the screening survey [21]. Gender was measured using 1 item from the Gender Identity in US Surveillance Group [22]. Sexual orientation was measured using 1 item from the Network for LGBT Health Equity at the Fenway Institute [23].

Mental Well-being and Social Support

The Patient-Reported Outcomes Measurement Information System (PROMIS)–Depression (PROMIS-depression) is an 8-item questionnaire that assesses depression symptom severity in the past 7 days [24]. The PROMIS-anxiety is an 8-item questionnaire that assesses anxiety symptom severity over the past 7 days [24]. Response options for both PROMIS scales were given on a 5-point Likert scale ranging from "1=at no time" to "5=all of the time." The Multidimensional Scale of Perceived Social Support (MSPSS) was used to assess overall perceived social support from friends, family, and significant others [25-27]. The response options were on a 7-point Likert scale ranging from "1=very strongly disagree" to "4=half and half" to "7=very strongly agree." As per standard practice, the PROMIS scale and MSPSS were scored as sums. We interpreted our summed scores according to the score translation table developed for these short forms [24], with raw scores of 32 to 40 indicating severe depressive and anxiety symptoms. For the MSPSS, we adhered to traditional cutoffs, with scores of 12 to 35 indicating low, 36 to 60 indicating medium, and 61 to 84 indicating high perceived social support.

Analysis

Semistructured interview summaries were abstracted directly from the recordings by a qualitatively trained PhD student who listened to the audio recordings and noted the participant answers. Quotes were transcribed verbatim directly into the summaries. All summaries were read by a team of 3 research assistants and the interviewer to verify accuracy and completeness; discrepancies were discussed and resolved. The agreed-upon summaries were then entered into the NVivo (QSR International) qualitative data analysis software [28]. Deductive codes were created based on question stems (Multimedia Appendix 2) and applied by the interviewer. Any discrepancies were discussed and resolved by a research assistant, the interviewer, and a coinvestigator. Codes in which adolescents discussed their changes in social media use, connectivity, and emotions were abstracted and reviewed in further detail by the senior coinvestigator and the study team for main themes; memos were written for each theme and shared with the research team, and inductive codes were applied to the data based on the discussion of the initial coding scheme. The themes were developed based on discussions among the entire team. Codes were also examined comparatively in NVivo according to participant groupings based on descriptive cutoffs for the quantitative self-report measures (described earlier) to identify any patterns among those who reported severe versus nonsevere depressive and anxiety symptoms and high versus medium perceived social support.

Descriptive statistics for the entire sample were calculated using SPSS (IBM Corp).

Results

Overview

Semistructured interviews were conducted with 39 adolescents aged 13 to 17 (mean 15.9) years; approximately half of the participants identified as women (49%), and most were

non-Hispanic (85%; see Table 1 for the demographics of our study sample). Table 1 shows the demographics and quantitative scores of our study sample. To protect our participants, we did not disclose specific demographic information for the quotes used in this paper; in summary, we used 31 quotes from 19 different participants who reflect the diversity of our overall sample.

Qualitative themes were examined comparatively between those who reported high (31/39, 79%) versus medium (8/39, 21%) social support on the MSPSS (no participants reported low social support). Data were also compared between those who reported severe (6/39, 15%) versus nonsevere (33/39, 85%) depressive symptoms and severe (8/39, 21%) versus nonsevere (31/39, 79%) anxiety symptoms. There were no observed differences in the themes based on social support or well-being.

We identified two overall themes related to technology and social media use during school closures: (1) perceived social connectivity and (2) emotions. Each is described in detail in subsequent sections. For theme 1, we found that connectivity had several subthemes, each related to the amount and relative ease of connectivity that adolescents experienced with their peers on the web. For theme 2, we found that emotions, particularly that of stress and happiness, were connected to whether adolescents actively or passively engaged with social media and technology.

Table 1. Demographics of study sample (N=39).

Demographic variables	Values, n (%)
Self-reported gender	
Man	10 (26)
Woman	19 (49)
Nonbinary or transgender	8 (21)
Prefer not to answer[a]	2 (5)
Sexual orientation	
Straight	11 (28)
Gay or lesbian	6 (15)
Bisexual	10 (26)
Other	12 (31)
Ethnicity	
Hispanic	6 (15)
Non-Hispanic	33 (85)
Race	
Asian	6 (15)
Black or African American	3 (8)
White	26 (67)
Mixed	4 (10)

[a]One participant chose not to disclose their gender.

Theme 1: The Effect of Technology and Social Media Use on Perceived Social Connectivity During School Closures

Overview

Overall, participants indicated that their use of technology and social media increased during school closures owing to the pandemic. The reasons provided for this increase included taking web-based classes, maintaining social relationships, and keeping oneself entertained during school closures. Staying in touch with friends and peers emerged as a primary reason for increased social media and technology use in all 39 interviews. Participants described how social media both facilitated and hindered their ability to connect with others during this period of school closure.

More Social Connectivity

Most participants reported that social media and technology helped them stay connected with others during school closures in the COVID-19 pandemic. Most participants said they stayed in touch with their core group of friends using social media and that their number of interactions increased. Many participants said that they also used social media to create new friendships during this period. There was considerable overlap between these 2 groups; 10 participants reported that they both spoke to the same friends and made new friends on the web. A few participants said that because of technology, they became distant from formerly close friends while becoming closer to people

they had not known well before. Many participants reported meeting new web-based friends through existing in-person friends:

> *I used to just hang out with the people who sat with me at lunch, but now it's friends of friends of friends* [on Discord]

Discord is an audio-, text-, and video-based communication application whose use originated with gamers [29]; it is particularly known for its voice chat and screensharing features that allow groups of users to speak and view content together in a real-time setting.

Easier Social Connectivity

Social media and technology were described as making communication easier according to approximately a quarter of the participants. Some highlighted that it was less intimidating to approach people on the web than in person, for example:

> *It's not as scary as interacting in person...before, some people may seem too intimidating to approach, but now it's like you don't really know how anyone looks like or sounds like so it's not really much of a fear.*

Others mentioned how social media made it easier to connect with people from around the world:

> *[Social media] allows me personally to talk to friends I may never have met previously who live all over the globe, which is quite cool.*

A few adolescents pointed out that when it comes to communicating with others via technology, emotions can be more accurately represented through GIFs, videos, and emojis than through words. Others talked about how being on the web allowed them to act more like their true selves without fear of judgment from others:

> *I'm really bad at talking to people in real life but online it's so much easier and I'm really extroverted online but really introverted in person...so it just makes it a lot easier to make friends and talk to people and share stuff.*

Although these advantages of technology may have become particularly salient during the pandemic, the participants also spoke about them generally, implying that they were true before the pandemic as well.

Notably, Discord was mentioned by most participants as a platform that facilitated web-based connectivity during school closures in the COVID-19 pandemic; 8 participants said that they first started using Discord during the pandemic. Of the participants who used Discord before school closures, most reported more frequent use during the pandemic compared with before school closures. Ten participants explicitly stated that they used Discord to game with friends; the others used it to voice chat with friends, video call friends, or message friends.

Hindered Social Connectivity

More than half of our participants talked about the difficulties in interpreting web-based communications during school closures (which many referred to as "quarantine"). For example, adolescents mentioned that being unable to see their peers made it harder to read the other person's feelings and intentions through text messages. They also mentioned that they missed the cues from physical touch:

> *Just that like when it is texting...I can't see what their facial expressions are...it gets kinda confusing if they're being sarcastic about something or not, or if they are joking and—I don't know, it's just nice to be near people...you know touch them on the shoulder or something and you can't do that now.*

Another participant spoke about how specifically the loss of familiarity with their peers' personalities made it more difficult to interpret tone and intentions over text messages:

> *Before [quarantine] I would see people during class...so I would know what their personality is like, but I don't know how they've changed during quarantine, so I don't know how they're thinking or if what they're saying lines up with my idea of them that I would see in-person.*

Many participants spoke about how communicating over social media and technology lacked emotional intimacy and sincerity overall, especially without in-person interactions to supplement:

> *I feel like on social media and stuff people are less real, it feels a little more disingenuous. Like I totally get it, but it's harder. It feels less real.*

Less Social Connectivity

Approximately a quarter of the participants described how social media and technology fell short when it came to reaching others during the pandemic. Five talked about difficulties getting in touch with others on the web owing to peers failing to or taking a while to respond:

> *There can be an element of stress to it, waiting for a text or "why isn't this friend texting me back."*

Compared with in-person conversations, web-based conversations can also be suddenly dropped by 1 party without explanation:

> *It can be a little unsatisfying if you actually want to have a conversation with a person and then you send a couple of texts back and forth and they stop responding and you're like "I don't know if it's that they don't want to talk to me or if they just had to take their dog for a walk or something" I don't know.*

Another handful of participants elaborated on situations where outdated technology or poor internet connections created challenges:

> *I started using Twitter...two months ago but that's mostly to talk with my friends...because some of their internet is really bad and they can't use Discord because it requires good internet.*

Furthermore, 1 participant described how they became more selective with who they chose to stay in touch with after-school closures:

> *So I stopped using it [Snapchat] also because we're not seeing people in person, I don't know if this makes sense, I don't feel like I'm obligated...some people I follow on Snap that I don't know as well, now that I don't see them it's like I don't really need to talk to them.*

Similarly, another participant spoke about becoming generally disconnected from peers during this time because of the lack of in-person contact:

> *I'm barely talking to anyone. After we got out for school, I kind of just accidentally cut everyone off and no one heard from me...because I didn't really talk to anyone outside of school. It was mostly just an in-person thing. So now that we weren't going to school it's like oh, sorry I didn't talk to you.*

Theme 2: The Effect of Technology and Social Media Use on Emotions During School Closures

Overview

Participants were asked specifically about whether social media and technology use made them feel stressed or happy and how these emotions may have changed after school closures. Most participants denied experiencing increased stress from screen use during school closures. The remaining participants stated that they felt more stressed after school closures owing to constantly being on social media, especially owing to engaging with pandemic-related content. Although only a little more than half of the participants directly addressed whether their screen-use–related happiness changed after school closures, most of those participants claimed that their happiness stayed the same, whereas a handful reported feeling happier and only 3 reported feeling less happy.

Participants were also asked separately about instances when they felt happy and stressed owing to social media and technology use. From their responses, inductive themes surfaced regarding active and passive social media and technology engagement. As these emerged, we defined "active engagement" as per others' work [30] as creating posts on platforms, directly messaging others, or engaging with others in web-based public spaces, such as comments sections, videogame servers, etc; for example, "I post pictures of cats because I like to make people feel happy." We defined "passive engagement" as browsing through posts, watching videos, and engaging in any web-based content alone, exemplified by "I see people who are working out, getting really fit and that really motivates me..." Each of the following subthemes about stress and happiness owing to screen use are organized into subsections for active and passive engagement.

Sources of Stress—Active Versus Passive Engagement

Active Engagement

Negative interactions with others on technology and social media were a common source of stress and being upset for participants during school closures, with over one-third of participants describing such cases. This included stress from receiving upsetting messages or comments while gaming, friend group drama, unpleasant conversations in group chats, and difficulty communicating with others over SMS text messaging. Although some of these experiences predated the pandemic, participants indicated that their frequency or tenor had increased or changed during school closures. Commonly described sources of conflict included politics; the pandemic; and lesbian, gay, bisexual, transgender, queer topics. Four participants explicitly underlined the stress from expressing differences on social media. One of them said that active engagement of social media for political purposes increased stress during the pandemic:

> *Just because I have been posting more political stuff so then people refute my political stuff and then I rebut that and then it's a thing.*

Another participant described an unpleasant time when they tried to clarify misinformation but were shut down:

> *During the BLM protests, there was a thing where people were...spreading a little misinformation...and I know that that isn't a thing, that's not a law, so I disagreed with them, but I ended up being downvoted a ton, so for disagreeing with facts, I was kind of censored.*

A few participants recalled receiving mean or upsetting messages from other social media users. Specifically, 1 participant mentioned that they were called "*ugly*" in typed comments from strangers on the web, and another said that they received mean text messages from friends of friends that they did not know well. Three participants explicitly spoke about being cyberbullied during school closures by people who had disliked their posts and moving or removing public social media accounts as a result. One of these adolescents felt that they were being cyberbullied more during the pandemic than before school closures; this participant said that cyberbullying in general has increased during the pandemic because everyone is on the web and "people are finding reasons to get upset over something very minor."

Passive Engagement

Negative passive engagement, which involves viewing but not participating in negative content or interactions on the web, was reported as a source of stress and upset for a large majority of participants. Participants described, for example, how passively observing negative posts or web-based arguments about politics; pandemic restrictions; and lesbian, gay, bisexual, transgender, queer topics generated more frequent stress than direct involvement in arguments or conflict:

> *I've been seeing a lot of not great stuff. I'm trans, so I've been seeing stuff about hate of trans people. That makes me uncomfortable. And Black Lives Matter stuff is stressful.*

One participant observed that especially during school closures, passive engagement with social media increased exposure to conflict:

> *In real life, you might get bad news that is more connected to you, with social media, depending on what you're looking at or connected to, you'll get bad news of everything in the world. And that can lead to heavy fatigue if you're not careful about it.*

Other sources of passive negative engagement described by participants to generate stress included witnessing incidents of cyberbullying, often involving racist or homophobic statements and political disagreements. Many participants observed that cyberbullying occurred more often during school closures owing to increased time spent on the web and lack of in-person contact with the survivors:

> *They don't have to be face to face with that person, so they send whatever they want. People aren't scared of what they say behind the screen.*

One participant (bystander) provided a specific example of this occurring to 1 survivor:

> *It [the bullying] definitely got more like aggressive now that they weren't seeing her [the victim] in-person every day.*

Web-based conflict and cyberbullying were also described to become more severe during the pandemic owing to the politically charged atmosphere and clashing of passionate and controversial opinions:

> *Before [the pandemic] it was just kinda making fun of people for maybe how they looked or if they did something stupid, but now it's about much more serious topics, presidential stuff, racism, xenophobia, homophobia, stuff like that. It's much more serious and people...feel much more passionately about it than they do about calling somebody stupid or saying that their dress is ugly.*

In particular, 1 participant felt upset when they saw cyberbullying related to living circumstances affected by the pandemic:

> *A lot of people are just like shaming people and judging people for what they have like how their families are surviving this. Because like a lot of people are out of their jobs, and people are like oh my gosh why can't you get another job. That's...not how it works.*

However, some participants also felt that more people were speaking up against cyberbullying than before the pandemic:

> *Now more than ever, people who used to not say anything are getting up and saying what they think. After COVID, they have realized, maybe it's better to stand up and say what they need to say. I feel a lot of new people are sharing their thoughts about conflict.*

In addition, many participants spoke about feeling stress from comparing themselves to others on the web, including comparisons to the amount of engagement that others were receiving. This source of passive stress reportedly increased during school closures as well. One participant explained as follows:

> *I feel like I followed more people since the pandemic started so I feel like I should get more likes and more comments and stuff. It kind of brings down your self-esteem too, like I base my worth on how my Instagram is.*

Some participants reported being upset when they juxtaposed their pandemic lives with others, such as 1 adolescent who said:

> *Sometimes it kind of makes me feel down because I see people out doing stuff and I'm like I wish I could be doing that.*

Finally, it was also stressful for 1 participant to see peers violating COVID-19 pandemic stay-at-home mandates:

> *I'm trying to be safe and then there are some people who just don't care. It's upsetting.*

Sources of Happiness—Active Versus Passive Engagement

Active Engagement

The types of active web-based engagement associated with self-described happiness or enjoyment during school closures included communicating, playing games, and sharing art with friends. For example, in reference to technology and social media, 1 adolescent said:

> *I enjoy to be on it now [more than before], like I follow a bunch of artists so I like it when I see them do artwork and like sharing...[their posts] with my friends.*

One participant mentioned that seeing their peers' social media posts made them feel happier now compared with before the pandemic:

> *I think I might actually like looking at people's pictures better now, because I didn't look at them that much before, like at all...it just gives me a way I can talk to people and even see how they're doing.*

Overall, many participants cited social connection as their reason for experiencing happiness from social media use, particularly in the context of school closures:

> *Seeing that my friends are doing okay, seeing that my friends are doing well and fine, and getting to wish them happy birthdays and stuff, like outside of in person, it makes me happy.*

This aligns with a general sentiment of appreciation toward social media and technology from participants who realized how crucial these platforms were for communication during the pandemic:

> *I didn't really acknowledge it [happiness from social media use] as much before COVID, but when I think about it I'm like what would I be doing in COVID without this.*

Over half of the participants expressed that they felt enjoyment or happiness from active technology and social media use during school closures.

Passive Engagement

Types of passive engagement associated with self-described happiness or enjoyment primarily involved viewing posts with positive or entertaining content (cat videos, artwork, inspiring messages, etc).

A handful of participants elaborated that they specifically used technology or social media passively when they were feeling down:

> *If I'm feeling really kind of upset and I just want to distract myself, if I just start watching cat videos or something, I tend to feel a bit better.*

Notably, this participant described this feeling to be distinct from that of happiness ("I wouldn't say it makes me feel happy, but I would say it makes me feel less bad").

Especially during school closures, participants seemed to rely on social media use as a source of support or coping strategy for pandemic-related stressors. One participant spoke about social media being a motivating force for them to get out of bed and be productive each morning. Similarly, another participant explained how social media was helpful when present issues became too overwhelming:

> *Sometimes it is good because if I need a break from reality, I can just connect with a page I really like, just look at their photos and that kind of reminds me of when stuff was open or when things were normal.*

Furthermore, a few participants mentioned that they purposefully customized their social media feed to only show content that made them feel happy. This was described as a particularly useful skill during the pandemic when social media was overrun with controversial discourse:

> *I realized you can basically remove people you don't want to hear anymore, so I began unfollowing people, I began deleting apps, and I began following people who kind of brought me happiness, so now whenever I go on social media, it's never like "oh no, I don't want to go on social media, why??" but now it's just like "oh look, another good story."*

Overall, passively engaging on social media was related to positive emotions in over one-third of the participants.

Discussion

Principal Findings

This first-of-its-kind qualitative analysis explores adolescents' own perspectives on the value and pitfalls of social media and technology use during the COVID-19 pandemic. Although many findings mirrored prepandemic challenges, we identified several novel and important themes in the context of the COVID-19 pandemic. We identified the following: (1) new use of old technologies (ie, Discord), (2) new sources of stress, (3) novel benefits of active and passive social media use, (4) changes in web-based social circles and social experiences, (5) persisting limitations of social media use, and (6) changes in cyberbullying and bystander behaviors. Surprisingly, we found that these challenges and opportunities cut across our demographically and experientially varied group of participants, with no major thematic differences according to gender, age, perceived social support, and emotional well-being. Although increased time on social media during the COVID-19 pandemic was previously associated with lower mood and poorer mental health outcomes in adolescents [12], our analysis provides insights into how the purpose and type of technology use influences both positive and negative outcomes.

Although much research discussion of adolescent technology use has focused on social media platforms such as TikTok, Instagram, and Snapchat [31], Discord emerged as being among the most used platforms for our participants. Discord is a web-based application that allows messaging, voice chat or video calls, and playing cooperative video games. Previous research on Discord has highlighted its use for multiplayer video games and its potential for subversion by White nationalists [32,33]. Others' work has evaluated Discord's potential to improve distance learning owing to its attractive user interface, completeness of features, and ease of maintaining social links between students and professors [34,35]. Our work confirms this potentially positive value of Discord for adolescents, particularly during the pandemic era. For example, the ability to combine voice and text chat may better emulate the feeling of in-person conversations without the associated pressures of being in-person [29]. Our findings therefore raise new insights about the value of nonvideo gaming uses of Discord for social connection; both observational and interventional research is needed. Finally, we urge researchers to include Discord as an identified social media platform in future survey research.

We identified novel web-based stressors during the pandemic, namely observing others' noncompliance with COVID-19 pandemic stay-at-home mandates and observing or participating in polarizing web-based discourses surrounding the pandemic, politics, and other controversial events during the time participants were interviewed. Although prepandemic research on active social media use suggests that active engagement is typically associated with positive well-being outcomes [7,36], emerging research suggests that active engagement with COVID-19 social media content during the pandemic is related to higher anxiety [37]. Illustratively, in our study, actively participating in web-based discussions about current events such as the Black Lives Matter protests and distancing regulations served as sources of stress that adolescents had not experienced previously. During the pandemic, social media platforms became the primary medium for the widespread dissemination of news, information (and misinformation), and political propaganda [38], which kept adolescents informed about current events and perspectives and provided endless fuel for discussion and debate. We suggest that the purpose behind active social media use (social connection vs political discourse) may influence its effects on well-being. Future work should quantitatively explore the relative influence of health or political mis- and disinformation and provide guidance to youth on how to proactively identify and manage potential political and social conflicts on the web.

The value of active social media engagement in this analysis parallels that of prepandemic research [7,36] such that participants reported that using social media to connect with others led them to experience positive affect. However, active engagement in social media during school closures was identified as qualitatively different from that before school closures. For example, some participants mentioned that messaging and sharing content with friends on social media was more enjoyable than in prepandemic times. Overall, social

connection with friends emerged as the predominant source of happiness from active screen use during the pandemic. Our analysis also aligns with prepandemic research reporting that passive social media use is more strongly correlated with negative well-being than active use [6,7,14], particularly around the impact of social comparison [15,39]. However, participants also mentioned that passively using social media as a distraction served a functional purpose by making them feel better when they were overwhelmed by pandemic-related stressors. One adolescent even used the phrase "a break from reality" to describe the reason for their screen use. In discussing how passive screen use improved their affect, participants more frequently recalled times when passive use relieved negative feelings rather than explicitly instilling positive feelings. These reports support previous work suggesting that avoidance coping using social media may sometimes be beneficial by allowing adolescents to engage with preferred content and, in turn, temporarily mitigate stress and other negative emotions [18].

Another prominent theme identified in our analysis was how technology both enhanced and hindered adolescents' ability to socially connect with friends and peers during school closures. Web-based platforms being the sole method of communication introduced unique challenges described by participants. Prepandemic research highlights the ways in which social media allows youth to create and maintain larger social circles [40,41]. Although this was still true for many of our participants during school closures, this analysis found that social media was sometimes deliberately or unintentionally used by adolescents to limit social circles during this period. Others reported naturally losing contact with peers over time. Before the pandemic, adolescents ascribed digitally disconnecting from others to web-based drama or bullying [42]; however, digital disconnect during the pandemic was attributed more simply to the lack of in-person contact to propel or supplement peer interactions. Nonetheless, our study sample reported a relatively high perceived social support level despite being during school closures, implying that feelings of social support can be maintained through social media in the absence of in-person interaction. We were unable to tease out perceived social support among adolescents who connected primarily with friends they already knew in person versus adolescents who made new friends during this period. Future studies should examine how adolescents' perceptions and acceptance of relationships that are only based on the web have changed over the course of the pandemic and how different forms of social support may influence the effects of social media use on well-being outcomes.

Although school closures enhanced some positive aspects of technology use for connectivity, they also exacerbated preexisting negative aspects of web-based communications such as the potential to misunderstand and misinterpret text messages [43]. Despite most participants seeming to acknowledge or even prefer the convenience of messaging others (compared with talking in person), they also reported missing the ability to complement text- or image-based communication with in-person interactions and physical touch. Adolescents previously acknowledged the superficiality of web-based communication in prepandemic times, stating that such communication lacked emotional depth [43], but school closures and the complete absence of in-person contact seemed to exacerbate the impersonal feeling of digital messages reported by many participants. Evidently, there are crucial aspects of connectivity that only web-based interactions cannot fulfill—an important consideration for schools and youth-based organizations as the world continues to digitalize.

Exposure to and experiences of web-based bullying also changed during the pandemic [44,45]. Participants described an increase in the frequency of bullying owing to increased screen time as well as new forms of bullying around pandemic-related behaviors. On the positive side, participants also mentioned an increase in their peers' use of prosocial bystander behaviors. Prepandemic research on bystander intervention in cyberbullying found that increased cognitive empathy (ie, mental perspective taking) and interethnic contact were related to prosocial bystander behaviors among adolescents [46,47]. The shared struggle of pandemic-related school closures and greater exposure to news and web-based slander targeting minority groups during this time may have increased overall empathy among adolescents and encouraged more of them to speak out against perceived injustices.

There were no differences in the patterns of themes based on participants' age, gender, mental well-being, or perceived social support. The overall study sample reported high perceived social support and low depressive and anxiety symptoms, indicating that adolescents' expressed stressors and benefits of using social media and technology during the pandemic did not depend on their personal well-being. The lack of thematic differences in our varied sample suggests that the aspects of social media and technology use that adolescents found challenging and helpful during the pandemic may be universal. Active and passive screen use were also associated with stress and happiness in a similar manner across groups, implying that it may be appropriate and effective to standardize education on the pros and cons of social media and technology use for all adolescents. However, it remains true that there were a variety of narratives describing unique experiences with social media and technology use during the pandemic. This suggests that there may be other variables influencing adolescents' connectivity and feelings of stress and happiness, such as their prepandemic familiarity or comfort with web-based communication or time spent using social media and technology during the pandemic.

Limitations

Although we interviewed participants until we reached content saturation, our small study sample may have contributed to the lack of observed differences in interview themes across participants. Furthermore, the nature of a qualitative analysis prevents it from establishing any causality or generalizability between themes and adolescents' self-reported well-being. Although efforts were made to recruit a nationally representative sample of adolescents across multiple levels of identity, the subset of participants who consented to and fully completed the interview was less diverse in terms of racial and ethnic identity; however, we were able to achieve greater representation of sexual and gender identities. The study sample was also recruited solely from Instagram, so it is possible that these

participants differed from those who do not use Instagram. Although these semistructured qualitative interviews were guided by a written agenda, the nature of qualitative research is such that comments and provided details varied based on participants' individual experiences. This analysis was conducted from written summaries with verbatim quotes but not from full transcripts. Summaries, however, clearly noted when participants had no comments on particular open-ended questions and are appropriate alternatives to transcription-based analysis [48]. This analysis provides a preliminary framework of adolescents' social media use during the pandemic on which further research should continue to refine and build on.

Conclusions

Overall, this qualitative analysis on changes in adolescent social media and technology use during pandemic-related school closures demonstrates that the effects of active and passive engagement on social connectivity and emotions during this time were complex and variable, depending on the purpose for which they were used. Although evidently useful for maintaining and facilitating social connections among peers, especially during a time when in-person contact was minimal, it was also clear that there were areas where social media and technology could not replace in-person communication. Furthermore, although active web-based engagement may have been more frequently associated with happiness during school closures compared with passive engagement, it was also a source of stress for participants who participated in polarizing web-based discourse. Although passive web-based engagement was more frequently related to stress during this time, adolescents also demonstrated that they could exert control over their passive social media use to make it more beneficial and uplifting. Researchers should seek to develop and disseminate strategies to guide adolescents on how to personalize their social media experiences to be more positive rather than negative. Consideration of how patterns of use and related outcomes shift as youth return to in-person school—and how to enhance positive impacts in case of short-term, future school closures—is also needed. Ultimately, social media and technology research and education for adolescents remains essential given the prominence of web-based environments in their social lives.

Acknowledgments

The authors would like to acknowledge the Technology and Adolescent Mental Wellness (TAM) Youth Advisory Board and Jennifer Leaño, a Clinical Research Program Coordinator at the Brown-Lifespan Center for Digital Health. This study was funded by grant 0000000136/132580194 from the TAM program at the University of Wisconsin-Madison. The content is solely the responsibility of the authors and does not necessarily represent the official views of the university or the TAM program. The components of these methods were supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development grant R21HD088739. The sponsors had no role in data collection, analysis, or review of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Abbreviated semistructured interview question guide.
[DOCX File , 15 KB - mental_v10i1e37711_app1.docx]

Multimedia Appendix 2
Deductive codes corresponding to the abbreviated semistructured interview question guide.
[DOCX File , 16 KB - mental_v10i1e37711_app2.docx]

References

1. Vogels EA, Gelles-Watnick R, Massarat N. Teens, social media and technology 2022. Pew Research Center. 2022. URL: https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/ [accessed 2022-08-30]
2. Mental health. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyyouth/mental-health/index.htm [accessed 2022-09-01]
3. Vannucci A, McCauley Ohannessian C. Social media use subgroups differentially predict psychosocial well-being during early adolescence. J Youth Adolesc 2019 Aug 29;48(8):1469-1493. [doi: 10.1007/s10964-019-01060-9] [Medline: 31256313]
4. Cole DA, Nick EA, Zelkowitz RL, Roeder KM, Spinelli T. Online social support for young people: does it recapitulate in-person social support; can it help? Comput Human Behav 2017 Mar;68:456-464 [FREE Full text] [doi: 10.1016/j.chb.2016.11.058] [Medline: 28993715]
5. Seabrook E, Kern M, Rickard N. Social networking sites, depression, and anxiety: a systematic review. JMIR Ment Health 2016 Nov 23;3(4):e50 [FREE Full text] [doi: 10.2196/mental.5842] [Medline: 27881357]
6. Verduyn P, Lee D, Park J, Shablack H, Orvell A, Bayer J, et al. Passive Facebook usage undermines affective well-being: experimental and longitudinal evidence. J Exp Psychol Gen 2015 Apr;144(2):480-488 [FREE Full text] [doi: 10.1037/xge0000057] [Medline: 25706656]

7. Frison E, Eggermont S. Exploring the relationships between different types of Facebook use, perceived online social support, and adolescents' depressed mood. Social Sci Comput Rev 2015 Jan 21;34(2):153-171 [FREE Full text] [doi: 10.1177/0894439314567449]
8. Wang K, Frison E, Eggermont S, Vandenbosch L. Active public Facebook use and adolescents' feelings of loneliness: evidence for a curvilinear relationship. J Adolesc 2018 Aug;67:35-44 [FREE Full text] [doi: 10.1016/j.adolescence.2018.05.008] [Medline: 29894889]
9. Pittman M, Reich B. Social media and loneliness: why an Instagram picture may be worth more than a thousand Twitter words. Comput Human Behav 2016 Sep;62:155-167 [FREE Full text] [doi: 10.1016/j.chb.2016.03.084]
10. Burke T, Kutok E, Dunsiger S, Nugent NR, Patena JV, Riese A, et al. A national snapshot of U.S. adolescents' mental health and changing technology use during COVID-19. Gen Hosp Psychiatry 2021;71:147-148 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.05.006] [Medline: 34090695]
11. Nearchou F, Flinn C, Niland R, Subramaniam S, Hennessy E. Exploring the impact of COVID-19 on mental health outcomes in children and adolescents: a systematic review. Int J Environ Res Public Health 2020 Nov 16;17(22):8479 [FREE Full text] [doi: 10.3390/ijerph17228479] [Medline: 33207689]
12. Nagata J, Cortez C, Cattle C, Ganson KT, Iyer P, Bibbins-Domingo K, et al. Screen time use among US adolescents during the COVID-19 pandemic: findings from the Adolescent Brain Cognitive Development (ABCD) study. JAMA Pediatr 2022 Jan 01;176(1):94-96 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.4334] [Medline: 34724543]
13. Canale N, Marino C, Lenzi M, Vieno A, Griffiths MD, Gaboardi M, et al. How communication technology fosters individual and social wellbeing during the COVID-19 pandemic: preliminary support for a digital interaction model. J Happiness Stud 2022;23(2):727-745 [FREE Full text] [doi: 10.1007/s10902-021-00421-1] [Medline: 34177363]
14. Pennington N. Communication outside of the home through social media during COVID-19. Comput Hum Behav Rep 2021;4:100118 [FREE Full text] [doi: 10.1016/j.chbr.2021.100118] [Medline: 34568638]
15. Masciantonio A, Bourguignon D, Bouchat P, Balty M, Rimé B. Don't put all social network sites in one basket: Facebook, Instagram, Twitter, TikTok, and their relations with well-being during the COVID-19 pandemic. PLoS One 2021;16(3):e0248384 [FREE Full text] [doi: 10.1371/journal.pone.0248384] [Medline: 33705462]
16. Eden A, Johnson B, Reinecke L, Grady S. Media for coping during COVID-19 social distancing: stress, anxiety, and psychological well-being. Front Psychol 2020;11:577639 [FREE Full text] [doi: 10.3389/fpsyg.2020.577639] [Medline: 33391094]
17. Cauberghe V, Van Wesenbeeck I, De Jans S, Hudders L, Ponnet K. How adolescents use social media to cope with feelings of loneliness and anxiety during COVID-19 lockdown. Cyberpsychol Behav Soc Netw 2021 Apr;24(4):250-257 [FREE Full text] [doi: 10.1089/cyber.2020.0478] [Medline: 33185488]
18. Vorderer P, Klimmt C. The Oxford Handbook of Entertainment Theory. Oxford, United Kingdom: Oxford University Press; Feb 10, 2021.
19. Guest G, MacQueen K, Namey E. Applied Thematic Analysis. Thousand Oaks, California, United States: SAGE Publications; 2012.
20. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
21. Harris K. Add Health (The National Longitudinal Study of Adolescent to Adult Health). The University of North Carolina. URL: https://addhealth.cpc.unc.edu/documentation/study-design/ [accessed 2021-06-14]
22. Best practices for asking questions to identify transgender and other gender minority respondents on population-based surveys. The Williams Institute. URL: https://williamsinstitute.law.ucla.edu/wp-content/uploads/Survey-Measures-Trans-GenIUSS-Sep-2014.pdf [accessed 2021-06-14]
23. Scout N. Transgender health: data requirements/policy implications. Network for LGBT Health Equity at The Fenway Institute. 2012. URL: https://www.cdc.gov/nchs/ppt/nchs2012/SS-37_SCOUT.pdf [accessed 2021-06-14]
24. Irwin D, Stucky B, Langer M, Thissen D, Dewitt EM, Lai JS, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. Qual Life Res 2010 May;19(4):595-607 [FREE Full text] [doi: 10.1007/s11136-010-9619-3] [Medline: 20213516]
25. Canty-Mitchell J, Zimet GD. Psychometric properties of the Multidimensional Scale of Perceived Social Support in urban adolescents. Am J Community Psychol 2000 Jun;28(3):391-400 [FREE Full text] [doi: 10.1023/A:1005109522457] [Medline: 10945123]
26. Zimet G, Powell S, Farley G, Werkman S, Berkoff K. Psychometric characteristics of the multidimensional scale of perceived social support. J Pers Assess 1990;55(3-4):610-617. [doi: 10.1080/00223891.1990.9674095] [Medline: 2280326]
27. Zimet G, Dahlem N, Zimet S, Farley G. The multidimensional scale of perceived social support. J Personality Assess 1988 Mar;52(1):30-41 [FREE Full text] [doi: 10.1207/s15327752jpa5201_2]
28. NVivo (Version 12). QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2022-02-15]
29. Discord homepage. Discord. URL: https://discord.com/ [accessed 2022-03-01]

30. Verduyn P, Ybarra O, Résibois M, Jonides J, Kross E. Do social network sites enhance or undermine subjective well-being? A critical review. Social Issues Policy Rev 2017 Jan 13;11(1):274-302 [FREE Full text] [doi: 10.1111/sipr.12033]
31. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2022-02-05]
32. Lavin T. Culture Warlords My Journey Into the Dark Web of White Supremacy. Victoria Embankment, London: Octopus; Nov 12, 2020.
33. Brown J, Hennis G. Hateware and the outsourcing of responsibility. In: Digital Ethics. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2019.
34. Di Marco L. User-centered evaluation of Discord in midwifery education during the COVID-19 pandemic: analysis of the adaptation of the tool to student needs. Eur J Midwifery 2021;5:51 [FREE Full text] [doi: 10.18332/ejm/142638] [Medline: 34782891]
35. Arifianto M, Izzudin I. Students' acceptance of discord as an alternative online learning media. Int J Emerging Technol Learning (iJET) 2021:179-195 [FREE Full text] [doi: 10.3991/ijet.v16i20.22917]
36. Beyens I, Pouwels J, van Driel II, Keijsers L, Valkenburg P. The effect of social media on well-being differs from adolescent to adolescent. Sci Rep 2020 Jul 01;10(1):10763 [FREE Full text] [doi: 10.1038/s41598-020-67727-7] [Medline: 32612108]
37. Vaterlaus J, Spruance L, Patten E. Social media behaviors and experiences during the COVID-19 pandemic: associations with anxiety, depression, and stress. J Social Behav Health Sci 2021 Sep 27;15(1) [FREE Full text] [doi: 10.5590/jsbhs.2021.15.1.15]
38. Nunziato D. Misinformation mayhem: social media platforms' efforts to combat medical and political misinformation. GW Law Faculty Publications & Other Works. 2020. URL: https://scholarship.law.gwu.edu/faculty_publications/1502 [accessed 2020-01-01]
39. Nesi J, Prinstein M. Using social media for social comparison and feedback-seeking: gender and popularity moderate associations with depressive symptoms. J Abnorm Child Psychol 2015 Nov;43(8):1427-1438 [FREE Full text] [doi: 10.1007/s10802-015-0020-0] [Medline: 25899879]
40. Allen K, Ryan T, Gray D, McInerney DM, Waters L. Social media use and social connectedness in adolescents: the positives and the potential pitfalls. Aus Educ Developmental Psychol 2020 Oct 19;31(1):18-31 [FREE Full text] [doi: 10.1017/edp.2014.2]
41. Lenhart A. Chapter 4: social media and friendships. Pew Research Center. 2015 Aug 6. URL: https://www.pewresearch.org/internet/2015/08/06/chapter-4-social-media-and-friendships/ [accessed 2022-02-05]
42. Anderson M, Jiang J. Teens' social media habits and experiences. Pew Research Center. 2018 Nov 28. URL: https://www.pewresearch.org/internet/2018/11/28/teens-social-media-habits-and-experiences/ [accessed 2022-02-18]
43. Oprea C, Stan A. Adolescents' perceptions of online communication. Procedia Social Behav Sci 2012;46:4089-4091 [FREE Full text] [doi: 10.1016/j.sbspro.2012.06.204]
44. Wright M, Wachs S, Harper B. The moderation of empathy in the longitudinal association between witnessing cyberbullying, depression, and anxiety. Cyberpsychology J Psychosocial Res Cyberspace 2018 [FREE Full text] [doi: 10.5817/CP2018-4-6]
45. Doumas D, Midgett A. Witnessing cyberbullying and internalizing symptoms among middle school students. Eur J Investig Health Psychol Educ 2020 Oct 04;10(4):957-966 [FREE Full text] [doi: 10.3390/ejihpe10040068] [Medline: 34542429]
46. Barlińska J, Szuster A, Winiewski M. Cyberbullying among adolescent bystanders: role of affective versus cognitive empathy in increasing prosocial cyberbystander behavior. Front Psychol 2018;9:799 [FREE Full text] [doi: 10.3389/fpsyg.2018.00799] [Medline: 29899715]
47. Abbott N, Cameron L. What makes a young assertive bystander? The effect of intergroup contact, empathy, cultural openness, and in-group bias on assertive bystander intervention intentions. J Social Issues 2014 Mar 06;70(1):167-182 [FREE Full text] [doi: 10.1111/josi.12053]
48. Halcomb E, Davidson P. Is verbatim transcription of interview data always necessary? Appl Nurs Res 2006 Feb;19(1):38-42 [FREE Full text] [doi: 10.1016/j.apnr.2005.06.001] [Medline: 16455440]

Abbreviations

MSPSS: Multidimensional Scale of Perceived Social Support
PROMIS: Patient-Reported Outcomes Measurement Information System
REDCap: Research Electronic Data Capture

Edited by J Torous; submitted 04.03.22; peer-reviewed by B Trifiro, J Torous; comments to author 25.08.22; revised version received 01.09.22; accepted 01.09.22; published 23.02.23.

Please cite as:
Liang E, Kutok ER, Rosen RK, Burke TA, Ranney ML
Effects of Social Media Use on Connectivity and Emotions During Pandemic-Induced School Closures: Qualitative Interview Study Among Adolescents
JMIR Ment Health 2023;10:e37711
URL: https://mental.jmir.org/2023/1/e37711
doi:10.2196/37711
PMID:36054613

Original Paper

Trends in Language Use During the COVID-19 Pandemic and Relationship Between Language Use and Mental Health: Text Analysis Based on Free Responses From a Longitudinal Study

Rachel Weger[1], BA; Juan Antonio Lossio-Ventura[2], PhD; Margaret Rose-McCandlish[1], BA; Jacob S Shaw[2], BA; Stephen Sinclair[2], PhD; Francisco Pereira[2], PhD; Joyce Y Chung[2], MD; Lauren Yvette Atlas[1,2,3], PhD

[1]National Center for Complementary and Integrative Health, National Institutes of Health, Bethesda, MD, United States

[2]National Institute of Mental Health, National Institutes of Health, Bethesda, MD, United States

[3]National Institute on Drug Abuse, National Institutes of Health, Baltimore, MD, United States

Corresponding Author:
Lauren Yvette Atlas, PhD
National Center for Complementary and Integrative Health
National Institutes of Health
10 Center Drive
Rm 4-1741
Bethesda, MD, 20892
United States
Phone: 1 301 827 0214
Email: lauren.atlas@nih.gov

Abstract

Background: The COVID-19 pandemic and its associated restrictions have been a major stressor that has exacerbated mental health worldwide. Qualitative data play a unique role in documenting mental states through both language features and content. Text analysis methods can provide insights into the associations between language use and mental health and reveal relevant themes that emerge organically in open-ended responses.

Objective: The aim of this web-based longitudinal study on mental health during the early COVID-19 pandemic was to use text analysis methods to analyze free responses to the question, “Is there anything else you would like to tell us that might be important that we did not ask about?” Our goals were to determine whether individuals who responded to the item differed from nonresponders, to determine whether there were associations between language use and psychological status, and to characterize the content of responses and how responses changed over time.

Methods: A total of 3655 individuals enrolled in the study were asked to complete self-reported measures of mental health and COVID-19 pandemic–related questions every 2 weeks for 6 months. Of these 3655 participants, 2497 (68.32%) provided at least 1 free response (9741 total responses). We used various text analysis methods to measure the links between language use and mental health and to characterize response themes over the first year of the pandemic.

Results: Response likelihood was influenced by demographic factors and health status: those who were male, Asian, Black, or Hispanic were less likely to respond, and the odds of responding increased with age and education as well as with a history of physical health conditions. Although mental health treatment history did not influence the overall likelihood of responding, it was associated with more negative sentiment, negative word use, and higher use of first-person singular pronouns. Responses were dynamically influenced by psychological status such that distress and loneliness were positively associated with an individual’s likelihood to respond at a given time point and were associated with more negativity. Finally, the responses were negative in valence overall and exhibited fluctuations linked with external events. The responses covered a variety of topics, with the most common being mental health and emotion, social or physical distancing, and policy and government.

Conclusions: Our results identify trends in language use during the first year of the pandemic and suggest that both the content of responses and overall sentiments are linked to mental health.

(JMIR Ment Health 2023;10:e40899) doi:10.2196/40899

KEYWORDS

COVID-19; mental health; natural language processing; sentiment analysis; free response; qualitative; text analysis; mental illness; text; mental state; language; pandemic; age; education

Introduction

Background

Times of crisis lead to increased psychological distress and mental health symptoms in the general population [1]. The literature from previous epidemics and emerging literature about the COVID-19 pandemic [2,3] provide an understanding of the mental health impacts of the COVID-19 pandemic. Increased psychological distress and mental illness is associated with a longer duration of quarantine [4,5], increased exposure to the virus or status as a health care worker [4,6,7], fear of infection of self or others [4,8], financial stress [4,9], preexisting mental illness [1,4,7], and social isolation [10,11]. It is critical to understand how this pandemic has affected mental health, document those effects, and prepare for future ones.

Language is one option for assessing mental health. Language and, more broadly, qualitative data can provide context for quantitative data and even point to new directions of research or uncover patterns that may not be found quantitatively. Language has been shown to predict states such as personality [12] and psychological constructs [13]. Research on mental health and language use has used machine learning to examine how language features correlate with or predict mental illness [14]. Other non–content-based metrics related to language have also been associated with mental illness, such as word count [15] and post counts on social media [16].

Study Aims

In this study, we examined data from a web-based mental health survey on COVID-19 stressors during the pandemic. The survey ended with an open-text free-response prompt, "Is there anything else you would like to tell us that might be important that we did not ask about?" (see Figure 1 for the study overview). Free-response questions have been shown to add context to and validate existing quantitative measures [17]. We used a variety of text analysis methods on these free responses to investigate the characteristics of our sample population, the content of responses, how responses changed over time, and how language use reflected the participants' mental state.

To characterize the participants, we measured whether the demographics of the participants who responded to the free-response question differed from those who chose not to respond. On the basis of prior literature on responses to free-text comments in surveys, we predicted that, relative to those who did not respond, respondents would be more likely to be women, be older, have more years of education, and have a preexisting health condition [17-19]. In addition, we were interested in whether prior mental illness affected responses and language features. On the basis of prior work, we hypothesized that individuals with a history of mental health conditions would be more likely to respond and provide longer responses than those without mental health histories [15,16,20]. We also hypothesized that mental health history would be associated with more negative sentiment [15,16], greater use of negative emotional words [21-25], and more first-person singular pronouns (FPSPs) [16,20-30]. The rationale behind these selections is that increased FPSP use is associated with increased self-focus and increased negative valence and negative emotional words are associated with negatively biased thinking patterns [31,32]. These patterns of thought are associated with several mental disorders, including depression [33].

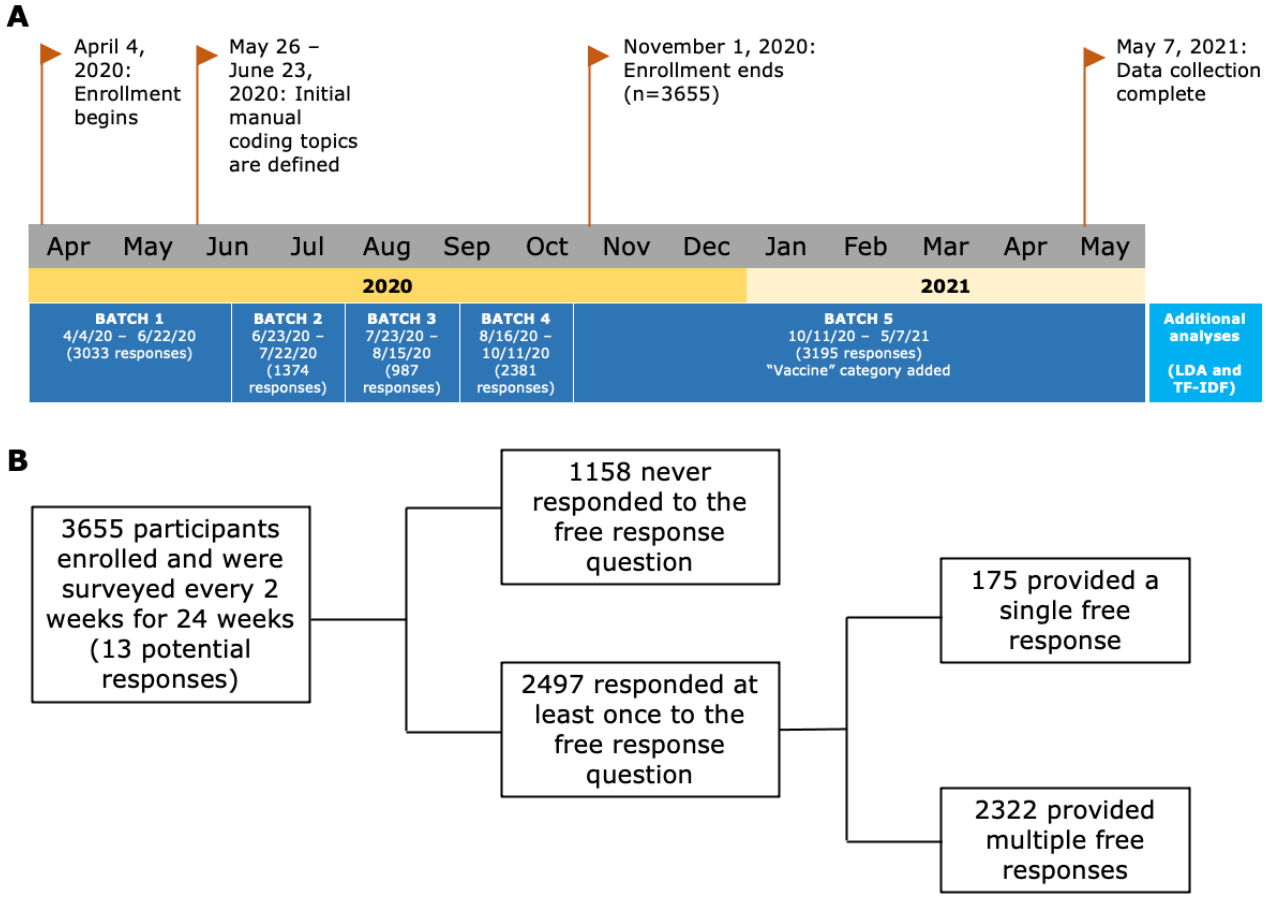
As for the content of responses, we first asked how the sentiment of responses varied over the course of the pandemic across all participants. On the basis of the literature from previous epidemics showing that distress increased with increased quarantine duration [4,5], we expected the sentiment to become more negative as social distancing and lockdown procedures remained in place. In addition, we expected emotional states to shape responses such that response likelihood and valence would be associated with fluctuations in self-reported loneliness, distress, and the presence of symptoms related to mental illness. Finally, we used various methods to categorize the responses.

Methods

Recruitment and Study Overview

A web-based, longitudinal study (NCT04339790) assessing the mental health impact of the COVID-19 pandemic was launched by investigators at the National Institute of Mental Health Intramural Research Program in early April 2020 (Figure 1). A convenience sample of adults aged ≥18 years was recruited via listserves, social media, word of mouth, flyers, and ClinicalTrials.gov (for more details, see the study by Chung et al [34]). After consenting on the web, participants completed self-report surveys upon enrollment and were then requested to respond to follow-up surveys every 2 weeks for 6 months. All survey data and responses were anonymized and associated with a unique ID.

Figure 1. Schematic of study timeline. (A) Study and analysis timeline. Enrollment in the 6-month study proceeded from April 4, 2020, through November 1, 2020, and the final data point was collected on May 7, 2021. Manual coding analysis was conducted in 5 batches during data collection, whereas additional analyses (eg, latent Dirichlet allocation [LDA]) were conducted using the entire sample after data collection was complete. The batch numbers are listed with the number and dates of the responses they contained. (B) Participant free-response rate; 68% of participants provided at least one free response during the 6-month study, with 93% of these respondents providing multiple responses. TF-IDF: term frequency–inverse document frequency.



Ethics Approval

This study was approved by the Institutional Review Board of the National Institutes of Health (NIH; 20 M-N085).

Questionnaires and Demographic Measures

At baseline, the participants completed various questionnaires assessing demographics, clinical history, and mental health symptoms (see the study by Chung et al [34] for a full list of study questionnaires). Then, they were invited to complete biweekly (ie, every 2 weeks) multiple-choice questionnaires for a 6-month period, including The Psychosocial Impact of COVID-19 Survey [35], which consisted of 45 multiple-choice questions that assess various attitudes, behaviors, and impacts surrounding the COVID-19 pandemic and a single free-response question ("Is there anything else you would like to tell us that might be important that we did not ask about?"). We analyzed responses to the free-response item and tested for associations with baseline demographics and clinical history questionnaires (see Multimedia Appendix 1 for details of classification of demographics and mental or physical health history [36-43]) as well as biweekly measures of loneliness, as measured by the University of California, Los Angeles 3-Item Loneliness Scale [44], and psychological distress, as measured by the Kessler-5 [45]. Participants could complete a maximum of 13 survey responses, one at each study time point. Of 2497 participants who provided free responses at any time point, 0.6% (n=15; range 2-6) of individuals provided duplicate responses across the study weeks. These individuals and their responses were included in the analyses because they covered stable concerns, such as employment, clinical conditions, physical health, and living situations.

Language Analyses

Sentiment Analysis and Analysis of Language Features

Sentiment analysis algorithms process text and automatically calculate the emotionality or sentiment of that text. They may simply report whether the overall text is positive or negative or use a continuous scale that quantifies both valence and intensity. To determine the optimal algorithm for the free-response data, the responses were tokenized into sentences, preprocessed, and inputted into 8 commonly used sentiment analysis applications: Stanza [46], VADER [47], LIWC2015 [48], SentiStrength [49], TextBlob [50], NLPTown model [51], Pysentimiento [52], and TweetEval [53]. We also used singular value decomposition (SVD) and a majority vote measure, which combined the outputs of 8 applications into a continuous and categorical aggregate score. Of all 10 possible options for sentiment analysis (8 different tools and aggregation of their predictions using either SVD or a majority vote), TweetEval performed the best. It obtained a precision of 0.76, recall of 0.75, F_1-score of 0.75,

and accuracy of 0.80 and was therefore selected to measure sentiment. TweetEval is a roBERTa-based model [54] trained on approximately 60 million tweets. TweetEval represents the sentiment of the text on a scale of −1 to 1, with −1 being the most negative and 1 being the most positive.

We performed a formal evaluation by assigning a polarity category to 130 sentences drawn at random, which were manually labeled by a separate observer. This allowed us to compute *F* score, precision, and recall to compare the 8 polarity scores of the algorithms; the SVD score; and a majority vote of the 8 polarity scores. The TweetEval score outperformed the other options and was therefore selected for further analyses.

The TweetEval values were aggregated by response so that each response had a score that was the mean of the sentence-level TweetEval values. Then, those scores were aggregated by date so that each date had a mean TweetEval value. The 7-day rolling averages of TweetEval and the number of responses were computed using the *zoo* package and plotted by date [55].

In addition to sentiment, we focused on 3 additional language features of interest: word count ("WC" in LIWC2015 software), percentage of negative emotional words ("negemo"), and percentage of FPSPs ("i category"), which were calculated using the output from LIWC2015 [48].

Manual Content Analysis

We used manual content analysis to evaluate the responses in addition to automated algorithms. Two clinicians and 4 other members of the research team (SS, JYC, LYA, Molly Cosgrove, RW, and MR-M) created initial manual content analysis categories. One clinician and 5 other members of the research team (SS, LYA, Molly Cosgrove, RW, MR-M, and JSS) annotated 4 small practice batches and met after each to discuss ambiguities and refine categories and definitions. In total, 36 categories and definitions were agreed upon and sorted into 6 overarching themes: mental health, physical health, social factors, career and finances, society (including government, community, or both), and other. A complete list of the categories is presented in Multimedia Appendix 1. Free responses were divided among the 4 coders (SS, MR-M, RW, and JSS), and each response was reviewed and scored by 2 randomly selected coders. Each coder labeled the responses based on their content as belonging to ≥1 manual categories. The responses were annotated in 5 batches. For the fifth batch, with a date range from October 11, 2020, to May 5, 2021, LYA annotated instead of MR-M and the category "Vaccines" was added based on a consensus of the coders after noting changes in the themes of responses. Responses such as "No," "NA," and "Nothing to report" were not categorized by any coder and were classified as nonresponses that were removed from subsequent analyses. Clinicians (SS and JYC) reviewed responses marked as clinically significant to evaluate severity.

To assess agreement between coders, interrater reliability (IRR) was calculated using the *irrCAC* package to find both the Fleiss κ and Gwet AC1 statistic [56]. The 2 methods were chosen to complement each other because the κ statistic is very commonly used for IRR, whereas the Gwet AC1 statistic overcomes some of the κ statistic's weakness with data with low variability [57,58]. For both measures, we evaluated agreement using the 1991 Altman interpretation of the κ statistic, in which <0.2 is poor, 0.2 to 0.4 is fair, 0.4 to 0.6 is moderate, 0.6 to 0.8 is good, and 0.8 to 1.0 is very good agreement [59].

Automated Topic Analysis

To supplement automated coding and manual scoring in predetermined categories, we used exploratory analyses to identify the topics that emerged in the responses over time. We focused on terms unique to each month. We used term frequency–inverse document frequency (TF-IDF), a technique that finds the words that appear the most frequently in 1 document (ie, all words for a given month) and the least frequently in the others (ie, all other months). Words were lemmatized using the *textstem* package [60], and TF-IDF was calculated using the *tidytext* package [61]. This analysis was performed independently from our manual content analysis to address topics that might have been omitted from our manual content analysis, for which categories were selected early in the pandemic and analysis proceeded in real time relative to data collection. The lemmas "coronavirus," "covid19," and "covid" were all classified as the lemma "covid," and the lemmas "vaccination," "vaccinate," and "vaccine," were all classified as "vaccine".

Finally, we used natural language processing methods, such as topic modeling and multiword expression extraction, to explore people's thoughts and concerns during the COVID-19 pandemic. Topic modeling automatically identifies clusters of words and themes from text data sets. One of the most popular methods is latent Dirichlet allocation (LDA), which seeks to classify text documents as a mixture of distinct topics [62] and has been widely used in automatic content analysis. The advantages of topic modeling are its high scalability and ability to infer topics or themes without being biased by users. However, a limitation is the potential lack of interpretability. This can occur because perplexity—a measure to evaluate the quality of topic modeling outputs and select model parameters—may be inversely correlated with human interpretability [63,64].

Therefore, the incorporation of additional evaluations or measures to validate the comprehensibility of topic-modeling outputs, such as human judgment [65,66], is necessary. In this study, we added human judgment in 3 steps to overcome the lack of interpretability. First, we used different values for the number of topics, calculated the performance for each iteration using perplexity, and compared the results to the manual number of themes found earlier by the annotators. Second, we calculated the agreement between the human and LDA topic assignments for over 100 sentences. We did this by randomly selecting 100 sentences and having 1 author sort them into the topics created by LDA. Then, the human- and LDA-selected topics were compared. Of the 100 sentences, 35 had complete agreement and 23 had weak agreement (the topic selections differed, but the author thought the LDA selection was reasonable or a closely related topic). There was no agreement among the remaining 42 sentences. Close inspection of sentences with disagreement revealed that in 14 cases, LDA-selected topics were based on keywords, but those words did not reflect the meaning of the sentence (eg, "DBT therapy has been big positive" was marked

as topic 13 [relating to test results] by LDA likely because of the word "positive," whereas the human coder rated it as topic 21 [relating to mental illness and medication]). Third, we intuitively evaluated the most representative keywords per topic (single-word terms extracted by LDA) by adding multiword terms to help represent topics better. Indeed, the output of topic modeling methods generally consists only of groupings of single-word terms. However, in natural languages, single-word terms are often part of multiword expressions and therefore do not provide complete context alone. Thus, an alternative to improve the identification of relevant topics is to incorporate multiword terms. These are expressions composed of ≥2 words with a grammatical structure and a specific meaning. Thus, we used LIDF-value [40], an information retrieval measure that extracts multiword terms. LIDF-value is based on several linguistic patterns also known as lexical categories such as nouns, adjectives, etc. Therefore, to automatically assess the content of the participants' responses, our approach consisted of four basic steps: (1) preprocessing, (2) topic modeling with LDA, (3) multiword term extraction with LIDF-value, and (4) word cloud creation. Further details regarding the automated topic analysis can be found in Multimedia Appendix 1.

Statistical Analyses

Comparing Respondents With Nonrespondents

A logistic regression was run comparing respondents and nonrespondents by gender, race, ethnicity, age, income, education, and preexisting mental health and medical conditions. Before running the logistic regression for age (the only continuous variable), the assumption of linearity between age and free-response response was tested. Participants' ages were divided into quantiles, and logits were plotted by age category. The relationship was monotonic, therefore meeting the assumptions of logistic regression. R was used for all analyses, and *ggplot2* within the package *tidyverse* was used for all figures, except where noted [67,68].

Evaluating the Likelihood of Free Response and Sentiment as a Function of Psychological State

We used a multilevel logistic regression implemented using the function *glmer* in the R package *lme4* [69] to determine whether loneliness (measured using the University of California, Los Angeles 3-Item Loneliness Scale total score) and distress (measured using the Kessler-5 overall score) influenced the likelihood of an individual providing a free response at a given time point and whether the likelihood of responding changed over time.

We also used linear mixed models restricted to participants who provided multiple free responses (n=2322) to determine whether loneliness or psychological distress influenced the mean sentiment of an individual's response at a given time point. Linear mixed models were implemented using the function *lmer* within the R package *lme4* [69]. We used a similar linear model to test whether free-response length (ie, number of words) varied over time and whether response length was related to loneliness or psychological distress. We also explored whether sentiment is associated with response length.

For each model, we included fixed effects of week (ie, time in the study relative to each participant's time of enrollment), modeled psychological state both within- and between-subjects (ie, mean-centered within individuals and grand mean–centered across individuals), and included interactions between within- and between-subjects factors to test whether individual differences moderated the effects over time. Intercepts and slopes were treated as random in linear models, whereas logistic models included only random intercepts because of issues with model convergence. Because psychological predictors were correlated, we analyzed both combined models (reported in the main manuscript) and models that separately evaluated associations with loneliness and distress (reported in Multimedia Appendix 1).

Correlations Between Patient Clinical History and Language Features

We were interested in whether those with a history of mental health treatment, termed patients, used language features differently from controls. We focused on 4 language features of interest, each aggregated to be the mean by subject across all their responses over the course of the entire survey: sentiment, word count, percentage of negative emotional words, and percentage of FPSPs.

Two methods were used to determine mental health status. We determined whether an individual had a history of mental health concerns using a clinical history questionnaire. Patients were defined as individuals who reported prior mental health treatment including hospitalization, psychotropic medication, or treatment for drug or alcohol use. We used 2-sample *t* tests (2-tailed) to assess whether patients differed from controls (ie, individuals with no prior treatment for mental illness) for each language feature. Hedges *g* was calculated using the package *effsize* to show the effect size [70]. One participant did not complete the clinical history question at baseline and was therefore excluded from this analysis.

We also explored associations between language features and a continuous measure of each individual's probability of being a patient, the patient probability score (PPS) [34]. PPS scores were trained on baseline questionnaire data from a subset of participants who were seen at the NIH before the pandemic and underwent a Structured Clinical Interview for DSM-5 (Diagnostic and Statistical Manual-5). Each participant who had not been seen at NIH was assigned a PPS value based on similarity to the patient or control group. For additional information and validation, refer to the study by Chung et al [34]. We used Spearman correlations to evaluate the associations between PPS and the 4 language features listed earlier. Seven participants were missing a PPS and were excluded from the analyses.

Results

Free-Response Sample

Of the 3655 participants enrolled in the study [34], 2497 participants responded at least once to the free-response item; these participants will be referred to as "respondents." Figure 2 depicts the distribution of the respondents as a function of the

number of times they provided free-response entries. The demographics of the total sample comparing respondents and nonrespondents (ie, participants who never provided free-response entries) are reported in Table 1. There was a total of 9738 free-response item responses.

Figure 2. Distribution of free responses. (A) Response frequency. Histogram showing the number of participants who responded from 1 to 13 times to the free-response question. (B) Response length. Histogram showing the distribution of word count across all responses, which ranged from 1 to 744 words.

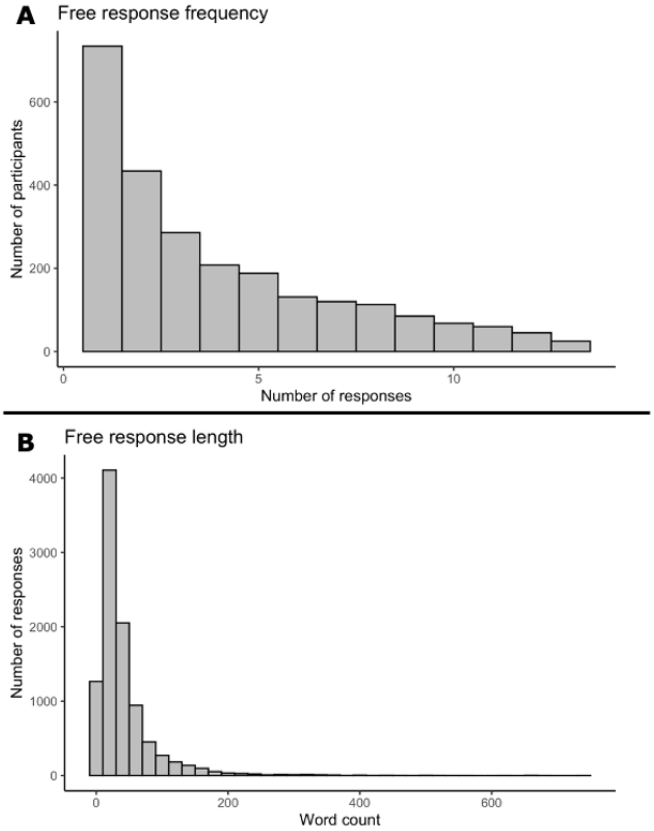


Table 1. Comparison of respondents and nonrespondents[a].

	Respondent, n (%)	Nonrespondent, n (%)	Responding, OR[b] (95% CI)
Gender			
Female	2065 (56.4)	866 (23.7)	1 (—[c])
Male	346 (9.5)	255 (7)	0.57[d] (0.48-0.68)
Nonbinary	58 (1.6)	21 (0.6)	1.16 (0.71-1.96)
Unknown	28 (0.8)	16 (0.4)	—
Ethnicity			
Hispanic or Latino	124 (3.4)	79 (2.2)	0.71[e] (0.53-0.95)
Not Hispanic or Latino	2271 (62.1)	1026 (28.1)	1 (—)
Unknown	102 (2.8)	53 (1.5)	—
Race			
American Indian or Alaska Native	7 (0.2)	4 (0.1)	—
Asian	43 (1.2)	47 (1.3)	0.40[d] (0.26-0.62)
Black or African American	62 (1.7)	44 (1.2)	0.62[e] (0.42-0.93)
White or Caucasian	2256 (61.7)	997 (27.3)	1 (—)
Hawaiian or Pacific Islander	0 (0)	0 (0)	—
Multiple races	92 (2.5)	48 (1.3)	0.85 (0.60-1.22)
Unknown	37 (1)	18 (0.5)	—
Age (years)			
Values, mean (SD)	48.0 (14.9)	43.7 (14.4)	1.02[d] (1.02-1.03)
Income (US $)			
<35,000	340 (9.3)	165 (4.5)	1 (—)
35,001-75,000	629 (17.2)	294 (8)	1.04 (0.82-1.31)
75,001-100,000	395 (10.8)	166 (4.5)	1.15 (0.89-1.50)
100,001-150,000	505 (16.4)	243 (6.6)	1.01 (0.79-1.28)
≥150,000	598 (16.4)	274 (7.5)	1.06 (0.84-1.34)
Unknown	30 (0.8)	16 (0.4)	—
Education			
Less than high school	4 (0.1)	6 (0.2)	—
High school graduate or above	52 (1.4)	41 (1.1)	1 (—)
Some college or above	179 (4.9)	130 (3.6)	1.09 (0.67-1.73)
Associate degree or above	111 (3)	71 (1.9)	1.23 (0.74-2.04)
Bachelor’s degree or above	788 (21.6)	366 (10)	1.70[e] (1.10-2.60)
Advanced or professional degree	1355 (37.1)	540 (14.8)	1.98[f] (1.29-3.01)
Unknown	8 (0.2)	4 (0.1)	—
Mental health status			
Mental health history	1384 (37.9)	609 (16.7)	1 (—)
No mental health history	1113 (30.5)	548 (15)	0.89 (0.78-1.03)
Unknown	0 (0)	1 (0)	—
Physical health status			
Has medical illness	1369 (37.5)	522 (14.3)	1 (—)

	Respondent, n (%)	Nonrespondent, n (%)	Responding, OR[b] (95% CI)
Does not have medical illness	1128 (30.9)	635 (17.4)	0.68[d] (0.59-0.78)
Unknown	0 (0)	1 (0)	—

[a]This table compares the demographics and clinical history of participants who responded at least once to the free-response question ("respondents") and those who did not ("nonrespondents"). The odds ratio of responding for each group compared with the reference group is shown in the third column. The reference groups were denoted by those with an odds ratio of 1.

[b]OR: odds ratio.

[c]Groups with too small a sample size and those whose demographics were unknown were not included in the logistic regression.

[d]Mean values <0.001.

[e]Mean values between 0.05 and 0.01.

[f]Mean values between 0.01 and 0.001.

Comparing Respondents With Nonrespondents

Logistic regressions indicated that the likelihood of responding was influenced by several demographic factors, including gender, race, ethnicity, education, and age, as reported in Table 1. For example, the odds of male participants responding compared with female participants were 43% lower (P<.001), and the odds of Asian and Black participants responding compared with White participants were 60% (P<.001) and 38% (P=.02) lower, respectively. Although education influenced the likelihood of responding such that the odds of participants with bachelor's or advanced degrees responding compared with participants who were high school graduates were 70% (P=.02) and 98% (P=.002) higher, respectively, we did not observe any influence of income. For additional demographic factors, please refer to Table 1.

Interestingly, there was no impact of mental health history on an individual's likelihood of providing free responses (Table 1). However, physical health history did influence an individual's likelihood of providing free responses, and the odds of participants without physical health conditions responding compared with those with these conditions was 32% lower (P<.001).

Impact of Psychological State on Likelihood of Providing a Free Response

We used multilevel models to evaluate the likelihood of an individual providing a response on a given week as a function of time and psychological state. All models revealed that an individual's likelihood of providing a free-response decreased over time (Table 2), although the effects were quite small based on odds ratios. Individuals were more likely to respond to the free-response item when feeling more distressed, as measured by the Kessler-5, and individuals with higher average distress were more likely to respond. Interestingly, we observed an interaction between within-subjects distress and between-subjects distress such that the effect of distress on the likelihood of responding for a given week was strongest for individuals with low average distress, perhaps because individuals with high average distress responded consistently over time. There was no effect of loneliness on the likelihood of responding when it was included in the same model as the distress measure; however, the fixed effects of loneliness and distress were correlated across individuals (r=0.633), and we therefore computed separate models for each predictor (Multimedia Appendix 1). Modeling distress alone confirmed the findings from the model that included all factors with similar coefficients. When loneliness was included in a separate model, we found that individuals were more likely to respond when they reported higher loneliness (B=0.05; P=.004) and that individuals who reported being more lonely on average were more likely to respond (B=0.09; P<.001).

Table 2. Multilevel logistic model examining association among distress, loneliness, and likelihood of response[a].

Predictor	OR[b] (95% CI)	*P* value
(Intercept)	0.37 (0.34-0.40)	<.001
Week	0.98 (0.98-0.98)	<.001
Distress	1.13 (1.11-1.15)	<.001
Mean distress	1.03 (1.01-1.06)	.005
Loneliness	0.98 (0.95-1.02)	.35
Mean loneliness	1.05 (0.99-1.10)	.10
Distress × mean distress	0.99 (0.98-0.99)	<.001
Loneliness × mean loneliness	0.98 (0.96-1.01)	.18

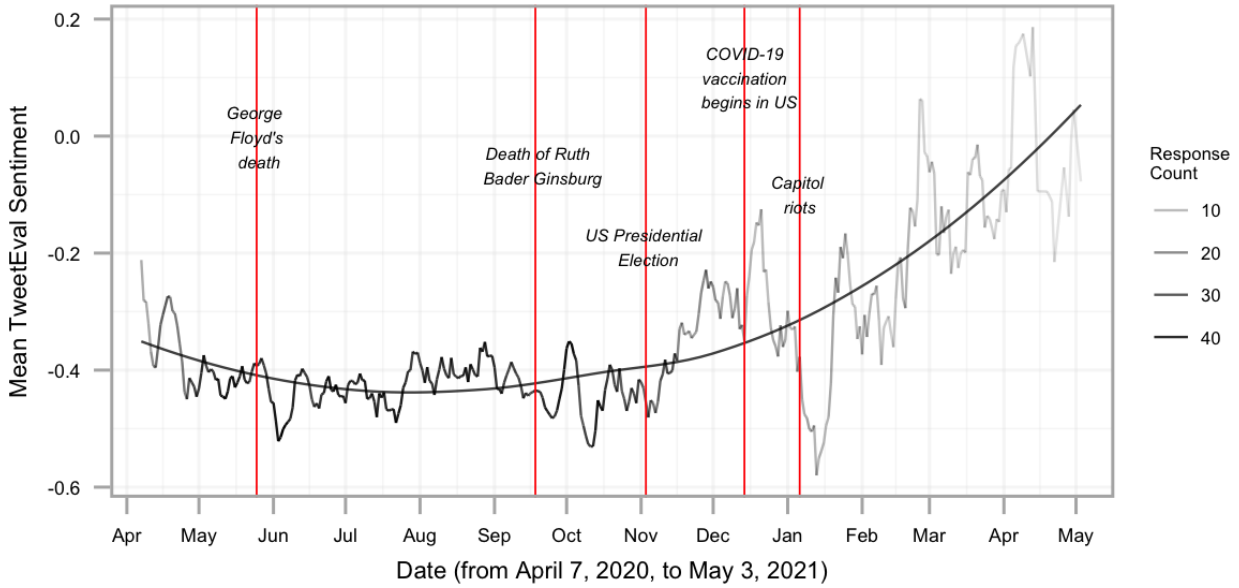
[a]This table presents the results of a multilevel logistic model examining the association between the likelihood of response on a given week and self-reported distress (measured using the Kessler-5) and loneliness (measured using the 3-item Loneliness Scale). Distress and loneliness were modeled both within (ie, dynamic fluctuations across intervals) and between participants (ie, mean distress and mean loneliness). There were 26,073 observations across 3163 individuals, with intraclass correlation coefficient=0.49, marginal R^2=0.017, conditional R^2=0.495, random error variance (σ^2)=3.29, and variance of random intercepts ($\tau_{00SUBJECT\_NUMBER}$)=3.11. The results of the models that separately analyzed distress and loneliness are presented in Multimedia Appendix 1.

[b]OR: odds ratio.

Sentiment During the Study Period

The results of sentiment analysis are shown in Figure 3. As the TweetEval scores range from −1 to 1, it is clear from the figure that the average sentiment of free responses remained negative for the entire study period. We observed a gradual upward tendency in sentiment starting in November, which coincides with announcements about the Pfizer vaccine (Figure 3). However, our sample size and proportion of responses were reduced at this time, and we did not run statistical analyses on the influence of time on sentiment; therefore, we do not make strong inferences about these overall patterns based on group averages.

Figure 3. Sentiment over time: this figure plots the 7-day rolling average of sentiment by day from April 7, 2020, to May 3, 2021 (responses from before April 7, 2020, or after May 3, 2021, are omitted due to the 7-day rolling average). The opacity of the line represents the 7-day rolling average of response count. TweetEval Sentiment below 0 is considered negative. Red vertical bars mark the dates of major national events in the United States, which emerged in free-response comments based on term frequency–inverse document frequency.



Important events throughout the pandemic that may have affected groupwide sentiment are marked in Figure 3. These events were selected based on the keywords seen in the TF-IDF analysis (see analysis below in Themes of Free Responses Across Time). The selected events were important events in the United States, given that most of the study participants came from the United States, with all 50 states represented; of free-response respondents, 2474 were based in the United States and 23 were international. The 5 events chosen were the death of George Floyd (May 25, 2020), the death of Ruth Bader Ginsburg (September 18, 2020), the 2020 US Presidential Election (November 3, 2020), the beginning of COVID-19

vaccination in the United States (December 14, 2020 [71]), and the US Capitol attack (January 6, 2021). As depicted in Figure 3, these events were followed by steep changes in the average sentiment of responses, as measured by TweetEval.

Association Between Psychological State and Sentiment of Responses

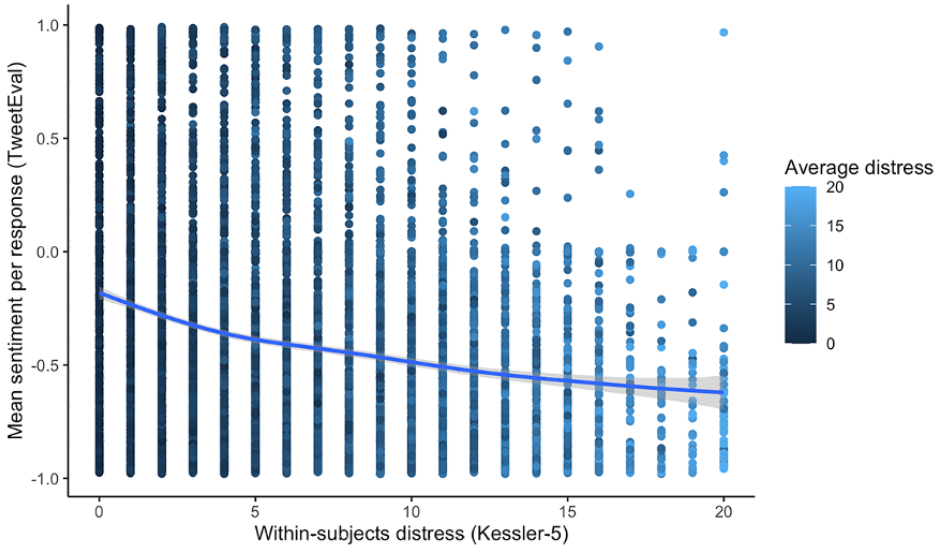
We used multilevel models to evaluate the dynamic association between self-reported psychological states and response sentiment, as measured by the mean TweetEval score per response. A model that combined distress and loneliness (Table 3) indicated that sentiment was negative on average, based on the intercept, and that sentiment increased over time within individuals, which is consistent with the overall average depicted in Figure 3. Responses were more negative at time points when individuals reported greater distress (Figure 4) or loneliness. We also observed that individuals with higher mean distress had more negative sentiment on average (Figure 4) and that there was a substantial interaction between within-subjects distress and between-subjects distress, such that the effect of distress on sentiment was strongest for those with low average distress scores. Between-subjects variations in loneliness did not influence sentiment when loneliness was included in the same model as distress; however, when loneliness and distress were modeled separately, we observed substantial associations with each measure, both within and between participants (Multimedia Appendix 1).

Table 3. Linear mixed model examining association among distress, loneliness, and likelihood of response[a].

Predictors	Estimates (95% CI)	*P* value
(Intercept)	−0.385 (−0.396 to −0.373)	<.001
Week	0.002 (0.001 to 0.004)	<.001
Distress	−0.038 (−0.043 to −0.033)	<.001
Mean distress	−0.021 (−0.024 to −0.017)	<.001
Loneliness	−0.019 (−0.029 to −0.010)	<.001
Mean loneliness	0.002 (−0.007 to 0.010)	.68
Distress × mean distress	0.002 (0.001 to 0.004)	<.001
Loneliness × mean loneliness	−0.002 (−0.009 to 0.004)	.49

[a]This table presents the results of a linear mixed model examining the associations between negative sentiment, self-reported distress, and loneliness (see the *Methods* section). There were 9253 observations across 2314 individuals, with intraclass correlation coefficient=0.14, marginal R^2=0.064, conditional R^2=0.199, random error variance (σ^2)=0.17, variance of random intercepts ($\tau_{00SUBJECT\_NUMBER}$)=0.03, and variance of random slopes=0. The results from the models that separately analyzed distress and loneliness are presented in Multimedia Appendix 1.

Figure 4. Association between distress (Kessler-5) and sentiment (TweetEval score) in free responses. Scatterplot illustrating association between biweekly measures of distress (as measured by Kessler-5) and mean sentiment of free responses as measured by TweetEval. Linear mixed models indicate that distress is negatively associated with sentiment within individuals, and that individuals with higher mean distress (visualized in lighter blue) use more negative language on average.



Association Between Psychological State and Response Length

We tested whether response length varied as a function of loneliness, distress, and time (Table 4). Response length ranged from 1 word to a maximum of 744 words (mean 41.63, SD 49.05; median 27; Figure 2). Linear mixed models indicated that response length decreased slightly over the course of an individual's participation, such that each biweekly interval was 0.5 words shorter on average. Response length was positively associated with distress (Table 4), such that an increase of 1 unit of distress on a given week was associated with an additional 0.8 words, and individuals who report higher distress provided responses that were 0.8 words longer on average. There was no effect of loneliness on response length when distress and loneliness were included in the same model, but separate analyses indicated that responses were longer in lonely individuals than in nonlonely individuals (P<.001), such that an increase of one unit in average loneliness was associated with 1.3 more words on average (Multimedia Appendix 1).

Table 4. Linear mixed model examining association among distress, loneliness, and response length[a].

Predictors	Estimates (95% CI)	*P* value
(Intercept)	36.726 (35.294 to 38.157)	<.001
Week	−0.249 (−0.358 to −0.140)	<.001
Distress	0.811 (0.361 to 1.261)	<.001
Mean distress	0.827 (0.371 to 1.282)	<.001
Loneliness	−0.081 (−0.988 to 0.826)	.86
Mean loneliness	0.166 (−0.867 to 1.200)	.75
Distress × mean distress	−0.044 (−0.162 to 0.073)	.46
Loneliness × mean loneliness	−0.235 (−0.851 to 0.381)	.46

[a]This table presents the results of a linear mixed model examining associations between word count, self-reported distress, and loneliness (see the *Methods* section). There were 9272 observations across 2314 individuals, with intraclass correlation coefficient=0.33, marginal R^2=0.009, conditional R^2=0.340, random error variance (σ^2)=1406.63, variance of random intercepts ($\tau_{00SUBJECT\_NUMBER}$)=705.36, variance of random slopes for distress=1.96, and variance random slopes for loneliness=17.25. The results from the models that separately analyzed distress and loneliness are presented in Multimedia Appendix 1.

Associations Between Mental Health History and Language Features

As reported in Table 5, individuals who reported prior mental health treatment had more negative sentiments (as measured by TweetEval), wrote longer responses, used more negative emotional words, and had higher frequencies of FPSP use. The effect sizes and P values are presented in Table 5. We also observed small but substantial associations with language features when we used our continuous PPS: PPS was associated with more negative sentiment (r=−0.12; P<.001), higher word counts (r=0.09; P<.001), more negative emotional words (r=0.04; P<.05), and higher FPSP use (r=0.13; P<.001).

Table 5. Relationship between mental health history and language use[a].

	Sentiment (TweetEval)	Word count	Negative emotional words (%)	First-person singular pronoun (%)
Mental health history, mean (SD)	−0.0044 (0.0034)	37.1 (33.8)	5.53 (6.15)	7.46 (4.77)
No mental health history, mean (SD)	−0.0038 (0.0036)	34.2 (30.2)	4.77 (4.42)	6.56 (4.46)
P value	<.001	.03	<.01	<.001
Hedges *g* effect size	−0.15	0.09	0.14	0.19

[a]A 2-sample *t* test (2-tailed) was run to compare the use of 4 language features by mental health history, as determined by mental health or drug or alcohol treatment or mental health hospitalization. The language features selected were the same as those used in the Spearman correlation analysis.

Themes of Free Responses Across Time

The results of the manual coding are reported in Table 6. The most frequently annotated categories were mental health or emotion (5159/9738, 53% of responses), social or physical distance (2475/9738, 25% of responses), and policy or government (1938/9738, 20% of responses). The Fleiss κ coefficient for IRR for all responses was 0.73 (95% CI 0.72-0.73), which is characterized as "good" agreement between raters [59]. The Gwet AC1 statistic coefficient for IRR for all responses was 0.96 (95% CI 0.96-0.96) or "very good" agreement [59]. Agreement for individual categories is presented in Multimedia Appendix 1. Only 2 categories ("non–health-related concern for the immediate circle" and "clarification of survey response") were characterized as "fair" by the Fleiss κ statistic; all others ranged from moderate to very good agreement.

Table 6. Manual coding of free-response topics[a].

Theme and category	Example response	Coding count (N=9738), n (%)
Mental health		
General negative mental health (ie, negative emotion or cognitive symptom)	"I have experienced a lot of physical symptoms of stress/anxiety, including fatigue and pain."	5159 (53)
Clinically significant (eg, mention of diagnosis, treatment, suicidality, or domestic violence)	"I have been on Prozac for 3 weeks now"	1837 (9)
Mood disorder	"I'm experiencing some depression, but I'm not having suicidal thoughts. I have a hard time thinking about a future that is different than it is now."	475 (5)
Anxiety disorder	"I took wellbutrin a few years ago, started a new RX this week for anxiety over Covid"	309 (3)
Other psychiatric diagnosis	"PTSD[b] symptoms re-activated by feeling trapped, uncertainty, amount of unknowns, untrustworthy authority figure"	275 (3)
Suicidality	"had to call suicide prevention hotline due to crisis"	130 (1)
Physical health		
Non–COVID-19–related physical health	"Symptoms listed above due to chronic asthma/allergies"	1444 (15)
Change in health behaviors, activities, or hobbies	"I am eating more and gaining weight"	798 (8)
Suspected or confirmed COVID-19 illness or self-test	"I did antibody testing with a home-kit because I was worried about my symptoms"	486 (5)
Sleep	"My sleep schedule has been completely thrown off."	366 (4)
Deferred medical care	"Close friend diagnosed with cancer and treatment was delayed due to pandemic so it spread faster than expected."	159 (2)
Drugs and alcohol	"Watching way too much TV; and smoking a heck of a lot more than I used to."	154 (2)
COVID-19–related risk factors in self	"Constant worry because I have asthma/COPD[c]"	149 (2)
Pregnancy	"I am pregnant with my first child and am very nervous about contracting COVID or my partner contracting COVID."	42 (1)
Social factors		
Experience with social or physical distance and masks	"extended family has different beliefs about social distancing which increase stress"	2475 (25)
Health condition or health-related concern about immediate circle	"Most of my stress is related to a sick family member (not Covid)"	1697 (17)
Providing care for dependents	"It is increasingly challenging to work full-time and parent children who are attending school at home."	570 (6)
Strained relationships	"My relationship with my spouse has been more rocky. It's been a lot to rely on one introverted person for my extroverted needs."	547 (6)
Non–health-related concern for immediate circle	"I'm anxious about the fears and anxieties of my closest friend. He's not handling the virus threat well at all."	513 (5)
Mention of non–COVID-19–related death	"Mother passed away from pancreatic cancer"	393 (4)
Loneliness or isolation	"I am beginning to notice the lack of and to miss physical presence and physical contact with people – besides my partner."	308 (3)
Positive relationships	"I live in a beautiful place with my wife, who is the love of my life."	172 (2)
Mention of COVID-19–related death	"My sister in law died from COVID-19 after 23 days in ICU[d]. Buried her last Saturday."	95 (1)
Career and finances		
Other work-related issues	"It has been more difficult to focus on work while working from home."	1545 (16)

Theme and category	Example response	Coding count (N=9738), n (%)
School-related changes (student or teacher)	"Back to work as a teacher. Very stressful because of fear of getting sick and because of changes in responsibilities and increased work load."	716 (7)
Experience or concern about reduction or loss in work or unemployment	"The biggest stressor is waiting to find out if I'm going to be laid off and worrying for my kids' futures in terms of getting a job."	491 (5)
Essential worker or in health care	"I am tested weekly for COVID due to my work in assisted living facilities"	357 (4)
Personal finances	"Approximate 85% of my COVID stress is related to financial uncertainty."	317 (3)
Society or government or community		
Policy or government	"My current distress and hopelessness is largely driven by the murder of George Floyd."	1938 (20)
Reopening or return to work and interactions with community	"My state is starting to reopen, but I think it's too soon."	1047 (11)
Effects of pandemic on the economy or society	"I feel so sad and scared for the world right now."	446 (5)
Other		
Positive aspects	"I feel extremely lucky to be healthy, housed and have no financial or relationship worries."	1291 (13)
Survey feedback	"I think inquiring about dietary intake and weight change would be interesting."	488 (5)
Other	"I am a veterinarian"	485 (5)
Vaccine	"I've received my first dose of the COVID vaccine, and a number of my family members are now fully vaccinated. That's given me some hope."	416 (4)
Clarification of survey response	"I checked the box for using marijuana, but I only use CBD[e] for sleep and back pain. CBD was not an option."	362 (4)
Minimal change to lifestyle	"I am a loner by nature so self-isolating is not a problem for me."	55 (1)

[a]This table presents the frequency and percentage of responses manually annotated as a function of category, as well as example responses. For the original category names, definitions, and interrater reliability by category, see Multimedia Appendix 1.

[b]PTSD: posttraumatic stress disorder.

[c]COPD: chronic obstructive pulmonary disease.

[d]ICU: intensive care unit.

[e]CBD: cannabidiol.

We also used TF-IDF to identify unique topics that emerged over time during the first year of the pandemic and therefore would not have been captured by our free-response categories. Table 7 presents the lemmas with the highest TF-IDF scores for each month. This analysis confirms the link between free responses and national and international events, such as the initial "lockdown" during the months of April and May 2020; responses to the killing of George Floyd in June 2020; wildfires and the death of Ruth Bader Ginsberg in September 2020; the lead up to the November 3, 2020, election; and the widespread availability of vaccines starting in February 2021. These events are visualized in Figure 3 to highlight the associations between events and the overall sentiment of the language. We also present word clouds of language use in Figure 5.

Table 7. Results of term frequency–inverse document frequency (TF-IDF) per month[a].

Month and lemma	Term frequency, n	TF-IDF
April 2020		
Time	146	0.0010
Home	128	0.0009
Distance	58	0.0008
Guilt	12	0.0008
Worry	108	0.0007
May 2020		
Distance	107	0.0008
Home	204	0.0008
Time	188	0.0007
Care	90	0.0007
Patient	29	0.0007
June 2020		
George	60	0.0045
Floyd	50	0.0038
Protest	159	0.0026
Brutality	30	0.0023
Riot	48	0.0020
July 2020		
Wear	115	0.0008
Care	105	0.0007
People	191	0.0006
Time	191	0.0006
Distance	88	0.0006
August 2020		
Time	187	0.0007
Distance	88	0.0007
Wear	84	0.0007
College	36	0.0006
Min	6	0.0006
September 2020		
Wildfire	37	0.0017
Rbg	20	0.0017
Fire	47	0.0015
Election	75	0.0014
Ruth	9	0.0013
October 2020		
Election	124	0.0025
Supreme	9	0.0011
Political	50	0.0007
Time	162	0.0007
Upcoming	25	0.0007

Month and lemma	Term frequency, n	TF-IDF
November 2020		
Election	301	0.0102
Thanksgiving	36	0.0025
Presidential	30	0.0021
Biden	29	0.0016
Holiday	29	0.0016
December 2020		
Christmas	49	0.0057
Holiday	45	0.0042
Thanksgiving	30	0.0035
Election	22	0.0012
Test	70	0.0009
January 2021		
Capitol	60	0.0215
Insurrection	32	0.0115
Inauguration	18	0.0065
Coup	13	0.0025
Riot	13	0.0020
February 2021		
Variant	15	0.0062
Moderna	8	0.0027
Shoot	23	0.0026
21	7	0.0008
Time	31	0.0008
March 2021		
Shot	9	0.0039
Moderna	8	0.0035
Shoot	20	0.0029
Pfizer	7	0.0021
People	31	0.0010

[a]The top 5 lemmas with the highest TF-IDF scores per month are listed.

Figure 5. Most frequent topics based on topic modeling: the top 10 most frequent topics, depicted as word clouds. The word clouds contain the top 5 single-word terms per topic as identified by latent Dirichlet allocation and the top 20 multiword terms as identified by Linguistic patterns, inverse document frequency, and C-value information.

Topic Model

The word clouds representing the 10 most frequent topics are shown in Figure 5. Although there is a substantial overlap with the most common topics found by manual coding (eg, topics relating to mental health, government, and social distancing), there are also several topics that build upon the results of manual coding by providing further context or identifying new topics (eg, various topics relating to work and family). All word clouds can be found in Multimedia Appendix 1.

Discussion

Principal Findings

We examined the use of a free-response question among individuals enrolled in a study on mental health during the COVID-19 pandemic, both in terms of the characteristics of its respondents and the content of their responses. Participants who responded to the free-response question were less likely to be male, Asian, Black, or Hispanic. They were also more likely to respond the older they were, the more education they had, and when they had physical health conditions. We found that, although participants with a mental health treatment history were not more likely to respond than those without prior histories, they had more negative sentiments, used more FPSPs and more negative emotional words, and provided longer responses. Participants responded more negatively and provided longer responses when they felt more distressed. Loneliness also predicted negative sentiment and an individual's likelihood of responding at a given time point. The overall sentiment of the sample responses remained stable and constantly negative over the period of most responses from May 2020 to November 2020. However, there were marked fluctuations that seem to have been related to external events, which were also mentioned in the free responses based on our analyses of unique topics per month. Finally, most of the responses addressed negative emotional and mental health impacts, experiences with social distancing, and political and national events. Here, we discuss these findings and their implications for mental health during the COVID-19 pandemic, as well as future studies on text analysis in mental illness.

Free Response Likelihood Varied Based on Demographic Factors

Similar to many other survey-based studies assessing the impact of the COVID-19 pandemic on physical and mental health and daily life [72-74], our overall study population was skewed, with respondents who were mainly White, educated, and female [34]. Interestingly, even within this skewed sample, those who responded to the free-response question were even more likely to be female, White, non-Hispanic, more educated, and older. Those who responded to the free-response question were also more likely to have a history of physical illness. These findings are consistent with previous studies of responses to free-response questions in patient satisfaction surveys [17-19]. Demographic differences must be considered when evaluating the generalizability of the results, and future qualitative studies should incorporate additional approaches to ensure equitable inclusion or to determine why some groups of participants may be more likely than others to complete open-ended free-response items. Nonetheless, we believe that these findings are meaningful in considering the dynamic relationship between mental health and language use within individuals and over time during the pandemic.

Mental Health History Impacts Language Use but Not Likelihood of Responding

In contrast to the hypotheses, individuals with a mental health treatment history were not more likely to respond than those without prior histories. However, individuals with mental health treatment histories had more negative sentiment, more FPSP use, higher word count, and used more negative emotional words. When judged according to traditional interpretations of correlation coefficients, correlations between the PPS (a continuous measure of each individual's likelihood of being a patient based on machine learning) and language features range from negligible to weak at best [75]. However, these associations seem more substantial when judged relative to other relationships between language features and individual differences. One study correlating Linguistic Inquiry and Word Count features and personality traits from Facebook posts highlighted all correlations with a magnitude greater than $r>0.1$ [76]. In addition, 2 meta-analyses of the relationship between FPSP use and depression found a correlation of 0.13 [26] and 0.19 [24]. Tølbøll [24] also reported a correlation of 0.12 between negative emotional word use and depression severity. Edwards and Holtzman [26] commented that the correlation between FPSP use and depression is among the largest when language features are related to individual differences. In this context, the correlation with language features, particularly FPSP use ($r=0.13$), in this study appears to be substantial. Our findings further expand on the literature by using a unique format of text, a sample with nondifferentiated mental health, and a continuous metric of mental health.

Psychological State Affects Free Response Probability and Sentiment of Responses

Language use was linked not only to mental health history and demographic factors but also to psychological state, as measured by self-reported distress and loneliness. Fluctuations in these factors predict whether an individual is likely to respond at a given time point, the sentiment of the response, and the response length. The fact that psychological state predicted the likelihood of responding suggests that perhaps responding to the free-response question was therapeutic, that is, an outlet to express negative emotions or a method of coping with loneliness. This is supported by the fact that the likelihood of responding and response length decreased over time within individuals, suggesting an overall reduction in motivation to provide free responses over the course of the 6-month study, but that within-subject fluctuations in distress and loneliness all increased response likelihood, response length, and sentiment. Furthermore, this was directly reported by some participants in their free responses, for example, "Thank you for listening"; "I'm disappointed that the study is ending while my world is still so broken. This is one of the few outlets I have to make people listen to how I'm being affected"; "I'm SO bummed this is ending. I've enjoyed doing this"; "I'm not happy that the study is ending while my life is still profoundly

diminished. This is one of the few places I have a voice to make it known how badly the restrictions and isolation are affecting me"; and "I have appreciated being a part of this research. It helped me to be in better touch with myself." Of course, other participants felt differently, for example, "I am so glad to finish this off. It was a poorly designed survey (what is 'normal'?) and it was so biased trying to show the panic over a deadly but unavoidable disease" and "Well well...we are at the end of this study. I hope the NIH will read it's [*sic*] own articles on the distress caused by wearing masks."

Average Sentiment Was Consistently Negative

Across individuals, not surprisingly, the sentiment over the course of the COVID-19 pandemic was generally negative. Other studies performing sentiment analyses on social media with overlapping time ranges during the pandemic have found overall negativity [77,78], overall positivity [73,79,80], or mixed results [81]. This variability is likely a result of the type of language assessed, location of the participants, period, and sentiment analysis algorithm. The consistent negativity we observed could reflect self-selection, that is, individuals may have responded to the question when they felt the worst. This interpretation is supported by results showing that participants were more likely to respond and more likely to have negative responses when they felt more distressed or lonelier. Although the sentiment of this study's sample cannot be quantitatively compared with others in similar studies, the high prevalence of mental health disorders in this sample (1384/2497, 55.43% of free-response respondents had a history of treatment for mental health) may also have contributed to negativity because of the pandemic's mental health exacerbations for those with preexisting mental illness [82].

We found that sentiment remained generally stable during the period of most responses and actually increased over time within individuals, contradicting the hypothesis that sentiment would become more negative as quarantine and infection control measures go on, Although increases in global sentiment were not evident until December, around the time of the Food and Drug administration approval of the Pfizer vaccine. The visualization of average sentiment over time suggests that the main source of variability came from major spikes in the TweetEval measure at several points throughout the pandemic, which may have been tied to US national events. For example, negative spikes followed the murder of George Floyd and the Capitol Riots, and positive spikes followed the news of vaccine efficacy and approval. These fluctuations occurred within 2 weeks of the event, which likely reflected the survey being e-mailed to participants every 2 weeks. Importantly, because most responses were between May 2020 and November 2020, interpretations of fluctuations and trends outside of that duration are less reliable. Those inflection points could be true spikes or a result of a few responses. However, these topics also emerged in our TF-IDF analysis, which quantifiably shows a timeline of events during the pandemic based on the unique words used each month. For example, words about the murder of George Floyd appeared in June 2020 and words about the US Capitol attack appeared in January 2021. This suggests that individuals used the free-response item to discuss timely topics of national importance, and the association with changes in sentiment suggests they reported on their feelings about these events.

Automated and Manual Scoring Provide Insights on Topics Relevant to the Pandemic

An interesting finding is that LDA was successful in identifying relevant topics despite the small size of the data set. Although we used perplexity as a measure to evaluate LDA outputs, the number of topics and topics were human interpretable. We evaluated them using human validations in 3 steps. In addition to the automated sentiment, TF-IDF, and topic model analyses, we used traditional manual coding to evaluate the presence of categories that were expected to be discussed. The most frequent topic across participants was negative impacts on mental health and emotion, including worries, stressors, and concerns (with a frequency of 5159/9741, 52.96%). This is likely to reflect the fact that our study was advertised by the National Institute of Mental Health and had a high (1384/2497, 55.43%) percentage of individuals with a history of mental health treatment. Individuals may have been primed to reflect on the mental health impacts of the pandemic because that was the focus of the overall study, so this result might not be generalizable to the broader population. Participants also frequently discussed social or physical distancing (2475/9741, 25.41%) and policy and government (1938/9741, 19.9%), which included mentions of protests and leadership. These results align with another study [83] that performed a content analysis on social media posts during the pandemic and found that the top negative topic was "Frustration due to life disruptions." Surprisingly, the number of respondents who spoke about COVID-19–related physical illness or testing and COVID-19–related death remained relatively small. These findings may indicate that what was most salient to many participants at the time of responding was the impact of the pandemic on their mental health and life rather than the infection itself. However, an alternative explanation could be that the number of people who reported contracting COVID-19 was small (of the full survey sample of 3655 participants, only 95 reported testing positive for COVID-19 at any point during the study) or that participants were able to provide all pertinent information using other questionnaire measures, thereby reducing the need to expand on this using the free-response item.

The IRR was calculated for manual coding. Although Fleiss κ and Gwet AC1 statistics were chosen to complement each other, both showed contradicting discrepancies. As expected, Fleiss κ was unusually low for the infrequently annotated categories. Fleiss κ has a notable weakness in dealing with data with low variability [57,58]. In contrast, Gwet AC1 statistic was >0.90 for all but the top 3 annotated categories, suggesting that it artificially inflated the score for categories with low variabilities. This contradiction is best represented by the "Minimal Change to Lifestyle" and "Clarification of survey response" categories, in which Fleiss κ is 0.45 and 0.38 and Gwet AC1 is >0.99 and 0.97, respectively. Similarly, 1 study [84] found large discrepancies between the κ and AC1 statistic with skewed data. In particular, Keener [84] noted that more disagreements between coders led to higher discrepancies between the 2 statistics, with Gwet AC1 statistic being higher, as in this study. Both statistics are reported here to portray IRR in full context.

Strengths, Limitations, and Outstanding Questions

There were several key strengths to this study. To our knowledge, this is the first study to analyze language during the COVID-19 pandemic for such a long duration, which facilitated the analysis of dynamic within-person changes in mental health and language use. Other studies that applied text mining methods to language during the COVID-19 pandemic had a much shorter duration, captured data from the early stages of the pandemic only, and drew their samples from social media [83,85-90]. In studies on language use in survey responses during the COVID-19 pandemic, this study appears to have the broadest sample and scope. Other studies that analyzed similar texts in surveys during the pandemic targeted more specific populations [74,91,92] or fields [93,94]. We integrated multiple unique and complementary approaches to text analysis, including sentiment analysis, automated text analysis, and human coding, which provided a fuller picture of the qualitative data. Having such a large sample of participants with a history of mental illness provided the power to compare language as a function of whether individuals had prior mental illness. In addition, the literature largely focuses on the relationship between language and depression, whereas our sample represents a broader spectrum of mental health.

Our study also differs from previous work in that we analyzed text from an open-ended free response, whereas most prior studies on mental health and language have been based on data collected from social media. Focusing on texts from social media may bias the findings in favor of respondents who openly post about their mental health on social media. A previous study showed that the same language features that predicted depression and posttraumatic stress disorder on Twitter (assessed through the user's self-declaration of diagnosis) overlapped with features that predicted demographics and personality [95]. This suggests that the language features most commonly used to predict mental health may be confounded by the language used by those who are more likely to post about their mental health on social media, where most such studies collect their data. Thus, it is important to assess the relationship between language features and mental health in a variety of texts and samples beyond social media and depression. We believe that our findings indicate that even an open-ended free-response measure can provide important insights into general distress and mental health.

There are also limitations to the study that may reduce generalizability or explain why some results did not replicate the literature. First, as mentioned earlier, the sample demographics for the survey were highly skewed, and the participants who self-selected to answer the free-response question were representative of neither the study nor the general population. This problem plagues all studies that analyze free-text comment responses [19]. Second, our study enrolled participants for a 6-month period starting in April 2020, and recruitment ceased in November 2020. We had higher enrollment at the start of the study than at the end; therefore, the response rate was lower after November 2020, reducing the sample size for interpretation after that date. This means that we cannot make reliable inferences on the sentiment or content of a language after the large-scale introduction of vaccines. Third, the variability in response length and frequency might limit the accuracy or generalizability of our findings. Finally, links with mental health relied on self-report measures and only a subset of our participants underwent in-person clinical evaluations. Future studies should compare these approaches in the context of longer free responses and well-characterized clinical samples. Nonetheless, the results replicating the relationships between mental health and various language features found in literature support the validity of this study's measures.

In conclusion, our analysis of free responses in the context of a study on mental health during the COVID-19 pandemic reveals associations between psychological factors, language, and sentiment and identifies pertinent topics of interest during the first year of the pandemic. We found that mental health history and current psychological state (ie, distress and loneliness) impacted language use, even within a simple open-ended item. In addition, the sentiments of responses varied over time as a function of both within-person dynamics and population-level events. This provides a unique window for mental health during the pandemic, complementing other longitudinal and population-based studies that rely strictly on quantitative measures. Both manual and automated methods were used to characterize the content of these responses, providing insight into the key themes participants raised regarding how the pandemic affected them. However, a skewed sample in various demographics was a limitation and raised questions about what makes one more likely to respond to a survey or a free-text question. Future research should consider adding or using existing free-response questions for quantitative and qualitative language analysis. Future research on language and mental health should also make a deliberate effort to build samples that are diverse in demographics, types of text, and mental health conditions.

Acknowledgments

This work was supported in part by the Intramural Research Programs of the National Institute of Mental Health (principal investigators JYC, ZIAMH002922 and FP, ZIC-MH002968) and the National Center for Complementary and Integrative Health (principal investigator LYA, ZIA-AT000035). The authors thank Molly Cosgrove for assistance with initial selection and validation of free-response coding categories, Alison Gibbons for overall study coordination, and the other members of the National Institute of Mental Health Office of the Clinical Director COVID-19 study team for helpful conversations. The authors also thank the many respondents who shared their experiences throughout the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information about methods of demographic classification, automated topic analysis and topic modeling, as well as supplemental results (ie, word clouds depicting results of topic modeling) and tables (manual coding categories and definitions; results of separate logistic and linear models).
[DOCX File , 1646 KB - mental_v10i1e40899_app1.docx]

References

1. Mental health in emergencies. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-in-emergencies [accessed 2021-08-06]
2. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
3. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
4. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
5. Pancani L, Marinucci M, Aureli N, Riva P. Forced social isolation and mental health: a study on 1,006 Italians under COVID-19 lockdown. Front Psychol 2021 May 21;12:663799 [FREE Full text] [doi: 10.3389/fpsyg.2021.663799] [Medline: 34093358]
6. Hennein R, Mew EJ, Lowe SR. Socio-ecological predictors of mental health outcomes among healthcare workers during the COVID-19 pandemic in the United States. PLoS One 2021 Feb 5;16(2):e0246602 [FREE Full text] [doi: 10.1371/journal.pone.0246602] [Medline: 33544761]
7. Kılınçel S, Kılınçel O, Muratdağı G, Aydın A, Usta MB. Factors affecting the anxiety levels of adolescents in home-quarantine during COVID-19 pandemic in Turkey. Asia Pac Psychiatry 2021 Jun 11;13(2):e12406 [FREE Full text] [doi: 10.1111/appy.12406] [Medline: 32783389]
8. Brülhart M, Klotzbücher V, Lalive R, Reich SK. Mental health concerns during the COVID-19 pandemic as revealed by helpline calls. Nature 2021 Dec;600(7887):121-126. [doi: 10.1038/s41586-021-04099-6] [Medline: 34789873]
9. Cheng Z, Mendolia S, Paloyo AR, Savage DA, Tani M. Working parents, financial insecurity, and childcare: mental health in the time of COVID-19 in the UK. Rev Econ Househ 2021 Jan 12;19(1):123-144 [FREE Full text] [doi: 10.1007/s11150-020-09538-3] [Medline: 33456425]
10. Kim H, Jung J. Social isolation and psychological distress during the COVID-19 pandemic: a cross-national analysis. Gerontologist 2021 Jan 21;61(1):103-113 [FREE Full text] [doi: 10.1093/geront/gnaa168] [Medline: 33125065]
11. Maunder RG. Was SARS a mental health catastrophe? Gen Hosp Psychiatry 2009 Jul;31(4):316-317 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.04.004] [Medline: 19555790]
12. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, et al. Personality, gender, and age in the language of social media: the open-vocabulary approach. PLoS One 2013;8(9):e73791 [FREE Full text] [doi: 10.1371/journal.pone.0073791] [Medline: 24086296]
13. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Language Social Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]
14. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020 Mar 24;3(1):43 [FREE Full text] [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
15. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Oct 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
16. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. Proc Seventh Int AAAI Conference Weblogs Social Media 2021;7(1):128-137. [doi: 10.1609/icwsm.v7i1.14432]
17. Rich JL, Chojenta C, Loxton D. Quality, rigour and usefulness of free-text comments collected by a large population based longitudinal study - ALSWH. PLoS One 2013 Jul 11;8(7):e68832 [FREE Full text] [doi: 10.1371/journal.pone.0068832] [Medline: 23874784]
18. Cunningham M, Wells M. Qualitative analysis of 6961 free-text comments from the first National Cancer Patient Experience Survey in Scotland. BMJ Open 2017 Jun 15;7(6):e015726. [doi: 10.1136/bmjopen-2016-015726] [Medline: 28619780]

19. Garcia J, Evans J, Reshaw M. ``Is there anything else you would like to tell us" – methodological issues in the use of free-text comments from postal surveys. Quality Quantity 2004 Apr;38(2):113-125. [doi: 10.1023/b:ququ.0000019394.78970.df]
20. Zhang Y, Lyu H, Liu Y, Zhang X, Wang Y, Luo J. Monitoring depression trends on Twitter during the COVID-19 pandemic: observational study. JMIR Infodemiol 2021;1(1):e26769 [FREE Full text] [doi: 10.2196/26769] [Medline: 34458682]
21. Al-Mosaiwi M, Johnstone T. In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. Clin Psychol Sci 2018 Jul;6(4):529-542 [FREE Full text] [doi: 10.1177/2167702617747074] [Medline: 30886766]
22. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: analyzing the language of mental health on twitter through self-reported diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun 5, 2015; Denver, Colorado. [doi: 10.3115/v1/w15-1201]
23. Lyons M, Aksayli ND, Brewer G. Mental distress and language use: linguistic analysis of discussion forum posts. Comput Human Behav 2018 Oct;87:207-211. [doi: 10.1016/j.chb.2018.05.035]
24. Tølbøll K. Linguistic features in depression: a meta-analysis. J Language Works - Sprogvidenskabeligt Studentertidsskrift 2019;4(2):39-59 [FREE Full text]
25. Wolf M, Sedway J, Bulik CM, Kordy H. Linguistic analyses of natural written language: unobtrusive assessment of cognitive style in eating disorders. Int J Eat Disord 2007 Dec;40(8):711-717. [doi: 10.1002/eat.20445] [Medline: 17683092]
26. Edwards T, Holtzman NS. A meta-analysis of correlations between depression and first person singular pronoun use. J Res Personality 2017 Jun;68:63-68 [FREE Full text] [doi: 10.1016/j.jrp.2017.02.005]
27. Sekulic I, Gjurković M, Šnajder J. Not just depressed: bipolar disorder prediction on Reddit. In: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 2018 Presented at: 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; Oct, 2018; Brussels, Belgium. [doi: 10.18653/v1/w18-6211]
28. Shen J, Rudzicz F. Detecting anxiety through Reddit. In: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality. 2017 Presented at: Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality; Aug, 2017; Vancouver, BC. [doi: 10.18653/v1/w17-3107]
29. Weerasinghe J, Morales K, Greenstadt R. “Because... I was told... So much”: linguistic indicators of mental health status on twitter. Proc Privacy Enhancing Technologies Symp 2019;2019(4):152-171. [doi: 10.2478/popets-2019-0063]
30. Stirman SW, Pennebaker JW. Word use in the poetry of suicidal and nonsuicidal poets. Psychosom Med 2001;63(4):517-522. [doi: 10.1097/00006842-200107000-00001] [Medline: 11485104]
31. Berry‐Blunt AK, Holtzman NS, Donnellan MB, Mehl MR. The story of “I” tracking: psychological implications of self‐referential language use. Soc Personal Psychol Compass 2021 Oct 19;15(12):e12647. [doi: 10.1111/spc3.12647]
32. Rude S, Gortner E, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cognit Emotion 2004 Dec;18(8):1121-1133. [doi: 10.1080/02699930441000030]
33. Mor N, Winquist J. Self-focused attention and negative affect: a meta-analysis. Psychol Bull 2002 Jul;128(4):638-662. [doi: 10.1037/0033-2909.128.4.638] [Medline: 12081086]
34. Chung J, Gibbons A, Atlas L, Ballard E, Ernst M, Japee S, et al. COVID-19 and mental health: predicted mental health status is associated with clinical symptoms and pandemic-related psychological and behavioral responses. medRxiv 2021 Oct 14:2021.10.12.21264902 (forthcoming) [FREE Full text] [doi: 10.1101/2021.10.12.21264902] [Medline: 34671781]
35. Psychosocial impact of COVID-19 survey. Phenx Toolkit. 2020 Feb 6. URL: https://www.phenxtoolkit.org/toolkit_content/PDF/Psychosocial_Impact_of_COVID-19_Survey.pdf [accessed 2023-02-06]
36. Parker K, Minkin R, Bennett J. Methodology. Pew Research Center. 2020 Sep 24. URL: https://www.pewresearch.org/social-trends/2020/09/24/covid-19-financial-hardships-methodology/ [accessed 2021-12-20]
37. Wei X, Croft W. LDA-based document models for ad-hoc retrieval. In: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. 2006 Presented at: SIGIR06: The 29th Annual International SIGIR Conference; Aug 6 - 11, 2006; Seattle Washington USA. [doi: 10.1145/1148170.1148204]
38. Grün B, Hornik K. topicmodels: an R Package for fitting topic models. J Stat Softw 2011;40(13):1-30. [doi: 10.18637/jss.v040.i13]
39. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. In: Proceedings of the 22nd international conference on World Wide Web. 2013 Presented at: WWW '13: 22nd International World Wide Web Conference; May 13 - 17, 2013; Rio de Janeiro Brazil. [doi: 10.1145/2488388.2488514]
40. Lossio-Ventura JA, Jonquet C, Roche M, Teisseire M. Biomedical term extraction: overview and a new methodology. Inf Retrieval J 2016;19(1-2):59-99. [doi: 10.1007/s10791-015-9262-2]
41. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 01;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]

42. Frantzi K, Ananiadou S, Mima H. Automatic recognition of multi-word terms:. the C-value/NC-value method. Int J Digit Libr 2000 Aug 1;3(2):115-130. [doi: 10.1007/s007999900023]
43. WordCloud for Python documentation. WordCloud. URL: https://amueller.github.io/word_cloud/ [accessed 2022-06-08]
44. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys. Res Aging 2016 Aug 19;26(6):655-672. [doi: 10.1177/0164027504268574]
45. Measuring the social and emotional wellbeing of Aboriginal and Torres Strait Islander peoples. Australian Institute of Health and Welfare. 2009 Jun 30. URL: https://www.aihw.gov.au/reports/indigenous-australians/measuring-the-social-and-emotional-wellbeing/summary [accessed 2022-06-12]
46. Qi P, Zhang Y, Zhang Y, Bolton J, Manning C. Stanza: a python natural language processing toolkit for many human languages. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2020 Presented at: 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations; Jul, 2020; Online. [doi: 10.18653/v1/2020.acl-demos.14]
47. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. Proc Int AAAI Conference Web Social Media 2014 May 16;8(1):216-225. [doi: 10.1609/icwsm.v8i1.14550]
48. Pennebaker J, Booth R, Boyd R, Francis M. Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates; 2015.
49. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. J Am Soc Inf Sci 2010 Dec 15;61(12):2544-2558. [doi: 10.1002/asi.21416]
50. TextBlob: Simplified Text Processing. TextBlob. URL: https://textblob.readthedocs.io/en/dev/ [accessed 2022-01-25]
51. Leys J, Peirsman Y. Bert-base-multilingual-uncased-sentiment. Hugging Face. URL: https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment [accessed 2022-05-19]
52. Pérez J, Giudici J, Luque F. pysentimiento: a python toolkit for sentiment analysis and SocialNLP tasks. arXiv 2021 Jun 17. [doi: 10.48550/ARXIV.2106.09462]
53. Barbieri F, Camacho-Collados J, Espinosa AL, Neves L. TweetEval: unified benchmark and comparative evaluation for tweet classification. arXiv 2020 Oct 26 [FREE Full text] [doi: 10.18653/v1/2020.findings-emnlp.148]
54. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv 2019 Jul 26.
55. Zeileis A, Grothendieck G. zoo: S3 infrastructure for regular and irregular time series. J Stat Softw 2005;14(6):1-27. [doi: 10.18637/jss.v014.i06]
56. Gwet KL. irrCAC: Computing Chance-Corrected Agreement Coefficients (CAC). CRAN R Project. 2019 Sep 23. URL: https://CRAN.R-project.org/package=irrCAC [accessed 2022-01-25]
57. Walsh P, Thornton J, Asato J, Walker N, McCoy G, Baal J, et al. Approaches to describing inter-rater reliability of the overall clinical appearance of febrile infants and toddlers in the emergency department. PeerJ 2014;2:e651 [FREE Full text] [doi: 10.7717/peerj.651] [Medline: 25401054]
58. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013 Apr 29;13(1):61 [FREE Full text] [doi: 10.1186/1471-2288-13-61] [Medline: 23627889]
59. Altman D. Practical Statistics for Medical Research. Boca Raton, Florida, United States: CRC Press; 1990.
60. Rinker T. textstem: Tools for stemming and lemmatizing text. version 0.1.4. GitHub. 2018. URL: http://github.com/trinker/textstem [accessed 2022-01-25]
61. Silge J, Robinson D. tidytext: text mining and analysis using tidy data principles in R. J Open Source Softw 2016 Jul;1(3):37. [doi: 10.21105/joss.00037]
62. Blei D, Ng A, Jordan M. Latent dirichlet allocation. J Mach Learn Res 2003;3(2):993-1022.
63. Chang J, Gerrish S, Wang C, Boyd-Graber J, Blei D. Reading tea leaves: how humans interpret topic models. In: Proceedings of the 22nd International Conference on Neural Information Processing Systems. 2009 Presented at: 22nd International Conference on Neural Information Processing Systems; Nov 9-12, 2015; Istanbul, Turkey.
64. Hoyle A, Goel P, Hian-Cheong A, Peskov D, Boyd-Graber J, Resnik P. Is automated topic model evaluation broken? The incoherence of coherence. arXiv 2021 Jul 5.
65. Mimno D, Wallach H, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: EMNLP '11: Proceedings of the Conference on Empirical Methods in Natural Language Processing; Jul 27 - 31, 2011; Edinburgh United Kingdom.
66. Hagen L. Content analysis of e-petitions with topic modeling: how to train and evaluate LDA models? Inform Process Manag 2018 Nov;54(6):1292-1307. [doi: 10.1016/j.ipm.2018.05.006]
67. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
68. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]
69. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]

70. Torchiano M. Effsize - a package for efficient effect size computation. Zenodo. 2016 Nov 13. URL: https://zenodo.org/record/1480624#.Y8rcV3ZBzIU [accessed 2021-08-10]
71. COVID-19 Vaccines. HHS.gov. URL: https://www.hhs.gov/coronavirus/covid-19-vaccines/index.html [accessed 2023-01-31]
72. Ali SH, Foreman J, Capasso A, Jones AM, Tozan Y, DiClemente RJ. Social media as a recruitment platform for a nationwide online survey of COVID-19 knowledge, beliefs, and practices in the United States: methodology and feasibility analysis. BMC Med Res Methodol 2020 May 13;20(1):116 [FREE Full text] [doi: 10.1186/s12874-020-01011-0] [Medline: 32404050]
73. Lossio-Ventura JA, Lee AY, Hancock JT, Linos N, Linos E. Identifying silver linings during the pandemic through natural language processing. Front Psychol 2021;12:712111 [FREE Full text] [doi: 10.3389/fpsyg.2021.712111] [Medline: 34539512]
74. Michaud K, Wipfler K, Shaw Y, Simon TA, Cornish A, England BR, et al. Experiences of patients with rheumatic diseases in the United States during early days of the COVID-19 pandemic. ACR Open Rheumatol 2020 Jun;2(6):335-343 [FREE Full text] [doi: 10.1002/acr2.11148] [Medline: 32311836]
75. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018 Sep;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
76. Kern ML, Eichstaedt JC, Schwartz HA, Dziurzynski L, Ungar LH, Stillwell DJ, et al. The online social self: an open vocabulary approach to personality. Assessment 2014 Apr 08;21(2):158-169. [doi: 10.1177/1073191113514104] [Medline: 24322010]
77. Chivers BR, Garad RM, Boyle JA, Skouteris H, Teede HJ, Harrison CL. Perinatal distress during COVID-19: thematic analysis of an online parenting forum. J Med Internet Res 2020 Sep 07;22(9):e22002 [FREE Full text] [doi: 10.2196/22002] [Medline: 32857707]
78. Garcia K, Berton L. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. Appl Soft Comput 2021 Mar;101:107057 [FREE Full text] [doi: 10.1016/j.asoc.2020.107057] [Medline: 33519326]
79. Chandra R, Krishna A. COVID-19 sentiment analysis via deep learning during the rise of novel cases. PLoS One 2021 Aug 19;16(8):e0255615 [FREE Full text] [doi: 10.1371/journal.pone.0255615] [Medline: 34411112]
80. Machuca CR, Gallardo C, Toasa RM. Twitter sentiment analysis on coronavirus: machine learning approach. J Phys Conf Ser 2021 Feb 01;1828(1):012104. [doi: 10.1088/1742-6596/1828/1/012104]
81. Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE. Sentiment analysis of COVID-19 tweets by Deep Learning Classifiers-A study to show how popularity is affecting accuracy in social media. Appl Soft Comput 2020 Dec;97:106754 [FREE Full text] [doi: 10.1016/j.asoc.2020.106754] [Medline: 33013254]
82. Gobbi S, Płomecka MB, Ashraf Z, Radziński P, Neckels R, Lazzeri S, et al. Worsening of preexisting psychiatric conditions during the COVID-19 pandemic. Front Psychiatry 2020 Dec 16;11:581426 [FREE Full text] [doi: 10.3389/fpsyt.2020.581426] [Medline: 33391049]
83. Oyebode O, Ndulue C, Adib A, Mulchandani D, Suruliraj B, Orji FA, et al. Health, psychosocial, and social issues emanating from the COVID-19 pandemic based on social media comments: text mining and thematic analysis approach. JMIR Med Inform 2021 Apr 06;9(4):e22734 [FREE Full text] [doi: 10.2196/22734] [Medline: 33684052]
84. Keener A. Comparison of Cohen's Kappa and Gwet's AC1 with a mass shooting classification index: a study of rater uncertainty. Shareok. 2020 May. URL: https://shareok.org/handle/11244/325442 [accessed 2022-04-16]
85. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
86. Koh JX, Liew TM. How loneliness is talked about in social media during COVID-19 pandemic: text mining of 4,492 Twitter feeds. J Psychiatr Res 2022 Jan;145:317-324 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.11.015] [Medline: 33190839]
87. Li I, Li Y, Li T, Alvarez-Napagao S, Garcia-Gasulla D, Suzumura T. What are we depressed about when we talk about COVID-19: mental health analysis on tweets using natural language processing. In: Artificial Intelligence XXXVII. Cham: Springer; 2020.
88. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: observational study. J Med Internet Res 2020 Oct 12;22(10):e22635 [FREE Full text] [doi: 10.2196/22635] [Medline: 32936777]
89. Osakwe ZT, Ikhapoh I, Arora BK, Bubu OM. Identifying public concerns and reactions during the COVID-19 pandemic on Twitter: a text-mining analysis. Public Health Nurs 2021 Mar 30;38(2):145-151 [FREE Full text] [doi: 10.1111/phn.12843] [Medline: 33258149]
90. Saha K, Torous J, Caine ED, De Choudhury M. Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. J Med Internet Res 2020 Nov 24;22(11):e22600 [FREE Full text] [doi: 10.2196/22600] [Medline: 33156805]
91. Davidson CJ, Lodge K, Kam A. The impact of the COVID-19 pandemic on autistic adults – a survey. AIA 2021 Feb 24;7(4):311-321. [doi: 10.1108/aia-10-2020-0057]
92. Dickerson J, Kelly B, Lockyer B, Bridges S, Cartwright C, Willan K, Bradford Institute for Health Research COVID-19 Scientific Advisory Group, Bradford Institute for Health Research Covid-19 Scientific Advisory Group. 'When will this end? Will it end?' The impact of the March-June 2020 UK COVID-19 lockdown response on mental health: a longitudinal

survey of mothers in the Born in Bradford study. BMJ Open 2022 Jan 11;12(1):e047748 [FREE Full text] [doi: 10.1136/bmjopen-2020-047748] [Medline: 35017230]
93. Shoesmith E, Shahab L, Kale D, Mills DS, Reeve C, Toner P, et al. The influence of human-animal interactions on mental and physical health during the first covid-19 lockdown phase in the U.K.: a qualitative exploration. Int J Environ Res Public Health 2021 Jan 22;18(3):976 [FREE Full text] [doi: 10.3390/ijerph18030976] [Medline: 33499354]
94. Varma P, Burge M, Meaklim H, Junge M, Jackson ML. Poor sleep quality and its relationship with individual characteristics, personal experiences and mental health during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Jun 03;18(11):6030 [FREE Full text] [doi: 10.3390/ijerph18116030] [Medline: 34205195]
95. Preoţiuc-Pietro D, Eichstaedt J, Park G, Sap M, Smith L, Tobolsky V, et al. The role of personality, age, and gender in tweeting about mental illness. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun 5, 2015; Denver, Colorado. [doi: 10.3115/v1/w15-1203]

Abbreviations

DSM-5: Diagnostic and Statistical Manual-5
FPSP: first-person singular pronoun
IRR: interrater reliability
LDA: latent Dirichlet allocation
NIH: National Institutes of Health
PPS: patient probability score
SVD: singular value decomposition
TF-IDF: term frequency–inverse document frequency

Edited by J Torous; submitted 08.07.22; peer-reviewed by H Lyu, D Low; comments to author 18.09.22; revised version received 29.11.22; accepted 15.12.22; published 01.03.23.

Please cite as:
Weger R, Lossio-Ventura JA, Rose-McCandlish M, Shaw JS, Sinclair S, Pereira F, Chung JY, Atlas LY
Trends in Language Use During the COVID-19 Pandemic and Relationship Between Language Use and Mental Health: Text Analysis Based on Free Responses From a Longitudinal Study
JMIR Ment Health 2023;10:e40899
URL: https://mental.jmir.org/2023/1/e40899
doi:10.2196/40899
PMID:36525362

Review

Determinants of Patient Use and Satisfaction With Synchronous Telemental Health Services During the COVID-19 Pandemic: Systematic Review

Ariana Neumann[1], MSc; Hans-Helmut König[1], MPH, Prof Dr; Josephine Bokermann[1], MSc; André Hajek[1], Prof Dr

Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author:
Ariana Neumann, MSc
Department of Health Economics and Health Services Research
University Medical Center Hamburg-Eppendorf
Martinistraße 52
Hamburg, 20246
Germany
Phone: 49 40 7410 54202
Fax: 49 40 7410 40261
Email: ar.neumann@uke.de

Abstract

Background: Several recent studies examined patient use and satisfaction with synchronous telemental health services in response to the widespread implementation during the COVID-19 pandemic. However, a systematic review of recent literature on the determinants of these outcomes is missing.

Objective: The aim of this systematic review was to give an extensive overview of the literature on and highlight the influential determinants of patient use and satisfaction with synchronous telemental health services during the COVID-19 pandemic.

Methods: This review satisfied the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and was registered in PROSPERO. Peer-reviewed, quantitative studies that observed the determinants of patient use or satisfaction with synchronous telemental health services during the COVID-19 pandemic were included. PubMed, PsycInfo, and Web of Science database searches were conducted in August 2022 for English and German language studies published from 2020 onward. Key steps were performed by 2 reviewers. Determinants were synthesized into major categories informed by the dimensions of the widely used and established Unified Theory of Acceptance and Use of Technology.

Results: Of the 20 included studies, 10 studies examined determinants of patient use, 7 examined determinants of patient satisfaction, and 3 observed both outcomes. The quality of the studies was mainly good or fair. There was substantial heterogeneity in the study designs, methods, and findings. Sociodemographic characteristics and health-related determinants were mostly considered. Some of the major dimensions of the Unified Theory of Acceptance and Use of Technology were neglected in recent studies. Although most findings were mixed or nonsignificant, some indications for potential relationships were found (eg, for sex, age, and symptom severity).

Conclusions: The findings revealed potential target groups (eg, female and young patients with mild symptoms) for future postpandemic telemental health interventions. However, they also identified patient groups that were harder to reach (eg, older patients with severe symptoms); efforts may be beneficial to address such groups. Future quantitative and qualitative research is needed to secure and expand on recent findings, which could help improve services.

Trial Registration: PROSPERO CRD42022351576; https://tinyurl.com/yr6zrva5

(JMIR Ment Health 2023;10:e46148) doi:10.2196/46148

KEYWORDS

telemedicine; digital health; teletherapy; mental health; use; satisfaction

Introduction

Background

Over the past 3 decades, health care services were usually delivered in person. Telemedicine is a promising, alternative service delivery model. The World Health Organization [1] summarized the four core characteristics of telemedicine as follows: (1) its purpose is to provide clinical support; (2) it is intended to overcome geographic barriers, connecting users who are not in the same physical location; (3) it involves the use of various types of information and communication technology; and (4) its goal is to improve health outcomes. Telemedicine benefits have been evaluated in the past and include, for example, reduced costs and improved access to services and information [2,3]. Evidence also suggests that telemedicine, in general, is a clinically and cost-effective tool with high satisfaction in patients and health care professionals [4]. However, the implementation of telemedicine has often been hindered by multiple barriers regarding reimbursement and clinical, legal, sustainability, and social issues [5,6].

In the wake of the COVID-19 pandemic, rapid changes in the delivery of health care services had to be made to prevent further spread of the virus, to protect people at higher risk of severe illness from COVID-19 (eg, patients with cancer, cardiovascular disease, or chronic respiratory disease), and to relieve the strain on the health care system. Consequently, telemedicine has been used worldwide across multiple specialties [7-9]. For instance, a large cohort study by Weiner et al [10] reported an increase in telemedicine use from 0.3% of ambulatory contacts between March and June 2019 to 23.6% between March and June 2020 among privately insured working-age individuals in the United States. Most telemedicine services were delivered via synchronous video or telephone calls during those periods [8].

The outbreak of the COVID-19 pandemic was also linked to stressors such as restrictions in everyday life, lifestyle changes, social isolation, and uncertainty and worries regarding health, finances, and work, which caused psychological burden [11]. Consequently, multiple studies have observed an increase in public mental health problems [12,13]. Liu et al [12] included 71 papers in their meta-analysis and detected an increased prevalence of anxiety (32.60%, 95% CI 29.10%-36.30%), depression (27.60%, 95% CI 24.00%-31.60%), insomnia (30.30%, 95% CI 24.60%-36.60%), and posttraumatic stress disorder (16.70%, 95% CI 8.90%-29.20%) during the pandemic. Moreover, preexisting mental health conditions were found to aggravate owing to the pandemic [14]. Therefore, patients with mental health conditions represented an especially vulnerable group during that time.

Telemental health services played an essential role in managing the increased public mental health burden and preventing the worsening of psychological symptoms. Mental health services are well suited for the remote format, as they do not require physical examination and can be delivered in multiple ways (eg, via telephone and video calls or mobile apps) [15]. In fact, telemental health services were found to be part of the medical specialty with the highest use rate during the pandemic [9]. The National Institute of Mental Health defined telemental health services as the use of telecommunications or videoconferencing technology to provide mental health services [16]. This can include synchronous (eg, videoconference and telephone) and asynchronous (eg, mobile apps and email) services. Regarding the effectiveness of telemental health services, an umbrella review of 19 systematic reviews on telemental health services before the pandemic suggested that remote mental health services produced at least moderate reductions in symptom severity and could be as effective as in-person formats [17]. They also found that user acceptance and satisfaction of telemental health services were comparable with those of in-person interventions. Recent reviews have also reported the effectiveness of and high patient and provider satisfaction with telemental health services during the pandemic [18,19]. Therefore, telemental health services seem to be a valuable addition to the treatment of mental illnesses of which implementation should be supported in the postpandemic future [20,21].

A crucial factor in the successful implementation of telemental health services is patient acceptance. In previous research, no universal definition of technology or telemedicine acceptance was identified. However, past definitions can be sorted into four main categories, which refer to (1) the effectiveness or efficiency of the services, (2) the use or adoption of the services, (3) the intention or willingness to use the services, and (4) consumer or provider satisfaction with the services [22-26]. To set a more precise focus, this systematic review concentrates only on patient use and satisfaction. In the course of this systematic review, patient use includes different measures of use behavior, such as the adoption of a new service, frequency of use, or attendance. Multiple definitions of patient satisfaction were introduced in the past and include various perspectives. For example, the expectancy-disconfirmation model defines consumer satisfaction as a function of expectation and expectancy disconfirmation, which can influence attitude change and purchase intention [27]. Although this definition is widely used, there is a lack of consensus regarding the definition of satisfaction [28]. The systematic review by Giese and Joseph [28] summarized three essential components of consumer satisfaction: (1) a summary affective response, which varies in intensity; (2) satisfaction, which focuses on product choice, purchase, and consumption; and (3) time of determination, which varies by situation but is generally limited in duration.

Different theories have been introduced to explain why patients accept telemedicine services. The Unified Theory of Acceptance and Use of Technology (UTAUT) [23] was thereby one of the most frequently used theories to predict patient acceptance of telemedicine [29]. In this theory, the key determinants of behavioral intention and technology use behavior are performance expectancy, effort expectancy, social influence, and facilitating conditions. In the context of telemedicine, performance expectancy is the degree to which an individual believes that using telemedicine could be helpful. Effort expectancy refers to the perceived ease of using the service, which also includes the effect of factors such as computer anxiety and computer self-efficacy. Furthermore, social influence means the degree to which an individual believes that others think that they should use telemedicine. Facilitating

conditions include perceived organizational and technical infrastructure to support the use of telemedicine. Additional influential constructs in this theory include gender, age, experience, and voluntariness of use. User satisfaction was also found to be associated with major UTAUT constructs and to potentially contribute to the service reuse intentions [30,31].

Objective

In addition to theoretical models, only few systematic reviews have summarized the determinants of patient use or satisfaction with telemental health services from prepandemic studies [32,33]. Potential determinants that were observed in these reviews were sex, age, education, socioeconomic status, living arrangement, cognitive function, experience with telehealth technology, comfort with using the internet, satisfaction with the health care provider, experience with the clinic, and cultural background [32,33]. Nevertheless, these reviews also highlighted the need for further research on this topic. The rapid, extensive implementation of synchronous telemental health services during the COVID-19 pandemic sparked international interest in the topic. Several studies examined the determinants of patient use and satisfaction with telemental health services since the pandemic. However, a systematic review of recent literature is missing.

Conducting such a systematic review may be helpful in identifying target groups, as well as groups that need further attention and support in relation to telemental health services. This could be of major importance to successfully implement postpandemic telemental health interventions and benefit from the remote format in the future, where it can be a valuable tool to deal with challenges, such as population aging (ie, shortage of health care professionals and increased demand for long-term care), stigma attached to visiting mental health facilities and undersupply in rural areas [34,35]. Moreover, it could be useful to identify gaps in the literature and guide future research. Therefore, the objective of this systematic review was to give an extensive overview of the literature on and highlight the influential determinants of patient use and satisfaction with synchronous telemental health services during the COVID-19 pandemic. In other words, this systematic review examined the following research question: what are the determinants of patient use of and satisfaction with synchronous telemental health services in studies conducted during the COVID-19 pandemic?

Methods

Overview

The systematic review protocol is available in PROSPERO (registration number: CRD42022351576). This manuscript was written in accordance with the most recent version of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [36].

Eligibility Criteria

For this systematic review, peer-reviewed quantitative studies in German or English that observed determinants of patient use or satisfaction with synchronous telemental health services during the COVID-19 pandemic were included. Only peer-reviewed quantitative studies were considered to assure high quality of the included studies. As most of the telemedicine services were delivered via synchronous services during the pandemic [8] and to assure comparability among the studies, only synchronous telemental health services were included. Mental health patients of all age groups (ie, children, adolescents, and middle- and older-aged adults) were considered to obtain as much information as possible from recent studies. Therefore, studies were excluded if they referred to (1) asynchronous services or eHealth interventions, (2) exclusively individuals with physical illnesses (to assure comparability among the samples), (3) data that were collected before the COVID-19 pandemic, (4) qualitative data, (5) outcomes that were not related to the use or satisfaction with telemental health services, or (6) studies that did not examine determinants of use or satisfaction with the services.

Search Strategy

We searched the PubMed, PsycInfo, and Web of Science databases for studies published from 2020 onward. The PubMed and Web of Science databases are well established and frequently used in medical and related research fields. Moreover, they have also been recommended for searching telemedicine-related studies [37]. In addition, the PsycInfo database was included to account for the mental health context. A predefined search query was used to filter the databases (see Table 1 for the PubMed search query). Moreover, reference lists of eligible studies were screened for additional relevant articles. A pretest including 100 titles and abstracts was conducted before the screening process started.

Table 1. Search strategy (PubMed).

Serial number	Search term	Limits (filter, limits, and refine)
1	*telepsychiatry OR online therap* OR telepsychology OR teleconferenc* OR teleconsult* OR online consult* OR videoconferenc* OR video consult* OR phone consultation* OR telephone OR telemental* OR teletherapy OR video call OR televideo OR telehealth OR telemedicine*	• Text word
2	*satisfaction OR utilization OR engagement OR usage OR adherence OR patient satisfaction OR patient engagement*	• All fields
3	*predict* OR determin* OR associat* OR correlat**	• All fields
4	#1 *AND* #2 *AND* #3	• Publication years: 2020-2022 • Language: English and German • Species: humans

Selection Process

In August 2022, all the results from the different databases were imported to EndNote (Clarivate), where duplicates were removed. For the next step, 2 reviewers (AN and JB) independently screened the titles and abstracts of the studies, followed by a full-text screening (Cohen κ=0.61). The Rayyan web application was used to support the double-screening process [38]. Disagreements (15/144, 10.4% of studies) were resolved via discussion and consultation with a third reviewer (AH) when needed.

Data Collection Process

Relevant data from articles that passed the full-text screening were extracted by 1 reviewer (JB) and crosschecked by a second reviewer (AN) using an Excel spreadsheet (Microsoft Corp). The information that was extracted included study characteristics (author, year, study design, country, study period, and data source), population characteristics (sample size, sex, and age), setting (psychiatric care setting and telemental health service type), outcome definition, determinants, analytic approach, and key findings. For missing information or for reasons of clarification, the corresponding authors of the studies were contacted.

Quality Assessment

The risk of bias was assessed by 2 reviewers independently (AN and JB) using the assessment tool for observational cohort and cross-sectional studies by the National Heart, Lung and Blood Institute [39]. Disagreements were resolved via discussion and consultation with a third reviewer (AH) when needed.

Synthesis Methods

A formal narrative synthesis of the study results was conducted following the current reporting guidelines for syntheses without meta-analysis in systematic reviews [40]. General study characteristics were summarized in a tabular format. Key findings concerning the determinants of patient use and satisfaction were grouped into categories based on the UTAUT constructs. The UTAUT constructs were adapted and extended depending on the focus of the different studies and the pandemic context. The final categories included performance expectancy, effort expectancy, facilitating conditions, and experience. Age and gender were included into a larger category that contained sociodemographic determinants. The social influence category was adapted to include psychosocial influence to account for the special pandemic situation. Owing to the pandemic circumstances, voluntariness of use was excluded as a category because there was often no option to choose between in-person and telemental health visits. In addition, health- and service-related factors were added as categories to account for potential satisfaction-specific determinants. A meta-analysis of the results was not conducted because of the high heterogeneity across the study designs, outcomes, and effect measures. However, regression coefficients, correlations, and odds ratios were reported when available. In addition, if available, related CIs were specified to assess the certainty of the findings.

Results

Quality Assessment

The ratings for study quality are summarized in Tables 2 and 3. Most studies were rated as being of either good (n=12) or fair (n=6) quality. The quality criteria that were most commonly not met in the different studies were the reporting of participation rates (20% fulfilled) and sample size justification, power description or variance, and effect estimates (10% fulfilled).

Table 2. Quality assessment for the included studies (studies [41-50]).

Criteria	Studies									
	Ainslie et al [41]	Ceniti et al [42]	Chakawa et al [43]	Connolly et al [44]	Guinart et al [45]	Haxhihamza et al [46]	Hutchison et al [47]	Lewis et al [48]	Lohmiller et al [49]	Lynch et al [50]
1. Was the research question or objective in this paper clearly stated?	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No
2. Was the study population clearly specified and defined?	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
3. Was the participation rate of eligible persons at least 50%?	N/A[a]	NR[b]	NR	N/A	No	CD[c]	No	Yes	NR	No
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No
5. Was a sample size justification, power description, or variance and effect estimates provided?	No	No	No	Yes	No	No	No	No	No	No
6. For the analyses in this paper, were the exposures of interest measured prior to the outcomes being measured?	N/A	N/A	N/A	N/A	N/A	N/A	Yes	N/A	N/A	N/A
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	N/A	N/A	N/A	N/A	N/A	N/A	Yes	N/A	N/A	N/A
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (eg, categories of exposure or exposure measured as continuous variable)?	Yes	Yes	Yes	Yes	Yes	CD	Yes	Yes	Yes	Yes
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes	Yes	No (patients vs providers)	Yes	No	Yes	Yes	Yes	Yes
10. Was the exposures assessed more than once over time?	Yes (2 waves)	No	Yes (2 waves)	Yes (2 waves)	No	No	Yes (before and after)	No	No	Yes (3 waves)
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12. Were the outcome assessors blinded to the exposure status of participants?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
13. Was loss to follow-up after baseline 20% or less?	Yes/	N/A	Yes	Yes	N/A	N/A	Yes	N/A	N/A	Yes
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposures and outcomes?	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes
Quality rating	Good	Good	Fair	Good	Fair	Poor	Fair	Good	Good	Fair

[a]N/A: not applicable.

[b]NR: not reported.

[c]CD: cannot determine.

Table 3. Quality assessment for the included studies (studies [51-60]).

Criteria	Studies									
	Meininger et al [51]	Michaels et al [52]	Miu et al [53]	Morgan et al [54]	Nesset et al [55]	Severe et al [56]	Sizer et al [57]	Ter Heide et al [58]	Tobin et al [59]	Vakil et al [60]
1. Was the research question or objective in this paper clearly stated?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Was the study population clearly specified and defined?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3. Was the participation rate of eligible persons at least 50%?	Yes	No	N/A[a]	NR[b]	No	No	Yes	Yes	N/A	N/A
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5. Was a sample size justification, power description, or variance and effect estimates provided?	No	No	No	Yes	No	No	No	No	No	No
6. For the analyses in this paper, were the exposures of interest measured prior to the outcomes being measured?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (eg, categories of exposure or exposure measured as continuous variable)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
10. Was the exposures assessed more than once over time?	No	No	No	No	No	No	No	No	Yes (3 waves)	Yes (2 waves)
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12. Were the outcome assessors blinded to the exposure status of participants?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
13. Was loss to follow-up after baseline 20% or less?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	Yes
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposures and outcomes?	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Quality rating	Good	Fair	Good	Good	Poor	Fair	Good	Good	Good	Good

[a]N/A: not applicable.

[b]NR: not reported.

Overview of Included Studies

After the study selection process, 20 studies remained for the final synthesis (Figure 1; see Multimedia Appendix 1 [41-60] for the citations of all included studies). The main characteristics of these studies are summarized in Tables 4 and 5.

The study samples were predominantly from North America (n=14, with 12 from the United States and 2 from Canada). Furthermore, 5 study samples were from Europe (2 from Germany, 1 from the Netherlands, 1 from Norway, and 1 from North Macedonia), and 1 study sample was from Asia (Israel). Data sources consisted of electronic medical records in 7 studies as well as samples recruited from mental health clinics and community centers in 12 studies. One study used data from a sample that was recruited through targeted emails to mental health organizations nationwide, provincial psychiatric and family physician associations, hospital newsletters, existing participant networks within Canadian Biomarker Integration Network in Depression, and social media. A total of 4 studies were published in 2020, 6 in 2021, 8 in 2022, and 2 in 2023. Although most of the data were collected during the first months of the pandemic, starting from March 2020, some studies also included data from later periods until December 2021.

Figure 1. Flow diagram of the study selection process.

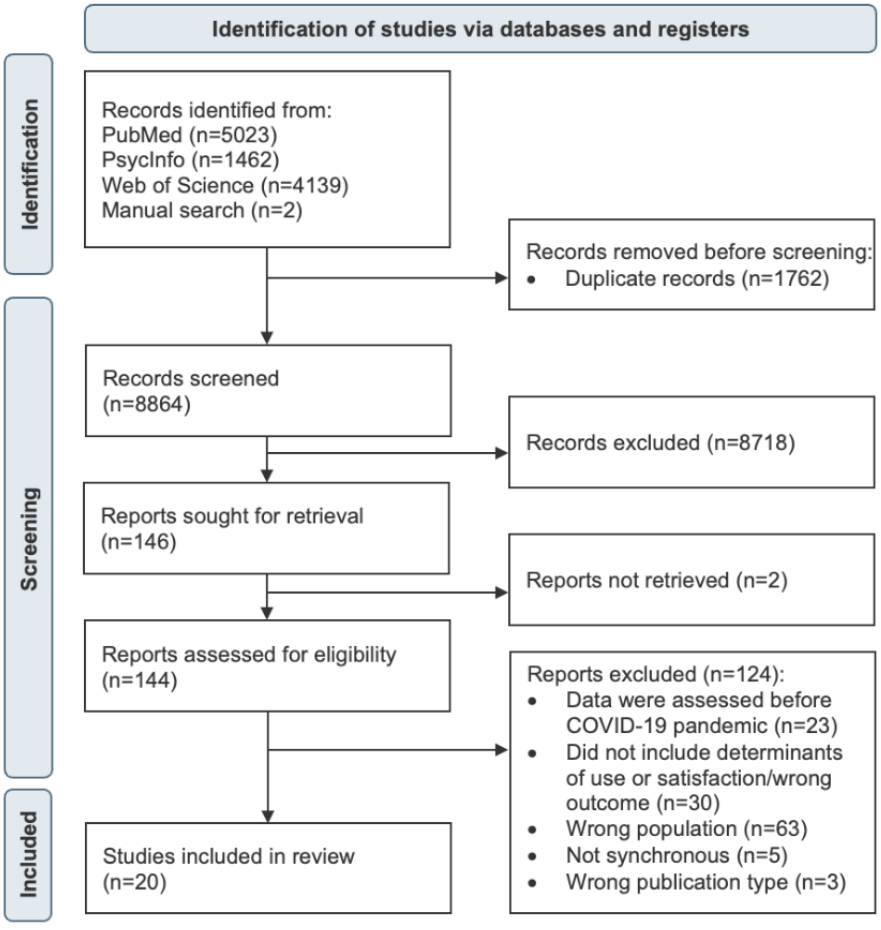


Table 4. Main characteristics of the included studies.

Study, year	Characteristics					
	Study design	Country	Study period	Data source	Population characteristics (sample size; sex: female; age [years], mean [SD])	Psychiatric care setting
Ainslie et al [41], 2022	Observational retrospective study	United States	• Study base period: December 1, 2019, to February 29, 2020 • Study retention period: April 1 to June 30, 2020 • Time trends comparison with study base period: December 1, 2018, to February 28, 2019 • Time trends comparison with study retention period: April 1 to June 30, 2019	Electronic medical record data	• Medicaid beneficiaries with SMIs[a], N=15,471 in 2020 • Sex: n=6792 (54.7%) • Age: mean age not reported	Community mental health centers
Ceniti et al [42], 2022	Cross-sectional, mixed methods study	Canada	• October 8, 2020, to February 4, 2021	Recruitment through targeted emails to mental health organizations nationwide, provincial psychiatric and family physician associations, hospital newsletters, existing participant networks within CAN-BIND[b], and social media	• Mental health care users N=332 • Sex: n=238 (71.7%) • Age: mean age not reported	General remote care experience
Chakawa et al [43], 2021	Comparative study	United States	• Before COVID-19: April to October 2019 • During COVID-19: April to October 2020	Recruited from clinic	• Children aged 1-19 years, N=226 (n=106 for in-person cases before COVID-19, n=120 for telehealth visits cases during COVID-19) • Sex: n=83 (36.7%) (before COVID-19 38.7%, during COVID-19 35%) • Age: 8.04 (4.30); before COVID-19 mean 7.0, during COVID-19 mean 8.0	Large, inner-city pediatric primary care clinic within a large regional children's hospital
Connolly et al [44], 2021	Cross-sectional study	United States	• Pre-COVID: October 1, 2017, to March 10, 2020 • COVID: March 11 to July 10, 2020	Electronic medical record	• US veterans with ≥1 mental health outpatient appointment N=2,480,119 before COVID-19; N=1,054,670 during COVID-19; N=954,704 cases from COVID-19 included in pre-COVID cases • Sex: before COVID-19 n=325,225 (13.5%), during COVID-19 n=163,186 (15.8%) • Age: mean age not reported	Department of Veterans Affairs
Guinart et al [45], 2020	Cross-sectional study	United States	• April to June 2020	Recruited from clinics and community centers	• Patients using telepsychiatry N=3052 • Sex: not reported • Age: mean age not reported	18 hospitals and community centers located in rural, suburban, small urban, and large urban areas

Study, year	Characteristics					
	Study design	Country	Study period	Data source	Population characteristics (sample size; sex: female; age [years], mean [SD])	Psychiatric care setting
Haxhihamza et al [46], 2021	Cross-sectional study	North Macedonia	• Not reported	Recruited from clinic	• Patients from the ward N=28 • Sex: n=11 (37.9%) • Age: mean 40.25 (19)	Daily hospital as a part of the University Clinic in Skopje
Hutchison et al [47], 2022	Cross-sectional study	United States	• October 2020 to June 2021	Recruited from clinic	• Adolescents aged 12-17 years, N=56 • Sex: n=37 (66.1%) • Age: 14.5 (1.6)	Community mental health clinic
Lewis et al [48], 2021	Cross-sectional study	Israel	• Mid-April to mid-May 2020	Recruited from clinic	• Eating disorder patients N=63 • Sex: n=57 (90.5%) • Age: 27.25 (11.47)	Hadarim Eating Disorders Treatment Center (part of the Shavata Mental Health Center)
Lohmiller et al [49], 2021	Cross-sectional study	Germany	• July 2020 to February 2021	Recruited from clinic	• Patients from the psychosomatic outpatient clinic N=278 • Sex: n=182 (83%) • Age: 31.5 (range 18-80)	Psychosomatic outpatient clinic at the University Hospital in Tübingen
Lynch et al [50], 2021	Cross-sectional, mixed methods study	United States	• February 2 to June 12, 2020 • Before COVID-19 (t1): February 2 to March 18, 2020 • After COVID-19 1 (t2): March 19 to April 30, 2020 • After COVID-19 2 (t3): May 1 to June 12, 2020	Recruited from clinic	• Adults with SMI (N=72; t1, n=60; t2, n=64; and t3, n=62) • Sex: t1 n=23 (38.3%), t2 n=20 (31.3%), and t3 n=21 (33.9%) • Age: t1 28.1 (10), t2 28.22 (10.7), and t3 28.45 (11.14)	Private university-affiliated outpatient psychiatric treatment center
Meininger et al [51], 2022	Cross-sectional study	Germany	• July 27 to October 22, 2020	Recruited from clinic	• Parents or caregivers answering for or with their children receiving teletherapy N=168 • Sex: n=61 (36.3%) • Age: 12.29 (4.01)	University Hospital Cologne—School for Child and Adolescent Cognitive Behavior Therapy
Michaels et al [52], 2022	Cross-sectional study	United States	• Not reported	Recruited from clinic	• College students in a postacute outpatient program who recently required psychiatric hospitalization N=101 • Sex: n=72 (74.5%) • Age: 22.5 (2.8)	Outpatient mental health clinic at a local psychiatric hospital that provides specialized postacute services to college students
Miu et al [53], 2021	Cross-sectional study	United States	• January 16 to April 30, 2020	Electronic medical record	• SMI and non-SMI patients N=1444 • Sex: n=970 (67.2%) • Age: mean age not reported	Outpatient psychiatry clinic of an urban, academic medical center
Morgan et al [54], 2021	Cross-sectional study	United States	• March 20 to June 10, 2020	Electronic medical record	• Clients in marriage and family training clinics (telehealth sample) N=142 • Sex: n=79 (55.6%) • Age: 32.56 (16.58)	A total of 2 marriage and family training clinics
		Norway				

Study, year	Characteristics				
	Study design	Country	Study period	Data source	Population characteristics (sample size; sex: female; age [years], mean [SD])	Psychiatric care setting
Nesset et al [55], 2023	Cross-sectional study		• October-December 2021	Recruited from clinic	• Patients from outpatient clinic who attended therapy for aggressive and violent behavior against their partners and children N=28 • Sex: n=7 (25%) • Age: mean age not reported	Outpatient clinic at St Olav's University Hospital, Center for Research and Education in Security, Prisons, and Forensic Psychiatry
Severe et al [56], 2020	Cross-sectional study	United States	• June-August 2020	Recruited from clinic	• Patients who had an in-person appointment date that fell in the first few weeks following the Michigan governor's stay-at-home edict, necessitating conversion to web-based visits or deferment of in-person care N=244 • Sex: n=167 (68.4%) • Age: mean age not reported	Outpatient Psychiatry Clinics at the University of Michigan
Sizer et al [57], 2022	Cross-sectional study	United States	• April 1, 2020, to March 31, 2021	Electronic medical record	• Patients from rural outpatient clinics N=1115 • Sex: n=623 (55.9%) • Age: not reported	A total of 6 Northeast Delta Human Services Authority outpatient behavioral health clinics
Ter Heide et al [58], 2021	Cross-sectional study	Netherlands	• June 3 to July 31, 2020	Recruited from clinic	• Patients with complex psychotrauma complaints N=318 • Sex: n=130 (40.9%) • Age: 52 (11.9)	ARQ Centrum '45 (National institute for diagnostics and treatment of complex psychotrauma complaints)
Tobin et al [59], 2022	Retrospective cohort study	United States	• January 1 to December 31, 2020 • Before COVID-19: January 1 to March 18, 2020 • Telehealth only: March 19 to May 31, 2020, December 1 to December 31, 2020 • Choice between telehealth and in-person services: June 1 to November 30, 2020	Electronic medical record	• Patients seen by integrated psychology team in general internal medicine N=1075 encounters • Sex: n=759 (70.6%) • Age: 49.73 (15.89)	Integrated psychology team within the general internal medicine primary care clinic at a large urban health system
Vakil et al [60], 2022	Retrospective cohort study	Canada	• Comparison sample: March 19, 2019, to March 18, 2020 • COVID-19 sample: March 19, 2020, to April 7, 2021	Electronic medical record	• Patients in need of urgent mental health assessment and treatment without referral N=3573 visits • Sex: n=1981 (55.4%) • Age: 33.9 (13.4)	Crisis Response Center

[a]SMI: serious mental illness.

[b]CAN-BIND: Canadian Biomarker Integration Network in Depression.

Table 5. Characteristics of the included studies.

Study, year	Characteristics				
	Telemental health service type (telephone vs video)	Outcome (use vs satisfaction and assessment)	Determinants	Analytic approach	Quality rating
Ainslie et al [41], 2022	All forms of telemental health services	Use: use from pandemic identified by service claim codes; categorized based on percentage of total treatment services during the retention period (low: <25%; medium: 25%-75%; high: >75%)	Sex, age group, diagnosis, and zip code (rural vs urban)	Chi-square test and logistic regression	Good
Ceniti et al [42], 2022	All forms of telemental health services	Use: number of remote visits Satisfaction: 7-point Likert scale (from total dissatisfied to total satisfied) for overall satisfaction with remote care, security, user-friendliness, speed of access and provision of care, continuity of care, convenience, maintenance of therapeutic rapport	Age, type of provider (psychiatrist or family physician vs other mental health care providers), level of connectedness with loved ones, living with others, province or territory, high-risk status for COVID-19, frequency of internet use, and number of people living at home	Chi-square test and Spearman correlation	Good
Chakawa et al [43], 2021	Video (or telephone or audio-only when there were technical problems)	Use: differences in service delivery modality use (in-person visit before COVID-19 vs telehealth use during COVID-19)	Sex, age, referral concern, health insurance type, race or ethnicity, language, controlling for primary care provider, visit control variable (assigned or familiar or not), and appointment type (first or follow-up visit)	Binominal logistic regression	Fair
Connolly et al [44], 2021	Telephone vs video vs in-person services	Use: having had any video experience (before COVID-19 vs during COVID-19); having had ≥50% of visits via phone vs video vs in person	Sex, age, socio economic status, race or ethnicity, rurality, marital status, ≥50% Department of Veterans Affairs disability rating, diagnosis, and history of mental health hospitalization	Binominal and multinomial logistic regression	Good
Guinart et al [45], 2020	Telephone vs video services	Satisfaction: overall experience (telephone or video), perceived helpfulness of remote sessions, challenges and advantages	Age and duration of care	Chi-square test	Fair
Haxhihamza et al [46], 2021	Not specified	Satisfaction: Patient Satisfaction Questionnaire (18 items with 7 dimensions of satisfaction with medical care measured by the Patient Satisfaction Questionnaire-III: general satisfaction, technical quality, interpersonal manner, communication, financial aspects, time spent with doctor, accessibility and convenience)	Age, gender, and place of living	Not specified	Poor
Hutchison et al [47], 2022	Video services	Use: attendance across sessions Satisfaction: Treatment Perception Questionnaire (10 items; general satisfaction and acceptability of mental health services); Internet Evaluation and Utility Questionnaire (15 items; ease of use, convenience, engagement, privacy, satisfaction and acceptability of an internet intervention)	Risk status for adverse mental and behavioral outcomes, and symptom severity	Bivariate correlation and *t* test	Fair

Study, year	Characteristics				
	Telemental health service type (telephone vs video)	Outcome (use vs satisfaction and assessment)	Determinants	Analytic approach	Quality rating
Lewis et al [48], 2021	Web-based platforms, not specified	Satisfaction: Telemedicine Satisfaction Questionnaire (15 items, 5-point Likert scale, 3 factors: quality of care, similarity of remote meetings to face-to-face meetings, perception of the interaction); perspective toward the transition to web-based treatment (6 self-developed statements, 1-5 Likert scale, perception of care, preference of web-based treatment to face-to-face treatment, promotion of this mode of therapy toward others)	Age, gender, education, BMI, duration of treatment in days, past eating disorder, hospitalization, Eating Disorder Examination Questionnaire, Depression, Anxiety and Stress Scales-21, Working Alliance Inventory-S, fear of COVID-19 scale-19S	*t* test and Pearson correlation	Good
Lohmiller et al [49], 2021	Telephone vs video vs in-person services	Satisfaction: self-developed questionnaire with 4 subject areas: patient characterization (10 items), assessment of therapeutic contact (12 items), therapeutic relationship (11 items), hurdles (5 items), 5 additional free-text items	Age, gender, and type of contact	Chi-square test, ANOVA, and hierarchical regression	Good
Lynch et al [50], 2021	Video services	Use: no show or cancellation frequency	Age, gender, race or ethnicity, primary diagnosis, and time period	Model building approach using generalized linear modeling with a Poisson log link (multilevel approach because of nested data structure was used)	Fair
Meininger et al [51], 2022	Video services	Satisfaction: self-developed questionnaire, 11 items: stable internet connection, overall satisfaction, intention to use teletherapy after pandemic=mean satisfaction score; changes in treatment satisfaction and changes in the therapeutic relationship=mean satisfaction change score	Corona Child Stress Scale, psychosocial functioning (Children's Global Assessment Scale, Child Behavior Checklist [6-18 R] and Youth Self Report [11-18 R]), Checklist for Screening Behavioral and Emotional Problems, and number of teletherapy sessions	Pearson correlation	Good
Michaels et al [52], 2022	Telephone vs video vs in-person services	Satisfaction: preferred telehealth method, overall experience (telephone or video), future telehealth use, perceived helpfulness of remote sessions	Sex, gender, race, and teletherapy format	Chi-square test, Mann-Whitney *U* test, and Kruskal-Wallis test	Fair
Miu et al [53], 2021	Video or telephone vs in-person services	Use: conversion rate to teletherapy for SMI[a] patients vs non-SMI patients, number of teletherapy sessions between SMI and non-SMI group, differences in new patients starting therapy via telehealth between SMI and non-SMI groups	Age, sex, ethnicity, previous engagement, and SMI vs non-SMI groups	Chi-square test and *t* test	Good
Morgan et al [54], 2021	Video and telephone services	Use: conversion to teletherapy (attendance of at least 1 teletherapy session vs opting out), engagement in teletherapy (number of teletherapy sessions)	Age, gender, race, ethnicity, relationship status, income, education, number of sessions before teletherapy, and case constellation (individual vs relational therapy)	*t* test, logistic regression, and multiple linear regression	Good
Nesset et al [55], 2023	Video services	Satisfaction: Client Satisfaction Questionnaire-8 (8 items measure respondents' perception of treatment quality)	Gender	*t* test	Poor
Severe et al [56], 2020	Video and telephone services	Use: visit type	Age, sex, race, health insurance type, and number of previous clinic visits	Multiple logistic regression	Fair

Study, year	Characteristics				
	Telemental health service type (telephone vs video)	Outcome (use vs satisfaction and assessment)	Determinants	Analytic approach	Quality rating
Sizer et al [57], 2022	Video and telephone services	Use: number of visits	Age, gender, education (number of school years), race, referral source, monthly income, discharge, chronic condition, number of diagnoses, primary diagnosis type	Negative binomial regression	Good
Ter Heide et al [58], 2021	Video services	Use: 1 item: how did you stay in touch with your therapist during the past 2 mo? (Multiple answers could be given: face-to-face, via videoconferencing, via telephone, through email or chat, not at all) Satisfaction: one item: how satisfied were you with this form of contact, rated on a scale from 0 (not at all satisfied) to 10 (as satisfied as can be)?	Age, gender, level of education, refugee status, Brief Symptom Inventory, Cantril Ladder (life satisfaction), COVID-19 stress level	Pearson product-moment correlation, MANCOVA[b], ANCOVA[c], chi-square test, binary logistic regression, and *t* test (2-tailed)	Good
Tobin et al [59], 2023	Telephone vs video vs in-person services	Use: visit type	Age, sex, race, and health insurance type	Logistic regression	Good
Vakil et al [60], 2022	Video or telephone vs in-person services	Use: visit type	Age, sex, distance to crisis response center, household income, prior visit to the center within 1 year, suicidal behavior, diagnosis, visit characteristics (day of the week, time of day, and period of pandemic)	Binary logistic regression	Good

[a]SMI: serious mental illness.

[b]MANCOVA: multivariate analysis of covariance.

[c]ANCOVA: analysis of covariance.

Patient use was examined in 10 studies [41,43,44,50,53,54,56,57,59,60], patient satisfaction in 7 studies [45,46,48,49,51,52,55], and both outcomes were observed in 3 studies [42,47,58]. Patient use was mostly defined as having at least 1 telemental health visit during the pandemic [43,44,53,54,56,58-60]. However, others have also considered the number of telemental health visits [42,53,54,57] and the percentage of telemental health services in overall mental health service use during the pandemic [41,44] or attendance [47,50]. For patient satisfaction, 6 studies used self-developed items and scales [42,45,49,51,52,58], whereas 4 studies used established instruments (ie, Telemedicine Satisfaction Questionnaire [61], Client Satisfaction Questionnaire [62], Patient Satisfaction Questionnaire [63], Treatment Perception Questionnaire [64], and Internet Evaluation and Utility Questionnaire [65]) [46-48,55]. The satisfaction questionnaires mainly focused on the overall satisfaction with the services. Nevertheless, specific satisfaction areas such as satisfaction with the therapeutic relationship and interaction, quality of care, technical aspects, or utility were also addressed.

Most samples included adult populations [42,44,46,49,50,52,55,57,58]. However, children or adolescents were also considered in other studies [41,45,48,53,54,56]. Moreover, some studies exclusively used data collected from children and adolescents [43,47,51]. The sample sizes ranged from 28 to 1,054,670 individuals, with 5 studies including less than 100 individuals, 8 including more than 100 individuals, and 7 including more than 1000 individuals. The proportion of female participants ranged from 15.8% (Department of Veterans Affairs [44]) to 90.5% (patients with an eating disorder [48]). The mean percentage of female participants in the included studies was approximately 55%.

Although none of the included studies used a theoretical model as a background for their analysis, the following sections are based on the UTAUT dimensions to allow for some theoretical context. This may guide future research in this area.

Patient Use

Overview

Key findings for the determinants of patient use of telemental health services are summarized in Table 6 (if reported, adjusted results are presented).

Table 6. Key findings of the included studies for determinants[a] of patient use.

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
Ainslie et al [41], 2022	• Female sex was negatively associated with going from low to either moderate or high telemedicine use (OR[b] 0.87, 95% CI 0.86-0.92). • Compared with patients aged ≥55 y, patients aged 0-12 y (OR 1.18, 95% CI 1.09-1.27) and 13-17 y (OR 1.16, 95% CI 1.09-1.25) had greater odds; patients aged 18-34 y (OR 0.74, 95% CI 0.70-0.79) and 35-54 y (OR 0.79, 95% CI 0.74-0.84) had lower odds of progressing from low to either moderate or high telemedicine use. • Living in an urban or rural area did not significantly change the probability for telemedicine use (*P*=.009).	• Except for bipolar disorder (OR 0.93, 95% CI 0.84-1.02), patients with diagnoses other than schizophrenia (reference) were negatively associated with progressing from low to either moderate or high use (major depression, OR 0.73, 95% CI 0.68-0.78; PTSD[c], OR 0.77, 95% CI 0.72-0.83; and anxiety or other disorders, OR 0.69, 95% CI 0.65-0.74).	__[d]	—	—
Ceniti et al [42], 2022	• There were no significant age differences (≥50 vs <50 y) regarding video vs telephone service use (detailed results, including numbers, were not reported).	—	—	—	—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
Chakawa et al [43], 2021	• Black children were less likely to have telemental health visits than White children compared with in-person visits (OR 0.35, 95% CI 0.16-0.76, *P*=.008). Hispanic (reference White: OR 0.45, 95% CI 0.17-1.19, *P*=.11) or other race or ethnicity (reference White: OR 0.57, 95% CI 0.21-1.53, *P*=.26) were not significantly associated with telemental health service use compared with in-person visits. • Telemental health service use was not significantly associated with sex (OR 0.64, 95% CI 0.33-1.23, *P*=.18), age (OR 1.22, 95% CI 0.62-2.51, *P*=.54), and language (OR 0.69, 95% CI 0.25-1.89, *P*=.47) compared with in-person visits.	• Children with internalizing problems were more likely to have telemental health visits than children with externalizing problems compared with in-person visits (OR 2.78, 95% CI 1.19-6.45, *P*=.02). Other primary referral concerns were not significantly associated with telemental health service use compared with in-person visits (reference; externalizing: OR 1.24, 95% CI 0.57-2.71, *P*=.59).	—	—	• Telemental health service use was not significantly associated with health insurance type compared with in-person visits (OR 1.68, 95% CI 0.74-3.82, *P*=.22).
Connolly et al [44], 2022			—	—	—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
	• Female sex was associated with having at least 1 video visit (OR 1.46, 95% CI 1.44-1.48, *P*<.01), having ≥50% of visits via video vs in-person (OR 1.64, 95% CI 1.60-1.68, *P*<.01), having ≥50% of visits via phone vs in-person (OR 1.17, 95% CI 1.15-1.19, *P*<.01), and having ≥50% of visits via video vs phone (OR 1.41, 95% CI 1.38-1.43, *P*<.01). • Age was negatively associated with having at least 1 video visit (ORs ranged from 0.27 to 0.92, *P*<.01), having ≥50% of visits via video vs in-person (ORs ranged from 0.23 to 0.91, *P*<.01), having ≥50% of visits via phone vs in-person (ORs ranged from 0.69 to 1.04, *P*<.01), and having ≥50% of visits via video vs phones (ORs ranged from 0.33 to 0.94, *P*<.01). • Race and ethnicity was associated with having at least 1 video visit (reference White and non-Hispanic: Black and non-Hispanic OR 0.97, 95% CI 0.96-0.98, *P*<.01; other race and non-Hispanic OR 1.21, 95% CI 1.18-1.25, *P*<.01; Hispanic OR 1.16, 95% CI 1.14-1.18, *P*<.01), having ≥50% of visits via video vs in-person (reference White and non-Hispanic: Black and non-Hispanic OR 0.86, 95% CI 0.84-0.88, *P*<.01; other race and non-Hispanic OR 1.18, 95% CI 1.13-1.23, *P*<.01; Hispanic OR 1.09, 95% CI 1.06-1.12, *P*<.01), having ≥50% of visits via phone vs in-person (reference White and non-Hispanic: Black and non-Hispanic, OR	• Schizophrenia diagnosis was negatively associated with having at least 1 video visit (OR 0.69, 95% CI 0.67-0.71, *P*<.01), having ≥50% of visits via video vs in-person (OR 0.36, 95% CI 0.34-0.37, *P*<.01), having ≥50% of visits via phone vs in-person (OR 0.64, 95% CI 0.62-0.65, *P*<.01), and having ≥50% of visits via video vs phone (OR 0.56, 95% CI 0.54-0.59, *P*<.01). • Depression diagnosis was associated with having at least 1 video visit (OR 1.06, 95% CI 1.05-1.07, *P*<.01), having ≥50% of visits via video vs in-person (OR 1.10, 95% CI 1.08-1.12, *P*<.01), and having ≥50% of visits via phone vs in-person (OR 1.10, 95% CI 1.09-1.12, *P*<.01). It was not significantly associated with having ≥50% of visits via video vs phone (OR 1.00, 95% CI 0.99-1.02, *P*>.05). • Anxiety disorder diagnosis was associated with having at least 1 video visit (OR 1.03, 95% CI 1.02-1.04, *P*<.01), having ≥50% of visits via video vs in-person (OR 1.03, 95% CI 1.02-1.05, *P*<.01), and having ≥50% of visits via phone vs in-person (OR 1.04, 95% CI 1.03-1.06, *P*<.01). It was not significantly associated with having ≥50% of visits via video vs phone (OR 1.00, 95% CI 0.98-1.01, *P*>.05). • Bipolar disorder diagnosis was negatively associated with having ≥50% of visits via video vs in-person (OR 0.89, 95% CI 0.86-0.91, *P*<.01) and having ≥50% of visits via video vs phone (OR 0.89, 95% CI 0.86-0.91, *P*<.01). It was not significantly associated with having at least 1 video			

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
	0.88, 95% CI 0.87-0.89, *P*<.01; other race and non-Hispanic, OR 0.93, 95% CI 0.90-0.96, *P*<.01; Hispanic, OR 0.95, 95% CI 0.93-0.97, *P*<.01), and having ≥50% of visits via video vs phone (reference White and non-Hispanic: Black and non-Hispanic, OR 0.97, 95% CI 0.96-0.99, *P*<.01; other race and non-Hispanic, OR 1.27, 95% CI 1.23-1.31, *P*<.01; Hispanic, OR 1.15, 95% CI 1.13-1.18, *P*<.01). • A low socioeconomic status (most disadvantaged tercile) was negatively associated with having at least 1 video visit (OR 0.68, 95% CI 0.67-0.69, *P*<.01), having ≥50% of visits via video vs in-person (OR 0.62, 95% CI 0.60-0.63, *P*<.01), having ≥50% of visits via phone vs in-person (OR 0.96, 95% CI 0.94-0.97, *P*<.01), and having ≥50% of visits via video vs phone (OR 0.64, 95% CI 0.63-0.65, *P*<.01).	visit (OR 1.00, 95% CI 0.98-1.02, *P*>.05) and having ≥50% of visits via phone vs in-person (OR 1.00, 95% CI 0.98-1.02, *P*>.05). • PTSD diagnosis was associated with having ≥50% of visits via video vs in-person (OR 1.11, 95% CI 1.09-1.13, *P*<.01), having ≥50% of visits via phone vs in-person (OR 1.16, 95% CI 1.15-1.18, *P*<.01), and negatively with having ≥50% of visits via video vs phone (OR 0.96, 95% CI 0.94-0.97, *P*<.01). It was not significantly associated with having at least 1 video visit (OR 1.01, 95% CI 0.98-1.01, *P*>.05). • Substance use disorder diagnosis was negatively associated with having ≥50% of visits via video vs in-person (OR 0.75, 95% CI 0.73-0.76, *P*<.01), having ≥50% of visits via phone vs in-person (OR 0.87, 95% CI 0.86-0.89, *P*<.01), and having ≥50% of visits via video vs phone (OR 0.86, 95% CI 0.84-0.87, *P*<.01). It was not significantly associated with having at least 1 video visit (OR 1.00, 95% CI 0.98-1.01, *P*>.05).			

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
	• Rurality was partly associated with having at least 1 video visit (reference urban: rural OR 1.14, 95% CI 1.12-1.16, *P*<.01; highly rural OR 1.22, 95% CI 1.14-1.31, *P*<.01), having ≥50% of visits via video vs in-person (reference urban: rural, OR 1.00, 95% CI 0.98-1.02, *P*>.05; highly rural, OR 1.24, 95% CI 1.13-1.36, *P*<.01), having ≥50% of visits via phone vs in-person (reference urban: rural, OR 1.14, 95% CI 1.12-1.16, *P*<.01; highly rural, OR 1.22, 95% CI 1.14-1.31, *P*<.01), and having ≥50% of visits via video vs phone (reference urban: rural, OR 0.88, 95% CI 0.86-0.89, *P*<.01; highly rural, OR 1.01, 95% CI 0.94-1.09, *P*>.05). • Not being married and being divorced, separated, or widowed compared with being married was negatively associated with having at least 1 video visit (ORs ranged from 0.92 to 0.93, *P*<.01), having ≥50% of visits via video vs in-person (ORs ranged from 0.82 to 0.83, *P*<.01), having ≥50% of visits via phone vs in-person (ORs ranged from 0.91 to 0.93, *P*<.01), and having ≥50% of visits via video vs phones (ORs ranged from 0.89 to 0.90, *P*<.01).	• Past mental health hospitalization was associated with having at least one video visit (OR 1.09, 95% CI 1.07-1.12, *P*<.01) and negatively associated with having ≥50% of visits via video vs in-person (OR 0.56, 95% CI 0.54-0.58, *P*<.01), having ≥50% of visits via phone vs in-person (OR 0.62, 95% CI 0.61-0.64, *P*<.01) and having ≥50% of visits via video vs phone (OR 0.88, 95% CI 0.86-0.91, *P*<.01). A disability rating of ≥50% was associated with having at least 1 video visit (OR 1.05, 95% CI 1.03-1.06, *P*<.01), having ≥50% of visits via video vs in-person (OR 1.07, 95% CI 1.05-1.10, *P*<.01), having ≥50% of visits via phone vs in-person (OR 1.02, 95% CI 1.00-1.04, *P*<.01), and having ≥50% of visits via video vs phone (OR 1.05, 95% CI 1.03-1.07, *P*<.01).			
Hutchison et al [47], 2022	—		—	—	—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
		• Increased baseline depression symptomology (r=−0.34, *P*<.05) and baseline anxiety symptomology (r=−0.32, *P*<.05) were associated with lower internet intervention use. Baseline somatic symptoms were not significantly correlated with internet intervention use (r=−0.26, *P*>.05). • No significant differences in attendance or retention rate were found for the moderate-risk and high-risk group (t=1.22, *P*=.23; t=0.20, *P*=.84).			
Lynch et al [50], 2021	• No significant associations with number of missed or cancelled sessions for age, gender, and race or ethnicity were found (detailed results, including numbers, were not reported).	• Having had at least 1 psychotic episode was associated with fewer missed or cancelled sessions (B=−0.49, *P*<.05).	—	• The mean no show or cancellation rate was 37% less during time 3 (post 2, week 13-18) compared with no show or cancellations while sessions were held in person (B=−0.47, *P*<.05).	—
Miu et al [53], 2021	• Older age was significantly associated with a smaller likelihood for conversion to teletherapy (B=−0.010, *P*=.01, OR 0.99, 95% CI 0.98-0.99). The SMI[e] status × age interaction was nonsignificant (B=0.021, *P*=.13, OR 1.02), meaning that the conversion for SMI and non-SMI groups did not depend on age. Nonsignificant predictors for conversion to teletherapy were sex (B=0.229, *P*=.13, OR 1.26, 95% CI 0.94-1.69) and ethnicity (reference non-Hispanic or Latino: Hispanic or Latino, B=−0.170, *P*=.44, OR 0.84, 95% CI 0.55-1.30; other, B=0.150, *P*=.56, OR 1.16, 95% CI 0.70-1.93).	• SMI status did not significantly predict conversion to telehealth (B=0.095, *P*=.63, OR 1.10, 95% CI 0.75-1.62). • The proportion of new patients starting teletherapy did not significantly differ by SMI status (χ21=1.2, *P*=.27). Patients with SMI had significantly higher numbers of telehealth visits compared with the non-SMI group (SMI: mean 1.47, SD 2.01; non-SMI: mean 1.04, SD 1.42; t251,154=−3.027, *P*=.003).	—	• Patients' previous engagement was not significantly associated with conversion to teletherapy (B=0.003, *P*=.41, OR 1.00, 95% CI 0.99-1.01). The SMI status × previous engagement interaction was nonsignificant (B=0.007, *P*=.43, OR 1.00), meaning that conversion for SMI and non-SMI groups did not depend on patients' previous engagement.	—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
Morgan et al [54], 2021	• Conversion to teletherapy was significantly associated with Hispanic ethnicity (χ21=6.7, *P*=.01, also in logistic regression model: B=2.425, *P*<.05, OR 11.30). • The association between conversion to teletherapy and the following demographic characteristics were not significant: age (t165=−1.474, *P*=.74), gender (χ21=2.1, *P*=.15), being a person of color (χ21=3.2, *P*=.07), poverty (χ21=3.0, *P*=.09), low educational attainment (χ21=0.1, *P*=.80), and household poverty status (χ21=1.2, *P*=.27). • Engagement in teletherapy was not significantly associated with ethnicity (B=1.15, β=.125, *P*>.05).	—	• Clients in individual therapy (individual vs relational case constellations) were more likely to convert to teletherapy (χ21=4.2, *P*=.04), also in logistic regression model (B=−1.38, *P*<.05, OR 0.25). • Engagement in teletherapy was associated with individual therapy (B=−2.34, *P*<.001, β=−.289).	• The number of sessions attended before the conversion to teletherapy was not significantly associated with the conversion to teletherapy (B=0.01, *P*>.05, OR 1.01) and engagement in teletherapy (B=0.02, *P*<.05, β=.179).	—
Severe et al [56], 2020	• Patient age was associated with the initial choice in visit type (*P*<.001). Patients aged ≥44 y were more likely than patients aged <44 y to choose telephone visits (RRR[f]=1.2; 95% CI 1.06-1.35). Sex (*P*=.99) and race (*P*=.06) were not significantly associated with the initial choice in visit type (delineating new and preexisting patients).	—	—	• The number of previous clinic visits was not significantly associated with the initial choice in visit type (delineating new and preexisting patients; *P*=.63).	• Health insurance type was not significantly associated with the initial choice in visit type (delineating new and preexisting patients; *P*=.08).
Sizer et al [57], 2022			• Discharge from clinic was negatively associated with the number of telehealth visits (IRR=0.55, *P*<.01). • No significant associations were found for referral source (self vs external source) and the number of telehealth visits (IRR=1.00, *P*>.10).	—	—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
	• Female sex was associated with an increased number of telehealth visits (reference male: IRR[g]=1.11, *P*<.05). • Age was negatively associated with the number of telehealth visits (reference>60 years: 18-30 y IRR=1.16, *P*<.10; 31-45 y IRR=1.22, *P*<.01; 46-60 y IRR=1.22, *P*<.01). • The number of school years was positively associated with the number of telehealth visits (IRR=1.01, *P*<.05). • No significant associations were found for race (IRR ranged from 0.74 to 0.99, *P*>.10) and monthly income (IRR=1.03, *P*>.10).	• The number of telehealth visits among patients with schizophrenia spectrum and other psychotic disorders decreased by 15% compared with patients with depressive disorders (IRR=0.85, *P*<.01). No significant results were found for other primary diagnosis types (IRR ranged from 0.903 to 0.959, *P*>.10). • The number of diagnosed mental illnesses was positively associated with the number of telehealth visits (IRR=1.07, *P*<.01). • The presence of other chronic health conditions was positively associated with the number of telehealth visits (IRR=1.10, *P*<.05).			
Ter Heide et al [58], 2021	• Refugee status was negatively associated with VCT[h] use (B=1.35, *P*<.01, OR 3.86, 95% CI 1.80-8.28).	• General psychopathology was negatively associated with VCT use (B=−0.58, *P*<.01, OR 0.56, 95% CI 0.39-0.56).	—	—	—
Tobin et al [59], 2023	• Older (OR 1.04, *P*<.001) and Black patients compared with White patients (OR 3.85, *P*<.05) were more likely to complete audio-only visits compared with video visits when only telehealth visits were offered (n=359). Gender was not significantly associated with telehealth visit type during that period (OR 1.04, *P*=.90). • No significant associations with demographic predictors were found (age, gender, and race; ORs ranged from 0.49 to 1.60, *P* value ranged from .07 to .25) when in-person and telehealth visits were offered (n=222).	—	—	—	• Patients with Medicare (OR 3.46, *P*<.001) and Medicaid (OR 3.43, *P*<.001) health insurance compared with private payers were more likely to complete audio-only visits than video visits when only telehealth visits were offered (n=359). • Health insurance type was not significantly associated with use of telehealth visits when in-person and telehealth visits were offered (n=222; reference private payer: Medicare, OR 2.01, *P*=.10; Medicaid, OR 1.07, *P*=.83).
					—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
Vakil et al [60], 2022	• Male sex was negatively associated with telehealth visit use (reference female: male, OR 0.76, 95% CI 0.64-0.91, *P*=.002; other, OR 0.43, 95% CI 0.13-1.45, *P*=.18). • Older age was positively associated with telehealth visit use (OR 1.01, 95% CI 1.00-1.01, *P*=.03). • Patients with income Q2[i] were more likely to use telehealth visits compared with the lowest income group Q1 (reference Q1: Q2, OR 1.32, 95% CI 1.01-1.74, *P*=.046). Other Qs did not significantly differ from the lowest income group Q1 concerning telehealth visit use (ORs ranged from 0.94 to 1.29, *P* values ranged from .10 to .69). • The distance between the individual's residence and the clinic was positively associated with telehealth visit use (OR 1.04, 95% CI 1.02-1.07, *P*=.001).	• Absence of suicidal behavior (reference none: ideation, OR 0.74, 95% CI 0.61-0.90, *P*=.003; planning, OR 0.55, 95% CI 0.38-0.79, *P*=.001; self-harm or attempt, OR 0.62, 95% CI 0.48-0.81, *P*<.001), substance use (reference none: OR 0.60, 95% CI 0.50-0.72, *P*<.001), psychotic symptoms (reference absent: OR 0.41, 95% CI 0.30-0.56, *P*<.001) and cognitive impairment (reference absent: OR 0.53, 95% CI 0.34-0.84, *P*=.007) were associated with telehealth visit use • Presence of personality problems (OR 1.13, 95% CI 0.92-1.40, *P*=.26), depressive or anxiety problems (OR 1.26, 95% CI 0.98-1.63, *P*=.07), bipolar spectrum disorders (OR 0.94, 95% CI 0.65-1.34, *P*=.30), and other mental illnesses (OR 0.82, 95% CI 0.57-1.19, *P*=.30) were not significantly associated with telehealth visit use.	• Each pandemic period after the first lockdown (reference lockdown 1: in-between period, OR 0.37, 95% CI 0.23-0.49, *P*<.001; lockdown 2, OR 0.39, 95% CI 0.30-0.52, *P*<.001; after lockdown 2, OR 0.35, 95% CI 0.26-0.49, *P*<.001), overnight visits (reference daytime visits: OR 0.48, 95% CI 0.34-0.67, *P*<.001), and weekend visits (reference weekday visits: OR 0.75, 95% CI 0.61-0.91, *P*=.004) were negatively associated with telehealth visit use.	• Patients with a prior visit in the last year were less likely to use telehealth visits (OR 0.75, 95% CI 0.61-0.91, *P*=.004).	

[a]Psychosocial influence, effort, and performance expectancy were not included as categories in this table because none of the included studies observed the relationship of these determinants with patient use.

[b]OR: odds ratio.

[c]PTSD: posttraumatic stress disorder.

[d]No information present in the study regarding this category of determinants.

[e]SMI: serious mental illness.

[f]RRR: relative risk reduction.

[g]IRR: incidence rate ratio.

[h]VCT: clinical videoconferencing.

[i]Q: income quintile (Q1: lowest and Q5: highest).

Sociodemographic Factors

In total, 11 studies examined the relationship between sex and patient use of telemental health services. Approximately half of these studies (n=6) did not find significant sex differences in use [43,50,53,54,56,59]. Nevertheless, 4 studies reported higher use rates in female participants [44,57,58,60]. In contrast, 1 study reported lower odds for female participants to go from low use rates (before the pandemic) to moderate or high use rates during the pandemic [41].

A total of 13 studies examined the relationship between age and patient use of telemental health services. Nearly half of these studies (n=6) found a nonsignificant association of age with patient use [42,43,50,54,58,59]. In contrast, 1 study found that older age was positively associated with telemental health service use [60] and 3 studies found that older patients were more likely to use audio-only formats (eg, telephone services) compared with video formats [44,56,59]. Nevertheless, 3 studies observed a negative association of age with telemental health service use [44,53,57]. Ainslie et al [41] reported mixed findings. In their sample, participants aged 0 to 17 years were more likely than those aged ≥55 years to go from having <25%

of mental health services in a remote format (low use) to having 25% to 75% (moderate use) or >75% (high use) of use. However, participants aged 18 to 54 years were less likely than those aged ≥55 years to go from low to moderate or high use.

In total, 8 studies examined the relationship between race or ethnicity and patient use of telemental health services. Of these, 5 studies did not find a significant association [50,53,56,57,59]. However, Tobin et al [59] reported that Black individuals were more likely to use audio-only services, which was also found in the study by Connolly et al [44]. In addition, 2 studies found that Black patients were less likely to use telemental health services and used them less frequently compared with White patients [43,44]. Connolly et al [44] found that other than Black races and Hispanic ethnicity compared with the White race, non-Hispanic race or ethnicity is positively associated with telemental health service use and frequency of video service use (but negatively associated with frequency of phone service use). Although being a person of color was a nonsignificant determinant for the conversion to teletherapy, a relationship between Hispanic ethnicity and the conversion was found in the sample of Morgan et al [54]. However, when examining engagement with teletherapy, no significant association with ethnicity was observed in their sample.

A total of 3 studies examined the relationship of area lived in and patient use of telemental health services. Findings suggested a positive association with rurality: 1 study found that individuals from (highly) rural areas were more likely to use telemental health services [44] and 1 study stated that telehealth users lived further away from the clinic [60]; however, 1 study found no significant association [41].

Other sociodemographic determinants of patient use were considered in very few studies. A low socioeconomic and financial status was associated with lower use in 2 studies [44,60] but failed to significantly predict telemental health service use in 2 other studies [54,57]. Years of schooling were positively associated with the number of visits in the sample by Sizer et al [57]; however, Morgan et al [54] did not find a significant association between educational attainment and opting out of teletherapy after clinical conversion from in-person therapy to teletherapy. In addition, being married was positively associated with telemental health service use and use frequency in 1 study [44]. Language was not significantly associated with use, and refugee status was associated with lower odds of telemental health use in single studies [43,58].

Health Factors

A total of 9 studies examined the relationship of psychological symptom severity or diagnosis and patient use of telemental health services. Most of these studies (n=5) found that individuals with higher symptom severity (eg, patients with schizophrenia) had lower use rates [44,47,57,58,60]. However, the number of diagnoses, depression, anxiety or posttraumatic stress disorder diagnosis, past psychotic episodes, and serious mental illness status were each associated with a higher use frequency or fewer missed sessions in single studies [44,50,53,57]. Similarly, Ainslie et al [41] reported that individuals with schizophrenia were more likely to go from low to moderate or high use than individuals with other diagnoses. Nevertheless, the risk status for adverse mental and behavioral outcomes and serious mental illness status were not significantly associated with use and visit intensity in single studies [47,53]. In addition, Chakawa et al [43] found that children with internalizing problems were more likely to have a telemental health visit than children with externalizing problems.

Furthermore, the presence of chronic health conditions was associated with a higher number of visits in the sample studied by Sizer et al [57]. A disability rating of ≥50% in US veterans was positively associated with telemental health service use and frequency of use in 1 study [44].

Service Factors

A total of 3 studies examined the relationship between service factors and patient use of telemental health services. Morgan et al [54] found that patients undergoing individual therapy were more likely to convert to telemental health services. Referral source (self vs external sources) was not significantly associated with use rates [57]. Regarding service times, Vakil et al [60] stated that telehealth visits were significantly less likely during each pandemic period after the first lockdown, for nighttime visits (compared with daytime visits) and weekend visits (compared with weekday visits).

Experience

A total of 5 studies examined the relationship between experience with telemental health services and patient use of telemental health services. Previous engagement in mental health services was found to be negatively associated with telehealth visit use in the sample studied by Vakil et al [60] but failed to predict use in 2 other studies [53,56]. Although the number of sessions attended before teletherapy was not significantly associated with conversion to teletherapy in the analysis by Morgan et al [54], it was found to significantly predict the number of telemental health visits in this sample. Moreover, Lynch et al [50] reported that longer duration of participation in telemental health services was associated with fewer missed sessions.

Facilitating Conditions

A total of 3 studies examined the relationship between facilitating conditions and patient use of telemental health services. Health insurance type was not significantly associated with patient use in these studies [43,56,59]. Nevertheless, Tobin et al [59] reported that Medicare- or Medicaid-insured individuals used audio-only formats more often than private payers.

Psychosocial Influence, Effort and Performance Expectancy

None of the included studies examined the relationship between psychosocial factors, effort or performance expectancy and patient use of telemental health services.

Patient Satisfaction

Overview

Key findings for the determinants of patient satisfaction with telemental health services are summarized in Table 7 (if reported, adjusted results are presented).

Table 7. Key findings of the included studies for determinants[a] of patient satisfaction.

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Psychosocial influence (what do families and peers think about program or psychosocial impact)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
Ceniti et al [42], 2022	• Satisfaction was not significantly associated with age (≥50 y vs <50 y) and number of people living in the household (detailed results, including numbers, were not reported). • Living with others was significantly associated with satisfaction (χ21=5.8, *P*=.02). • Satisfaction was greater in users from Ontario compared with those from other Canadian provinces (χ21=3.9, *P*=.047).	• Satisfaction was not significantly associated with high-risk status for COVID-19 (detailed results, including numbers, were not reported).	• Video services (compared with telephone) were associated with greater satisfaction (User-MD[b] χ21=6.1, *P*=.01; User-HCP[c] χ21=6.6, *P*=.01). • No significant differences between user-groups (psychiatrists or family physicians vs other mental health care providers) in overall satisfaction were found (detailed results, including numbers, were not reported).	—[d]	• Level of connectedness with loved ones was positively correlated with overall remote care satisfaction (*r*=.197, *P*=.007) and satisfaction with therapeutic rapport (*r*=.155, *P*=.03).	• Satisfaction was not significantly associated with frequency of internet use (detailed results, including numbers, were not reported).
Guinart et al [45], 2020	• Significant age differences for telephone services were found (χ224=46.3, *P*=.004). A lower proportion of patients aged 55-64 y described their experience as excellent compared with other age groups (χ24=12.8, *P*=.01). A higher proportion of patients aged 45-54 y rated their experience as poor compared with other age groups (χ24=10.5, *P*=.03).	—	—	• Patients under care for <1 y endorsed missing the clinic and feeling connected to it less frequently than other groups (χ^2_6=21.5, *P*=.002).	—	—
Haxhihamza et al [46], 2021	• Satisfaction was not significantly associated with gender, age, and place of living (detailed results, including numbers, were not reported).	—	—	—	—	—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Psychosocial influence (what do families and peers think about program or psychosocial impact)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
Hutchison et al [47], 2022	—	• None of the baseline psychological symptoms were correlated with treatment satisfaction (r values ranged between −0.13 and 0.01, *P*>.05). • Adolescents in the moderate-risk group reported significantly higher satisfaction with the intervention than those in the high-risk group (t=2.03, *P*<.05, Cohen d=0.60).	—	—	—	—
Lewis et al [48], 2021	• No significant correlations with the views toward the transition to web-based therapy for age (r=.036, *P*=.78), gender (r=.006, *P*=.96), and education (r=.092, *P*=.47) were found. • No significant correlations with the TSQ[e] for age (similarity scale: r=.182, *P*=.15; quality scale: r=−.047, *P*=.72), gender (similarity scale: r=.067, *P*=.60; quality scale: r=.146, *P*=.25), and education (similarity scale: r=.093, *P*=.47; quality scale: r=−.017, *P*=.89) were found.		—	• Treatment duration correlated with positive views toward the transition to online therapy (r=0.291, *P*=.02). • Treatment duration was not significantly correlated with the TSQ (similarity scale: r=0.124, *P*=.34; quality scale: r=−0.144, *P*=.26).	• The fear of COVID-19 scale-19S score correlated with positive views toward the transition to web-based therapy for (r=0.276, *P*=.03). The fear of COVID-19 scale-19S score was not significantly associated with the TSQ (similarity scale: r=−0.193, *P*=.13; quality scale: r=−0.143, *P*=.26).	—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Psychosocial influence (what do families and peers think about program or psychosocial impact)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
		• No significant correlations with the views toward the transition to web-based therapy for past eating disorder hospitalization (t=0.152, *P*=.24), EDE-Q[f] scales (r values ranged from −0.168 to −0.094, *P* values ranged from .19 to .47) and Depression, Anxiety and Stress Scales-21 scales (r values ranged from −0.162 to −0.080, *P* values ranged from .21 to .53) were observed. • No significant associations of the TSQ with past eating disorder hospitalization (similarity scale: t=0.149, *P*=.24; quality scale: t=0.061, *P*=.63), EDE-Q scales (r values ranged from −0.100 to 0.101, *P* values ranged from .43 to .77) and Depression, Anxiety and Stress Scales-21 scales (r values ranged from −0.121 to 0.094, *P* values ranged from .34 to .84) were found. TSQ scores did not significantly differ between eating disorder diagnoses (detailed results, including numbers, were not reported). • No significant correlation of the views toward the transition to web-based therapy with BMI (r=0.226, *P*=.08) were found. BMI was not significantly correlated with the TSQ (similarity scale: r=0.221, *P*=.09; quality scale: r=−0.011, *P*=.93).				

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Psychosocial influence (what do families and peers think about program or psychosocial impact)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
Lohmiller et al [49], 2021	• No significant associations of age (F1,277=0.18, *P*=.67) and gender (detailed results, including numbers, were not reported) with the overall "assessment of therapeutic contact" were found. • However, a significant association of age with the item "assessment of therapeutic contact as personal" was found (F1,277=4.50, *P*=.04) indicating that older individuals perceived the video format as partly more impersonal. • When looking at "hurdles" age was significantly associated with single items: "The necessary technology/framework conditions overwhelmed me" (F1,277=7.85, *P*=.005) indicating that older individuals perceived the video format as more challenging and "I was able to fully concentrate on the content of the conversation" (F1,277=14.85, *P*<.001) indicating that older individuals perceived the video format to be more impersonal and depersonalized.	—	• Significant differences were found in the items: global judgment conversation contact (F2,275=3.39, *P*=.04), pleasantness (F2,275=3.35, *P*=.04), friendliness (F2,275=5.55, *P*=.004), and feeling comfortable (F2, 275=8.49, *P*<.001), all favoring video consultations compared with phone and office consultation. The other items of the "assessment of therapeutic contact" showed no significant differences (P≥.05). • The assessment of the "therapeutic relationship" did not significantly differ between groups, except for the item "I have recently started to feel better" (F2,275=4.97, *P*=.008), favoring phone and video contacts.	—	—	—
Meininger et al [51], 2022	—	• There were no significant correlations between parent-rated treatment satisfaction and the severity of patients' symptoms, stress, and psychosocial functioning (detailed results, including numbers, were not reported).	—	• Treatment duration correlated positively with parent-rated treatment satisfaction (mean satisfaction score: r=.20, *P*<.02).	—	—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Psychosocial influence (what do families and peers think about program or psychosocial impact)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
Michaels et al [52], 2022	• No sex-based differences in the preferred telehealth method (P=.67), experiences using telephone (P=.92) or video (P=.58), whether patients would use telehealth in the future (P=.11) and was perceived as helpful as in-person treatment (P=.38) were found. • No gender-based differences in the preferred telehealth method (P=.64), experiences using telephone (P=.63) or video (P=.53), whether patients would use telehealth in the future (P=.52) and was perceived as helpful as in-person treatment (P=.13) were found. • No race-based differences in the preferred telehealth method (P=.21), experiences using telephone (P=.29) or video (P=.99) and whether patients would use telehealth in the future (P=.15) were found.	—	• No between-group differences in preferences for telehealth methods were found (H 1=0.46, P=.49). Most of the college therapy and medication group (63/78, 81%) and college medication–only group (20/23, 87%) reported a strong preference for the video format.	—	—	—
Nesset et al [55], 2023	• No significant sex differences were found for Client Satisfaction Questionnaire-8 scores (P values ranged from .10 to .57 for the different items), except that female participants were more content with the length of the therapy (item 5: mean 3.86, SD 0.378 vs male, mean 2.90, SD 1.01, P=.03).	—	—	—	—	—
Ter Heide et al [58], 2021				—		—

Study, year	Sociodemographic factors (eg, sex, age, race, education, and area lived in)	Health factors (eg, diagnosis, symptoms, and symptom severity)	Service factors (eg, video or telephone and duration of treatment)	Experience (with mental health services)	Psychosocial influence (what do families and peers think about program or psychosocial impact)	Facilitating conditions (eg, electronic devices, internet connection, and insurance)
	• There was a significant main effect of gender, with female participants reporting significantly higher VCT[g] satisfaction than male participants (F1,196=10.60, *P*<.01). • No significant associations with VCT satisfaction were found for age, refugee status and level of education (detailed results, including numbers, were not reported).	• There was a significant main effect of general psychopathology, with general psychopathology being negatively associated with VCT satisfaction (F1,196=6.61, *P*<.05). Among those who reported using VCT, a small, negative correlation between VCT satisfaction and general psychopathology was found (r=−0.18, *P*<.01, n=221).	• There was no significant difference in treatment satisfaction between the VCT group and the non-VCT group (t276=−0.237, *P*=.81, n=278).		• Among those who reported using VCT, a small, negative correlation between satisfaction and coronavirus stress level was found (r=−0.21, *P*<.01, n=228). A small, positive correlation between VCT satisfaction and life satisfaction was found (r=0.27, *P*<.001, n=228).	

[a]Effort and performance expectancy were not included as categories in this table because none of the included studies observed a relationship between these determinants and patient satisfaction.

[b]User-MD: mental health care users who saw an MD provider (psychiatrist or family physician).

[c]User-HCP: mental health care users who saw another mental health care provider (eg, psychotherapist).

[d]No information present in the study regarding this category of determinants.

[e]TSQ: telemedicine satisfaction questionnaire.

[f]EDE-Q: Eating Disorder Examination Questionnaire.

[g]VCT: clinical videoconferencing.

Sociodemographic Factors

A total of 5 studies examined the relationship between sex and patient satisfaction with telemental health services, and all of them did not find a significant association of sex with the satisfaction scores [46,48,49,52,55].

A total of 6 studies examined the relationship between age and patient satisfaction with telemental health services. Most studies (n=4) did not find a significant association between age and satisfaction [42,46,48,58]. Lohmiller et al [49] also did not find a significant association between age and the overall satisfaction with therapeutic contact. However, older age was associated with lower satisfaction for some items, meaning that older individuals perceived the video intervention as less personal and more challenging and found it harder to fully concentrate on the content of the conversation. Guinart et al [45] found lower satisfaction ratings for telephone services among older patients.

One study observed a nonsignificant relationship between race and patient satisfaction with telemental health services [52].

In total, 2 studies examined the relationship between area lived in and patient satisfaction with telemental health services. While Haxhihamza et al [46] did not find a significant association, Ceniti et al [42] reported greater satisfaction ratings in users from Ontario compared with those in other Canadian provinces.

Other sociodemographic determinants of patient satisfaction were considered in some studies. Educational level was observed in 2 studies and was not significantly associated with satisfaction in these samples [48,58]. In addition, Ter Heide et al [58] reported that refugee status is not significantly associated with satisfaction. Moreover, Ceniti et al [42] included living situation of participants as a potential determinant. While the number of people living in the household was not significantly associated with remote care satisfaction, living with others showed a significant association with this outcome.

Health Factors

A total of 4 studies examined the relationship between psychological symptom severity and patient satisfaction with telemental health services. Only 1 study found a significant association between symptom severity and satisfaction. In the sample studied by Hutchison et al [47], patients at moderate risk were more satisfied than patients who were at high risk for adverse mental and behavioral outcomes. However, the other 3 studies did not observe significant relationships [42,48,51].

A total of 2 studies examined the relationship between physical health and patient satisfaction with telemental health services. Nonsignificant relationships were found between BMI and high-risk status for COVID-19, with satisfaction in single studies [42,48].

Service Factors

A total of 7 studies examined the relationship between service factors and patient satisfaction with telemental health services. Of these, 3 studies reported that telemental health services delivered via video services were associated with higher patient satisfaction than those delivered via telephone services [42,49,52]. However, Ter Heide et al [58] could not find this relationship. Furthermore, the therapeutic alliance bond was associated with higher satisfaction ratings in 1 study [48]. The provider type (psychiatrists or family physicians vs other mental health care providers) was not significantly associated with patient satisfaction in the study by Ceniti et al [42].

Experience

A total of 3 studies examined the relationship between experience with telemental health services and patient satisfaction with telemental health services. In the study by Lewis et al [48], longer treatment duration was associated with higher satisfaction, while Guinart et al [45] observed that patients who were under care for less than a year perceived the transition to telemental health services as less negative (missed the clinic less and did not feel less connected). Moreover, the number of telemental health sessions was associated with higher satisfaction ratings in 1 study [51].

Psychosocial Influence

A total of 3 studies examined the relationship between psychosocial factors and patient satisfaction with telemental health services. Level of connectedness with loved ones and life satisfaction were associated with greater patient satisfaction [42,58]. Moreover, COVID-19–related aspects were considered in single studies. The COVID-19 stress level had a small negative correlation with satisfaction [58], and fear of COVID-19 was associated with positive views toward the transition to teletherapy but was not significantly associated with overall satisfaction scores [48].

Facilitating Conditions

One study examined the relationship between facilitating conditions and patient satisfaction with telemental health services. Ceniti et al [42] reported that the frequency of internet use was not significantly associated with patient satisfaction.

Effort and Performance Expectancy

None of the included studies examined the relationship between effort or performance expectancy and patient satisfaction with telemental health services.

Discussion

Principal Findings

Overview

This systematic review aimed to provide an extensive overview of the literature on and highlight the influential determinants of patient use and satisfaction with synchronous telemental health services during the COVID-19 pandemic. Various determinants of patient use and satisfaction were considered. Sociodemographic characteristics were most frequently examined. Nevertheless, health- and service-related determinants also received considerable attention. Major dimensions of the UTAUT, such as effort and performance expectancy, were neglected in recent studies. Although most associations were mixed or nonsignificant, some indications for potential relationships were found (eg, for sex, age, and symptom severity). This systematic review is the first to examine the determinants of patient use and satisfaction with synchronous telemental health services during the pandemic, thus markedly extending our current knowledge.

Sociodemographic Factors

Regarding sociodemographic factors, a variety of determinants were observed in the included studies. Most studies found that sex was not significantly associated with patient use and satisfaction. However, some studies with large samples found that female participants were more likely to use telemental health services. This suggests that previous findings regarding greater use of mental health services among female participants may also apply to the field of telemental health [66-69]. Moreover, this could explain the finding that women were less likely to go from low to either moderate or high telemedicine use [41], as they already had higher use rates before the occurrence of the pandemic.

When looking at patient age, mostly nonsignificant associations with the outcomes were found. Nevertheless, some large-sample studies found that older age was negatively associated with the outcomes and that older patients were more likely to use audio-only services compared with video services. This could be not only because of the lower likelihood of older adults using mental health care services [70] but also because of the digital divide in mobile health [71]. However, audio-only formats seem to be a promising alternative to video consultations for older adults, which was also found in other telemedicine areas during the pandemic (eg, academic medical center outpatient visits and oncological care) [72-74].

Race, ethnicity, area lived in (ie, rurality and province lived in), education, and other determinants (eg, refugee status, financial status, and living situation) were observed in only few studies and led to mainly nonsignificant or mixed associations with the outcomes. More research regarding these sociodemographic determinants is needed in the future. In summary, sociodemographic factors tend to play a role in patient use of telemental health services. In particular, sex and age appear to be potential determinants that were frequently observed. For patient satisfaction, mainly nonsignificant or mixed findings were reported.

Health Factors

Regarding health factors, symptom severity was observed in some studies and was mostly associated with lower use rates in patients with mental health conditions. This is in contrast to in-person mental health services research, where symptom severity was associated with an increased likelihood of seeking treatment [69]. A potential reason for this could be that patients with very severe symptoms were preferably kept in an in-person setting despite the pandemic to assure appropriate treatment. However, findings on engagement or attendance were mixed, with some studies suggesting that more severe symptoms were

associated with an increased frequency of telemental health visits. This could mean that individuals with more severe symptoms were less likely to start teletherapy, but once they were participating in telemental health services, they used it more frequently than patients with less severe symptoms. For satisfaction, most of the associations were nonsignificant. In conclusion, the associations with determinants were mostly observed for patient use. Although psychological symptom severity seemed to be negatively associated with the likelihood of telemental health service use, some indications for a positive association with use frequency were observed.

Service Factors

With regard to service factors, various determinants were observed in different studies. For patient use, there was great heterogeneity in the observed aspects. Therefore, it is challenging to compare the results of these studies. More research in this field is clearly needed. Nevertheless, services that were delivered in video format seemed to be associated with higher patient satisfaction than services delivered via telephone. A qualitative study in primary care highlighted potential reasons for the preference of video services, including nonverbal cues and reassurance, lower risk of miscommunication, more personal experience, and increased focus [75]. A recent systematic review on using telephone and video services for mental health treatment also emphasized the strengths of the video format [76]. However, they also stated that the telephone format can be superior to the video format in some cases (eg, fewer technological challenges [76]).

Experience

With regard to the experience with telemental health services, previous engagement in mental health services was not significantly associated with patient use. This could potentially mean that telemental health use rather depends on need factors than on experience. Regarding patient satisfaction, findings for the treatment duration were mixed. However, the number of telehealth sessions attended seemed to be associated with fewer missed sessions and higher satisfaction ratings. Therefore, patients might have got used to the new situation over time and had adapted to the remote format.

Psychosocial Influence

With regard to psychosocial factors, no determinants of patient use were observed. For patient satisfaction, significant determinants were only observed in single studies. Further research, including on psychosocial determinants, is urgently required. Especially factors such as personality (eg, neuroticism or conscientiousness) and social determinants (eg, loneliness) could be of interest for the future of telemental health, considering their impact on health care use [77,78].

Facilitating Conditions

With regard to facilitating conditions, the health insurance type was not significantly associated with patient use in some studies. The frequency of internet use was also not significantly associated with patient satisfaction in 1 single study. More research is needed in this area to identify potential facilitators of telemental health use and satisfaction.

Effort and Performance Expectancy

With regard to effort and performance expectancy, no study included determinants from these constructs. Considering that these 2 dimensions are key elements of the UTAUT, future research should urgently include determinants from this area.

Study Quality

Overall, the quality of the included studies was mainly good or fair and did not vary substantially between the different studies. Most studies included large samples and some included even very large electronic medical record data sets [41,44]. However, the generalizability of our results is limited considering that the evidence mainly came from North America and Western countries and because of differences in psychiatric care and telemental health services. Most studies did not provide participation rates, sample size justification, power description, or variance and effect estimates, which are important information sources for the interpretation of the associations and the detection of potential biases (eg, selection bias).

Future Research

Considering the findings of our systematic review, multiple research gaps were identified. In general, the inclusion of theoretical models is needed in future studies to set a more consistent focus on important determinants and to assure comparability of the studies. Future research should consider different types of use behavior (eg, frequency of use, adoption, and attendance) and satisfaction (different scales or areas). Established scales should be used to measure the outcomes rather than single items (especially for satisfaction) because single items are more prone to bias. Moreover, to improve the understanding of the relationships between the different determinants and their effects on patient use and satisfaction, future studies that examine the influencing chain and process behind the outcomes are needed. In addition, future studies should explore whether certain telemental health formats (eg, telephone, video, or asynchronous formats) are especially suited for the treatment of specific diagnoses (eg, depression, anxiety, or schizophrenia). Furthermore, longitudinal studies are needed to verify the findings and test for potential changes over time. Longitudinal studies are also of interest to see whether findings regarding use and satisfaction during the pandemic also apply to postpandemic circumstances. For instance, a recent qualitative study found that remote services were only seen as a good alternative to in-person mental health services during extreme circumstances [79]. Additional qualitative research is needed, for example, to explore the barriers of users who do not indicate high use or satisfaction rates to make telemental health services more accessible and user friendly in the future.

With regard to the UTAUT dimensions, major research gaps were revealed. In particular, for the dimensions effort and performance expectancy, psychosocial influence and facilitating conditions research is missing in the respective literature. However, these dimensions could be valuable starting points for interventions, as they could potentially be influenced or adapted over time to improve use rates and satisfaction with telemental health services.

Strengths and Limitations

Our systematic review was registered in PROSPERO and conducted in accordance with PRISMA guidelines to ensure the quality and transparency of the manuscript. A double-screening approach was used to screen 3 databases, which generally was found to be advanced in comparison with single screening and lead to fewer missed studies in the screening process [80]. In addition, data extraction and study quality assessment were performed by 2 reviewers. Furthermore, this review is the first to evaluate the existing literature on the determinants of use and satisfaction with synchronous telemental health services during the COVID-19 pandemic.

However, this study has some limitations. Only peer-reviewed quantitative studies were included. Therefore, potentially meaningful studies were not considered (eg, from the gray literature). Nevertheless, this step promoted the quality of the included studies and the comparability of the findings. In addition, only German and English language articles were screened, whereby relevant articles in other languages could have been missed. Finally, no meta-analysis was performed because of the high heterogeneity in study designs, outcomes, and effect measures.

Conclusions

The extensive implementation of synchronous telemental health services during the pandemic triggered new research in this field. This systematic review was the first to synthesize studies that observed the determinants of patient use and satisfaction with these services. Significant heterogeneity was observed among the included studies. The findings revealed potential target groups (eg, female and young patients with mild symptoms) for future postpandemic telemental health interventions. However, the findings also revealed that patient groups that were especially burdened during the pandemic (such as older patients with severe symptoms) were harder to reach, and efforts are required to address such groups. Finally, knowledge gaps in the recent literature were highlighted, which call for future quantitative and qualitative research to secure and expand the recent findings. This could help to better understand barriers as well as individual preferences and eventually improve telemental health services in the future.

Acknowledgments

The authors acknowledge financial support from the Open Access Publication Fund of UKE (Universitätsklinikum Hamburg-Eppendorf) and DFG (German Research Foundation).

Authors' Contributions

AN, AH, and H-HK developed the concept and search strategy for this systematic review. AH supervised the study. Study selection, data extraction, and quality assessment were performed by AN and JB; AH was consulted in case of any disagreement in these processes. The manuscript was written by AN and critically revised by AH and H-HK. Text and tables were formatted by AN and JB. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of all included studies in the narrative synthesis.
[DOCX File , 24 KB - mental_v10i1e46148_app1.docx]

References

1. Telemedicine: opportunities and developments in member states: report on the second global survey on eHealth. World Health Organization. 2010. URL: https://apps.who.int/iris/handle/10665/44497 [accessed 2022-10-14]
2. Hailey D, Roine R, Ohinmaa A. Systematic review of evidence for the benefits of telemedicine. J Telemed Telecare 2002;8 Suppl 1:1-30. [doi: 10.1258/1357633021937604] [Medline: 12020415]
3. Hjelm NM. Benefits and drawbacks of telemedicine. J Telemed Telecare 2005;11(2):60-70. [doi: 10.1258/1357633053499886] [Medline: 15829049]
4. Goharinejad S, Hajesmaeel-Gohari S, Jannati N, Goharinejad S, Bahaadinbeigy K. Review of systematic reviews in the field of telemedicine. Med J Islam Repub Iran 2021 Dec 29;35:184 [FREE Full text] [doi: 10.47176/mjiri.35.184] [Medline: 36042824]
5. Saliba V, Legido-Quigley H, Hallik R, Aaviksoo A, Car J, McKee M. Telemedicine across borders: a systematic review of factors that hinder or support implementation. Int J Med Inform 2012 Dec;81(12):793-809. [doi: 10.1016/j.ijmedinf.2012.08.003] [Medline: 22975018]
6. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
7. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of telehealth during the COVID-19 pandemic: scoping review. J Med Internet Res 2020 Dec 01;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]

8. Hincapié MA, Gallego JC, Gempeler A, Piñeros JA, Nasner D, Escobar MF. Implementation and usefulness of telemedicine during the COVID-19 pandemic: a scoping review. J Prim Care Community Health 2020 Jan;11:2150132720980612 [FREE Full text] [doi: 10.1177/2150132720980612] [Medline: 33300414]
9. Harju A, Neufeld J. Telehealth utilization during the COVID-19 pandemic: a preliminary selective review. Telemed Rep 2022 Feb 03;3(1):38-47 [FREE Full text] [doi: 10.1089/tmr.2021.0040] [Medline: 35720447]
10. Weiner JP, Bandeian S, Hatef E, Lans D, Liu A, Lemke KW. In-person and telehealth ambulatory contacts and costs in a large US insured cohort before and during the COVID-19 pandemic. JAMA Netw Open 2021 Mar 01;4(3):e212618 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2618] [Medline: 33755167]
11. Ciciurkaite G, Marquez-Velarde G, Brown RL. Stressors associated with the COVID-19 pandemic, disability, and mental health: Considerations from the Intermountain West. Stress Health 2022 Apr;38(2):304-317 [FREE Full text] [doi: 10.1002/smi.3091] [Medline: 34382736]
12. Liu X, Zhu M, Zhang R, Zhang J, Zhang C, Liu P, et al. Public mental health problems during COVID-19 pandemic: a large-scale meta-analysis of the evidence. Transl Psychiatry 2021 Jul 09;11(1):384 [FREE Full text] [doi: 10.1038/s41398-021-01501-9] [Medline: 34244469]
13. Wu T, Jia X, Shi H, Niu J, Yin X, Xie J, et al. Prevalence of mental health problems during the COVID-19 pandemic: a systematic review and meta-analysis. J Affect Disord 2021 Feb 15;281:91-98 [FREE Full text] [doi: 10.1016/j.jad.2020.11.117] [Medline: 33310451]
14. Murphy L, Markey K, O' Donnell C, Moloney M, Doody O. The impact of the COVID-19 pandemic and its related restrictions on people with pre-existent mental health conditions: a scoping review. Arch Psychiatr Nurs 2021 Aug;35(4):375-394 [FREE Full text] [doi: 10.1016/j.apnu.2021.05.002] [Medline: 34176579]
15. Villarreal-Zegarra D, Alarcon-Ruiz CA, Melendez-Torres GJ, Torres-Puente R, Navarro-Flores A, Cavero V, et al. Development of a framework for the implementation of synchronous digital mental health: realist synthesis of systematic reviews. JMIR Ment Health 2022 Mar 29;9(3):e34760 [FREE Full text] [doi: 10.2196/34760] [Medline: 35348469]
16. What is telemental health? National Institute of Mental Health. URL: https://www.nimh.nih.gov/health/publications/what-is-telemental-health#:~:text=Telemental%20health%20is%20the%20use,to%20as%20telepsychiatry%20or%20telepsychology [accessed 2023-04-10]
17. Barnett P, Goulding L, Casetta C, Jordan H, Sheridan-Rains L, Steare T, et al. Implementation of telemental health services before COVID-19: rapid umbrella review of systematic reviews. J Med Internet Res 2021 Jul 20;23(7):e26492 [FREE Full text] [doi: 10.2196/26492] [Medline: 34061758]
18. Li H, Glecia A, Kent-Wilkinson A, Leidl D, Kleib M, Risling T. Transition of mental health service delivery to telepsychiatry in response to COVID-19: a literature review. Psychiatr Q 2022 Mar;93(1):181-197 [FREE Full text] [doi: 10.1007/s11126-021-09926-7] [Medline: 34101075]
19. Hao X, Qin Y, Lv M, Zhao X, Wu S, Li K. Effectiveness of telehealth interventions on psychological outcomes and quality of life in community adults during the COVID-19 pandemic: a systematic review and meta-analysis. Int J Ment Health Nurs (Forthcoming) 2023 Feb 20. [doi: 10.1111/inm.13126] [Medline: 36808863]
20. Langarizadeh M, Tabatabaei MS, Tavakol K, Naghipour M, Rostami A, Moghbeli F. Telemental health care, an effective alternative to conventional mental care: a systematic review. Acta Inform Med 2017 Dec;25(4):240-246 [FREE Full text] [doi: 10.5455/aim.2017.25.240-246] [Medline: 29284913]
21. Zangani C, Ostinelli EG, Smith KA, Hong JS, Macdonald O, Reen G, et al. Impact of the COVID-19 pandemic on the global delivery of mental health services and telemental health: systematic review. JMIR Ment Health 2022 Aug 22;9(8):e38600 [FREE Full text] [doi: 10.2196/38600] [Medline: 35994310]
22. Karsh B. Beyond usability: designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394 [FREE Full text] [doi: 10.1136/qhc.13.5.388] [Medline: 15465944]
23. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
24. Whitten PS, Richardson JD. A scientific approach to the assessment of telemedicine acceptance. J Telemed Telecare 2002;8(4):246-248. [doi: 10.1258/135763302320272257] [Medline: 12217111]
25. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989 Aug;35(8):982-1003 [FREE Full text] [doi: 10.1287/mnsc.35.8.982]
26. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Jul;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
27. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. J Mark Res 1980 Nov;17(4):460-469 [FREE Full text] [doi: 10.2307/3150499]
28. Giese JL, Cote JA. Defining consumer satisfaction. Acad Mark Sci Rev 2000;4:1-24 [FREE Full text]
29. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
30. Maillet É, Mathieu L, Sicotte C. Modeling factors explaining the acceptance, actual use and satisfaction of nurses using an electronic patient record in acute care settings: an extension of the utaut. Int J Med Inform 2015 Jan;84(1):36-47. [doi: 10.1016/j.ijmedinf.2014.09.004] [Medline: 25288192]

31. Wang Q, Khan MS, Khan MK. Predicting user perceived satisfaction and reuse intentions toward massive open online courses (MOOCs) in the COVID-19 pandemic. Int J Acad Res Bus Soc Sci 2021 Mar 21;10(2):1-11 [FREE Full text] [doi: 10.20525/ijrbs.v10i2.1045]
32. Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and governing e-mental health care in Australia: a systematic review of challenges and a call for policy-focussed research. J Med Internet Res 2016 Jan 13;18(1):e10 [FREE Full text] [doi: 10.2196/jmir.4827] [Medline: 26764181]
33. Harerimana B, Forchuk C, O'Regan T. The use of technology for mental healthcare delivery among older adults with depressive symptoms: a systematic literature review. Int J Ment Health Nurs 2019 Jun;28(3):657-670. [doi: 10.1111/inm.12571] [Medline: 30666762]
34. Appleton R, Williams J, Vera San Juan N, Needle JJ, Schlief M, Jordan H, et al. Implementation, adoption, and perceptions of telemental health during the COVID-19 pandemic: systematic review. J Med Internet Res 2021 Dec 09;23(12):e31746 [FREE Full text] [doi: 10.2196/31746] [Medline: 34709179]
35. Schlief M, Saunders KR, Appleton R, Barnett P, Vera San Juan N, Foye U, et al. Synthesis of the evidence on what works for whom in telemental health: rapid realist review. Interact J Med Res 2022 Sep 29;11(2):e38239 [FREE Full text] [doi: 10.2196/38239] [Medline: 35767691]
36. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
37. Ahmadi M, Sarabi RE, Orak RJ, Bahaadinbeigy K. Information retrieval in telemedicine: a comparative study on bibliographic databases. Acta Inform Med 2015 Jun;23(3):172-176 [FREE Full text] [doi: 10.5455/aim.2015.23.172-176] [Medline: 26236086]
38. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
39. Study quality assessment tools. National Heart, Lung, and Blood Institute. 2021 Jul. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2022-08-05]
40. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890 [FREE Full text] [doi: 10.1136/bmj.l6890] [Medline: 31948937]
41. Ainslie M, Brunette MF, Capozzoli M. Treatment interruptions and telemedicine utilization in serious mental illness: retrospective longitudinal claims analysis. JMIR Ment Health 2022 Mar 21;9(3):e33092 [FREE Full text] [doi: 10.2196/33092] [Medline: 35311673]
42. Ceniti AK, Abdelmoemin WR, Ho K, Kang Y, Placenza F, Laframboise R, et al. "One degree of separation": a mixed-methods evaluation of Canadian mental health care user and provider experiences with remote care during COVID-19. Can J Psychiatry 2022 Sep;67(9):712-722. [doi: 10.1177/07067437211070656] [Medline: 34986035]
43. Chakawa A, Belzer LT, Perez-Crawford T, Yeh HW. COVID-19, telehealth, and pediatric integrated primary care: disparities in service use. J Pediatr Psychol 2021 Sep 27;46(9):1063-1075 [FREE Full text] [doi: 10.1093/jpepsy/jsab077] [Medline: 34343329]
44. Connolly SL, Stolzmann KL, Heyworth L, Sullivan JL, Shimada SL, Weaver KR, et al. Patient and provider predictors of telemental health use prior to and during the COVID-19 pandemic within the department of veterans affairs. Am Psychol 2022 Feb;77(2):249-261 [FREE Full text] [doi: 10.1037/amp0000895] [Medline: 34941310]
45. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Patient attitudes toward telepsychiatry during the COVID-19 pandemic: a nationwide, multisite survey. JMIR Ment Health 2020 Dec 22;7(12):e24761 [FREE Full text] [doi: 10.2196/24761] [Medline: 33302254]
46. Haxhihamza K, Arsova S, Bajraktarov S, Kalpak G, Stefanovski B, Novotni A, et al. Patient satisfaction with use of telemedicine in university clinic of psychiatry: Skopje, North Macedonia during COVID-19 pandemic. Telemed J E Health 2021 Apr;27(4):464-467. [doi: 10.1089/tmj.2020.0256] [Medline: 32809916]
47. Hutchison M, Russell BS, Gans KM, Starkweather AR. Online administration of a pilot mindfulness-based intervention for adolescents: feasibility, treatment perception and satisfaction. Curr Psychol 2022 Apr 01:1-13. [doi: 10.1007/s12144-022-03025-x] [Medline: 35382039]
48. Lewis YD, Elran-Barak R, Grundman-Shem Tov R, Zubery E. The abrupt transition from face-to-face to online treatment for eating disorders: a pilot examination of patients' perspectives during the COVID-19 lockdown. J Eat Disord 2021 Mar 05;9(1):31 [FREE Full text] [doi: 10.1186/s40337-021-00383-y] [Medline: 33673876]
49. Lohmiller J, Schäffeler N, Zipfel S, Stengel A. Higher acceptance of videotelephonic counseling formats in psychosomatic medicine in times of the COVID-19 pandemic. Front Psychiatry 2021 Oct 27;12:747648 [FREE Full text] [doi: 10.3389/fpsyt.2021.747648] [Medline: 34777051]
50. Lynch DA, Stefancic A, Cabassa LJ, Medalia A. Client, clinician, and administrator factors associated with the successful acceptance of a telehealth comprehensive recovery service: a mixed methods study. Psychiatry Res 2021 Jun;300:113871 [FREE Full text] [doi: 10.1016/j.psychres.2021.113871] [Medline: 33887517]

51. Meininger L, Adam J, von Wirth E, Viefhaus P, Woitecki K, Walter D, et al. Cognitive-behavioral teletherapy for children and adolescents with mental disorders and their families during the COVID-19 pandemic: a survey on acceptance and satisfaction. Child Adolesc Psychiatry Ment Health 2022 Jul 28;16(1):61 [FREE Full text] [doi: 10.1186/s13034-022-00494-7] [Medline: 35902959]
52. Michaels TI, Singal S, Marcy P, Hauser M, Braider L, Guinart D, et al. Post-acute college student satisfaction with telepsychiatry during the COVID-19 pandemic. J Psychiatr Res 2022 Jul;151:1-7 [FREE Full text] [doi: 10.1016/j.jpsychires.2022.03.035] [Medline: 35427873]
53. Miu AS, Vo HT, Palka JM, Glowacki CR, Robinson RJ. Teletherapy with serious mental illness populations during COVID-19: telehealth conversion and engagement. Couns Psychol Q 2021;34(3-4):704-721 [FREE Full text] [doi: 10.1080/09515070.2020.1791800]
54. Morgan AA, Landers AL, Simpson JE, Russon JM, Case Pease J, Dolbin-MacNab ML, et al. The transition to teletherapy in marriage and family therapy training settings during COVID-19: what do the data tell us? J Marital Fam Ther 2021 Apr;47(2):320-341 [FREE Full text] [doi: 10.1111/jmft.12502] [Medline: 33742728]
55. Nesset MB, Lauvrud C, Meisingset A, Nyhus E, Palmstierna T, Lara-Cabrera ML. Development of nurse-led videoconference-delivered cognitive behavioural therapy for domestic violence: feasibility and acceptability. J Adv Nurs 2023 Apr;79(4):1503-1512. [doi: 10.1111/jan.15347] [Medline: 35774003]
56. Severe J, Tang R, Horbatch F, Onishchenko R, Naini V, Blazek MC. Factors influencing patients' initial decisions regarding telepsychiatry participation during the COVID-19 pandemic: telephone-based survey. JMIR Form Res 2020 Dec 22;4(12):e25469 [FREE Full text] [doi: 10.2196/25469] [Medline: 33320823]
57. Sizer MA, Bhatta D, Acharya B, Paudel KP. Determinants of telehealth service use among mental health patients: a case of rural Louisiana. Int J Environ Res Public Health 2022 Jun 06;19(11):6930 [FREE Full text] [doi: 10.3390/ijerph19116930] [Medline: 35682512]
58. Ter Heide FJ, de la Rie S, de Haan A, Boeschoten M, Nijdam MJ, Smid G, et al. Wellbeing and clinical videoconferencing satisfaction among patients in psychotrauma treatment during the coronavirus pandemic: cross-sectional study. Eur J Psychotraumatol 2021 May 11;12(1):1906021 [FREE Full text] [doi: 10.1080/20008198.2021.1906021] [Medline: 34025924]
59. Tobin ET, Hadwiger A, DiChiara A, Entz A, Miller-Matero LR. Demographic predictors of telehealth use for integrated psychological services in primary care during the COVID-19 pandemic. J Racial Ethn Health Disparities 2023 Jun;10(3):1492-1498. [doi: 10.1007/s40615-022-01334-2] [Medline: 35794514]
60. Vakil T, Svenne DC, Bolton JM, Jiang D, Svenne S, Hensel JM. Analysis of the uptake and associated factors for virtual crisis care during the pandemic at a 24-h mental health crisis centre in Manitoba, Canada. BMC Psychiatry 2022 Aug 04;22(1):527 [FREE Full text] [doi: 10.1186/s12888-022-04166-w] [Medline: 35927672]
61. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the telemedicine satisfaction questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [doi: 10.1258/135763303321159693] [Medline: 12641893]
62. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-x] [Medline: 10259963]
63. Marshall GN, Hays RD. The patient satisfaction questionnaire short-form (PSQ-18). RAND Corporation. Santa Monica, CA: RAND Corporation; 1994. URL: https://www.rand.org/content/dam/rand/pubs/papers/2006/P7865.pdf [accessed 2023-04-26]
64. Marsden J, Stewart D, Gossop M, Rolfe A, Bacchus L, Griffiths P, et al. Assessing client satisfaction with treatment for substance use problems and the development of the Treatment Perceptions Questionnaire (TPQ). Addict Res 2000;8(5):455-470 [FREE Full text] [doi: 10.3109/16066350009005590]
65. Ritterband LM, Ardalan K, Thorndike FP, Magee JC, Saylor DK, Cox DJ, et al. Real world use of an Internet intervention for pediatric encopresis. J Med Internet Res 2008 Jun 30;10(2):e16 [FREE Full text] [doi: 10.2196/jmir.1081] [Medline: 18653440]
66. Leaf PJ, Bruce ML. Gender differences in the use of mental health-related services: a re-examination. J Health Soc Behav 1987 Jun;28(2):171-183 [FREE Full text] [doi: 10.2307/2137130]
67. Rhodes AE, Goering PN, To T, Williams JI. Gender and outpatient mental health service use. Soc Sci Med 2002 Jan;54(1):1-10. [doi: 10.1016/s0277-9536(01)00002-8] [Medline: 11820673]
68. Kovess-Masfety V, Boyd A, van de Velde S, de Graaf R, Vilagut G, Haro JM, EU-WMH investigators. Are there gender differences in service use for mental disorders across countries in the European Union? Results from the EU-world mental health survey. J Epidemiol Community Health 2014 Jul;68(7):649-656. [doi: 10.1136/jech-2013-202962] [Medline: 24616352]
69. Roberts T, Miguel Esponda G, Krupchanka D, Shidhaye R, Patel V, Rathod S. Factors associated with health service utilisation for common mental disorders: a systematic review. BMC Psychiatry 2018 Aug 22;18(1):262 [FREE Full text] [doi: 10.1186/s12888-018-1837-1] [Medline: 30134869]
70. Crabb R, Hunsley J. Utilization of mental health care services among older adults with depression. J Clin Psychol 2006 Mar;62(3):299-312. [doi: 10.1002/jclp.20231] [Medline: 16400646]

71. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Inf Syst J 2018 Jan 29;28(6):995-1019 [FREE Full text] [doi: 10.1111/isj.12179]
72. Chen J, Li KY, Andino J, Hill CE, Ng S, Steppe E, et al. Predictors of audio-only versus video telehealth visits during the COVID-19 pandemic. J Gen Intern Med 2022 Apr;37(5):1138-1144 [FREE Full text] [doi: 10.1007/s11606-021-07172-y] [Medline: 34791589]
73. Cousins MM, Van Til M, Steppe E, Ng S, Ellimoottil C, Sun Y, et al. Age, race, insurance type, and digital divide index are associated with video visit completion for patients seen for oncologic care in a large hospital system during the COVID-19 pandemic. PLoS One 2022 Nov 17;17(11):e0277617 [FREE Full text] [doi: 10.1371/journal.pone.0277617] [Medline: 36395112]
74. Karimi M, Lee EC, Couture SJ, Gonzales A, Grigorescu V, Smith SR, et al. National trends in telehealth use in 2021: disparities in utilization and audio vs. video services. Department of Health and Human Services. 2022. URL: https://aspe.hhs.gov/sites/default/files/documents/4e1853c0b4885112b2994680a58af9ed/telehealth-hps-ib.pdf [accessed 2023-01-10]
75. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Aug 29;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
76. Chen PV, Helm A, Caloudas SG, Ecker A, Day G, Hogan J, et al. Evidence of phone vs video-conferencing for mental health treatments: a review of the literature. Curr Psychiatry Rep 2022 Oct;24(10):529-539 [FREE Full text] [doi: 10.1007/s11920-022-01359-8] [Medline: 36053400]
77. Hajek A, Kretzler B, König HH. Personality, healthcare use and costs-a systematic review. Healthcare (Basel) 2020 Sep 09;8(3):329 [FREE Full text] [doi: 10.3390/healthcare8030329] [Medline: 32916927]
78. Sirois FM, Owens J. A meta-analysis of loneliness and use of primary health care. Health Psychol Rev 2023 Jun;17(2):193-210. [doi: 10.1080/17437199.2021.1986417] [Medline: 34581240]
79. Vera San Juan N, Shah P, Schlief M, Appleton R, Nyikavaranda P, Birken M, et al. Service user experiences and views regarding telemental health during the COVID-19 pandemic: a co-produced framework analysis. PLoS One 2021 Sep 16;16(9):e0257270 [FREE Full text] [doi: 10.1371/journal.pone.0257270] [Medline: 34529705]
80. Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. BMC Med Res Methodol 2019 Jun 28;19(1):132 [FREE Full text] [doi: 10.1186/s12874-019-0782-0] [Medline: 31253092]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by J Torous; submitted 01.02.23; peer-reviewed by J Severe, S Bidmon; comments to author 22.03.23; revised version received 03.05.23; accepted 31.05.23; published 18.08.23.

Please cite as:
Neumann A, König HH, Bokermann J, Hajek A
Determinants of Patient Use and Satisfaction With Synchronous Telemental Health Services During the COVID-19 Pandemic: Systematic Review
JMIR Ment Health 2023;10:e46148
URL: https://mental.jmir.org/2023/1/e46148
doi:10.2196/46148
PMID:37594785

Original Paper

Telehealth and In-Person Behavioral Health Services in Rural Communities Before and During the COVID-19 Pandemic: Multisite Prospective Cohort Study

Marcia M Ward[1], PhD; Fred Ullrich[1], BA; Divya Bhagianadh[2], PhD; Eve-Lynn Nelson[3], PhD; James P Marcin[4], MD; Knute D Carter[5], PhD; Kari Beth Law[6], MD; Carly McCord[7], PhD; Jonathan Neufeld[8], PhD; Kimberly A S Merchant[1], MA

[1]Department of Health Management and Policy, University of Iowa, Iowa City, IA, United States
[2]School of Social Work, Rutgers University, New Brunswick, NJ, United States
[3]Department of Pediatrics, Kansas University Medical Center, Kansas City, KS, United States
[4]Department of Pediatrics, University of California Davis School of Medicine, Sacramento, CA, United States
[5]Department of Biostatistics, University of Iowa, Iowa City, IA, United States
[6]Department of Behavioral Medicine and Psychiatry, West Virginia University, Morgantown, WV, United States
[7]Department of Psychiatry and Behavioral Sciences and Educational Psychology, Texas A&M University, College Station, TX, United States
[8]Institute for Health Informatics, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Marcia M Ward, PhD
Department of Health Management and Policy
University of Iowa
College of Public Health Building S236
145 Riverside Dr
Iowa City, IA, 52242
United States
Phone: 1 319 384 3815
Fax: 1 319 384 4371
Email: marcia-m-ward@uiowa.edu

Abstract

Background: The COVID-19 pandemic triggered widespread adjustments across the US health care system. Telehealth use showed a substantial increase in mental health conditions and services due to acute public health emergency (PHE) behavioral health needs on top of long-standing gaps in access to behavioral health services. How health systems that were already providing behavioral telehealth services adjusted services and staffing during this period has not been well documented, particularly in rural areas with chronic shortages of behavioral health providers and services.

Objective: This study investigates patient and treatment characteristic changes from before the COVID-19 PHE to during the PHE within both telehealth and in-person behavioral health services provided in 95 rural communities across the United States.

Methods: We used a nonrandomized, prospective, multisite research design involving 2 active treatment groups. The telehealth cohort included all patients who initiated telehealth treatment regimens during the data collection period. A comparison group included a cohort of patients who initiated in-person treatment regimen. Patient enrollment occurred on a rolling basis, and data collection was extended for 3 months after treatment initiation for each patient. Chi-square tests compared changes from pre-PHE to PHE time periods within telehealth and in-person treatment cohorts. The dependent measures included patient diagnosis, clinicians providing treatment services, and type of treatment services provided at each encounter. The 4780 patients in the telehealth cohort and the 6457 patients in the in-person cohort had an average of 3.5 encounters during the 3-month follow-up period.

Results: The encounters involving anxiety, dissociative, and stress-related disorders in the telehealth cohort increased from 30% (698/2352) in the pre-PHE period to 35% (4632/12,853) in the PHE period (P<.001), and encounters involving substance use disorders in the in-person cohort increased from 11% (468/4249) in the pre-PHE period to 18% (3048/17,047) in the PHE period (P<.001). The encounters involving treatment service codes for alcohol, drug, and medication-assisted therapy in the

telehealth cohort increased from 1% (22/2352) in the pre-PHE period to 11% (1470/13,387) in the PHE period (P<.001); likewise, encounters for this type of service in the in-person cohort increased from 0% (0/4249) in the pre-PHE period to 16% (2687/17,047) in the PHE period (P<.001). From the pre-PHE to the PHE period, encounters involving 60-minute psychotherapy in the telehealth cohort increased from 8% (190/2352) to 14% (1802/13,387; P<.001), while encounters involving group therapy in the in-person cohort decreased from 12% (502/4249) to 4% (739/17,047; P<.001).

Conclusions: The COVID-19 pandemic challenged health service providers, and they adjusted the way both telehealth and in-person behavioral therapy services were delivered. Looking forward, future research is needed to explicate the interaction of patient, provider, setting, and intervention factors that influenced the patterns observed as a result of the COVID-19 pandemic.

(JMIR Ment Health 2023;10:e47047) doi:10.2196/47047

KEYWORDS

anxiety; behavior; behavioral health; COVID-19; depression; digital health; eHealth; mental health; mHealth; pandemic; rural health services; rural; telehealth; telemedicine

Introduction

The COVID-19 pandemic triggered widespread adjustments across the US health care system. Once the public health emergency (PHE) was declared, several major legislative initiatives and policy waivers were enacted to expand web-based access to needed health care [1]. The PHE and associated policy changes prompted providers and patients to consider telehealth as a specific treatment option to address the need for social distancing and related reasons [2]. Many providers adopted telehealth for the first time or expanded their telehealth offerings [3-6].

As the COVID-19 pandemic unfolded, behavioral health concerns rose rapidly [7-9]. The largest increase in telehealth use was for mental health conditions and services due to rising acute behavioral health needs on top of longstanding gaps in access to behavioral health services [4,5,7,10-12]. Since the onset of the COVID-19 pandemic, publications have appeared describing how health systems implemented behavioral telehealth to meet this new demand [13]. In particular, multiple publications highlighted the challenges to ramping up behavioral telehealth services and the perceptions among both patients and providers toward their adoption [4,5,7,10-12,14] but the literature on the pandemic lacks information on the changes in existing behavioral health services regarding adjustments to provider type and services to meet the increased demand and changing patient needs triggered by the pandemic, particularly in rural areas with chronic shortages of behavioral health providers and services [15]. Previous population studies reported increases in certain psychiatric disorders, such as anxiety and substance use, during the COVID-19 pandemic [16-19]. We hypothesize that if those findings are replicated in this study, then the treatment types used by behavioral health providers would have adjusted accordingly to meet the needs of the changing patient population. In addition, particularly lacking are studies that examine changes in both telehealth and in-person behavioral health services and those serving rural populations. Thus, the purpose of this analysis is to compare the profile of telehealth and in-person treatments across a large national sample of rural behavioral health care providers during the time period immediately before the PHE with those during the early phases of the PHE.

Methods

Overview

This study pools data from 2 grant programs funded by the Office for the Advancement of Telehealth in the Health Resources and Services Administration of the US Department of Health and Human Services: the Substance Abuse Treatment Telehealth Network Grant Program funded from September 2017 to August 2020 and the Evidence-Based Tele-Behavioral Health Network Program funded from September 2018 to August 2021. Through these 2 programs, 17 grantees provided telebehavioral health services to 95 rural communities in 13 states (California, Indiana, Kansas, Kentucky, Maryland, Massachusetts, Minnesota, Missouri, Oregon, Pennsylvania, South Dakota, Texas, and West Virginia). For both grant programs, “rural” was defined by Health Resources and Services Administration as all nonmetro counties, all metro census tracts with rural-urban commuting area (RUCA) codes 4-10, and large area metro census tracts of at least 400 square miles in area with population densities of 35 or less per square mile with RUCA codes 2-3. The originating sites included behavioral health clinics, substance use treatment facilities, primary care outpatient clinics, rural health clinics, federally qualified health clinics, critical access hospitals, and schools; for telehealth, the sites providing services included these and larger medical centers. The grantees all had established telehealth networks before grant funding that delivered behavioral telehealth services through synchronous video connections. According to funding guidance, the implemented behavioral health services followed standard therapeutic guidelines tailored to the needs, resources, and capacities of the grantees and their rural communities.

This study was a multisite, nonrandomized, prospective research design that involved 2 active usual-care treatment cohorts. The telehealth cohort included all patients who began telehealth-based treatment as part of either grant program during the data collection periods. Grantees were asked to provide data on patients who began in-person treatment, had similar primary complaints or diagnoses, demographics, and received comparable in-person treatment to their telehealth cohort. This group of patients forms the in-person cohort. Patients were enrolled on a rolling basis. The data collection at each of the sites focused on initial visits and all encounters that followed in the first 3 months of treatment for each patient.

The Rural Telehealth Research Center was funded by the Office for the Advancement of Telehealth to serve as a data coordinating center for the programs. We identified 26 component data elements [20], including patient and treatment characteristics, and created a data element dictionary, study protocol, and Excel (Microsoft Corp)-based tool for data collection.

Data and Analysis

Grantees transmitted data to the coordinating center quarterly. Overall, 2 time periods relevant to the COVID-19 PHE were defined. Data transmitted from October 2019 to March 2020 (5 months) were considered in the pre-PHE time period, and data transmitted from April 2020 to July 2021 (15 months) were considered in the PHE time period. Patient and treatment characteristics during each encounter were compared across the 2 time periods with chi-square tests using Stata 16 within the telehealth cohort and the in-person cohort. The dependent measures were collected at each encounter and included patient diagnosis, clinician type providing treatment services, and type of treatment services provided. Patient diagnosis was recorded as International Classification of Disease-10 codes; clinician providing treatment were recorded from a checklist or open text field; and treatment service type was recorded in the form of current procedural terminology (CPT) or Health Common Procedure Coding System (HCPCS) codes.

Ethical Considerations

Given that no experimental interventions were involved and all data were deidentified before transfer to the data coordinating center, the study protocol for data analysis was deemed not human subjects research by the University of Iowa's institutional review board (IRB) chair (IRB-01 #201912016). The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Study protocols at each contributing grantee were reviewed by that grantee's IRB.

Data transfer and use agreements between each grantee and the University of Iowa were established to protect the confidentiality of the data being analyzed. We performed data monitoring and management activities to verify data accuracy, completeness, consistency, and timeliness. Office of Management and Budget clearance was received in October 2019, and grantees provided data from then until July 2021.

Results

Grantees provided data on both the telehealth and in-person cohorts using the data collection tool, resulting in a data set with 4780 patients in the telehealth cohort and 6457 patients in the in-person cohort. During the first 3 months of treatment, individuals in the telehealth cohort averaged 3.5 encounters and those in the in-person cohort averaged 3.4 encounters. The number of encounters did not differ significantly between time periods for either the telehealth or in-person cohorts (P=.63 and P=.58, respectively). Differences within the telehealth and in-person cohorts across the 2 time periods with respect to patient demographic characteristics are presented in Multimedia Appendix 1.

As shown in Table 1, the patient diagnosis recorded at each encounter differed within each treatment cohort between time periods. Within the telehealth cohort, there was an increase in the percent of encounters involving anxiety, dissociative, and stress-related disorders from 30% (698/2352) in the pre-PHE period to 35% (4632/12,853) in the PHE period (P<.001), while the percent of encounters involving non–mental health diagnoses decreased from 15% (353/2352) in the pre-PHE period to 6% (763/12,853) in the PHE period (P<.001). Within the in-person cohort, the percent of encounters involving substance use disorders increased from 11% (468/4249) in the pre-PHE period to 18% (3048/17,047) in the PHE period (P<.001), while the percent of encounters involving mood (affective) disorders decreased from 37% (1550/4249) in the pre-PHE period to 28% (4685/17,047) in the PHE period (P<.001).

As shown in Table 2, the treatment service type for encounters in the form of CPT or HCPCS codes differed significantly within the 2 treatment cohorts between time periods. Within the telehealth cohort, from the pre-PHE period to the PHE period, the percent of encounters involving HCPCS codes for alcohol, drug, and medication-assisted therapy increased substantially from 1% (22/2352) to 11% (1470/13,387; P<.001) as did the percent of encounters involving 60-minute psychotherapy, which increased from 8% (190/2352) to 14% (1802/13,387; P<.001), while the percent of encounters involving health and behavior interviews decreased from 8% (197/2352) to 0.2% (25/13,387; P<.001). Within the in-person cohort, from the pre-PHE period to the PHE period, the percent of encounters involving HCPCS codes for alcohol, drug, and medication-assisted therapy increased substantially from 0% (0/4249) to 16% (2687/17,047; P<.001) while the percent of encounters involving group therapy decreased from 12% (502/4249) to 4% (739/17,047; P<.001).

As shown in Table 3, the clinician type for encounters significantly differed between time periods within the 2 treatment cohorts. From the pre-PHE period to the PHE period, within the telehealth cohort, encounters involving licensed or provisionally licensed counselors increased from 22% (508/2353) to 27% (3674/13,387; P<.001) and from 12% (293/2353) to 15% (2040/13,387; P<.001) for clinical psychologists, while the percent of encounters involving clinical social workers decreased from 32% (783/2352) to 22% (2913/13,387; P<.001) and from 18% (416/2352) to 9% (1209/13,387; P<.001) for psychiatric and mental health advanced practice providers. From the pre-PHE period to the PHE period within the in-person cohort, encounters involving licensed or provisionally licensed professional counselors increased from 0% (0/4249) to 11% (1881/17,047; P<.001), from 2% (88/4249) to 11% (1797/17,047; P<.001) for clinical psychologists, and from 7% (299/4249) to 9% (1477/17,047; P<.001) for psychiatrists and other physicians. At the same time, the percent of encounters involving clinical social workers decreased substantially from 86% (3649/4249) to 55% (9313/17,047; P<.001) from the pre-PHE period to the PHE period. In both cohorts, the percent of unknown clinician type was almost 0% (16/6601) in the pre-PHE period but exceeded 10% (3215/30,434) in the PHE period (P<.001).

Table 1. Frequencies and percentages of encounters with each diagnostic category for the 2 treatment cohorts by time period. Categories with low n (ie, less than 3%) have been combined in to "Other mental health primary diagnoses."

Patient diagnostic category	Telehealth cohort		In-person cohort	
	Pre-PHE[a], n (%)	PHE, n (%)	Pre-PHE, n (%)	PHE, n (%)
Anxiety, dissociative, stress- related, somatoform, and other nonpsychotic mental disorders (F40-F48)	698 (29.7)	4632 (34.6)	1762 (41.5)	6936 (40.7)
Mood (affective) disorders (F30-F39)	686 (29.2)	4028 (30.1)	1550 (36.5)	4685 (27.5)
Mental and behavioral disorders due to psychoactive substance use (F10-F19)	405 (17.2)	2374 (17.7)	468 (11)	3048 (17.9)
Behavioral and emotional disorders with onset usually occurring in childhood and adolescence (F90-F98)	152 (6.5)	941 (7)	152 (3.7)	941 (5.5)
Other mental health primary diagnoses	58 (2.5)	115 (4.8)	250 (5.9)	720 (4.2)
Nonmental health primary diagnosis	353 (15)	763 (5.7)	67 (1.6)	717 (4.2)

[a]PHE: public health emergency.

Table 2. Frequencies and percentages of encounters with each treatment service type for the 2 treatment cohorts by time period. Categories with low n (ie, less than 2%) have been combined into the "Other" category.

Treatment service type	Telehealth cohort		In-person cohort	
	Pre-PHE[a], n (%)	PHE, n (%)	Pre-PHE, n (%)	PHE, n (%)
90832 Individual psychotherapy (30 min)	233 (9.9)	1478 (11)	221 (5.2)	1562 (9.2)
90834 Individual psychotherapy (45 min)	478 (20.3)	2371 (17.7)	1055 (24.8)	2676 (15.7)
90837 Individual psychotherapy (60 min)	190 (8.1)	1802 (13.5)	1273 (30)	4788 (28.1)
90853 Group psychotherapy (60 min)	129 (5.5)	720 (5.4)	502 (11.8)	739 (4.3)
90791-90792 Psychiatric diagnostic evaluation	314 (13.3)	1458 (10.9)	701 (16.5)	2189 (12.9)
96152 Health and behavior intervention	197 (8.4)	25 (0.2)	5 (0.1)	0 (0)
99213-99214 Evaluation and management	341 (14.5)	240 (15.2)	315 (7.4)	1060 (6.2)
HCPCS[b] codes for alcohol, drug therapy	22 (0.9)	1470 (11)	0 (0)	2687 (15.8)
Other	448 (19.1)	4063 (15.1)	177 (4.2)	1346 (7.8)

[a]PHE: public health emergency.

[b]HCPCS: Health Common Procedure Coding System.

Table 3. Frequencies and percentages of encounters with each clinician type for the 2 treatment cohorts by time period. Categories with a low n have been combined into the "Other" category.

Clinician type	Telehealth cohort		In-person cohort	
	Pre-PHE[a], n (%)	PHE, n (%)	Pre-PHE, n (%)	PHE, n (%)
Clinical social worker	783 (32)	2913 (21.8)	3649 (85.9)	9313 (54.7)
Licensed or provisionally licensed counselor	508 (21.6)	3674 (27.4)	0 (0)	1881 (11)
Clinical psychologist	293 (12.5)	2040 (15.2)	88 (2.1)	1797 (10.5)
Psychiatrist and other physician	294 (12.5)	1815 (13.6)	299 (7)	1477 (8.7)
Psychiatric and mental health advanced practice provider	416 (17.7)	1209 (9)	213 (5)	829 (4.9)
Other	72 (3.1)	250 (1.9)	0 (0)	21 (0.1)
Unknown	16 (0.7)	1486 (11.1)	0 (0)	1729 (10.1)

[a]PHE: public health emergency.

Discussion

Overview

Behavioral health concerns rose rapidly once the COVID-19 pandemic occurred [7-9]. How behavioral health interventions adjusted to meet this rising need is important to understand. The current pooled data across behavioral health services in 95 rural communities identified multiple changes from the pre-PHE time period to the PHE time period for both the telehealth and in-person cohorts, and these changes included both patient characteristics and treatment processes. In particular, from the pre-PHE period to the PHE period, the telehealth cohort encounters involving anxiety, dissociative, and stress-related disorders increased from 30% (698/2352) to 35%(4632/12,853), and in-person cohort encounters involving substance use disorders increased from 11% (468/4249) to 18% (3048/17,047). Associated with these changes, from the pre-PHE period to the PHE period, telehealth cohort encounters involving treatment service codes for alcohol, drug, and medication-assisted therapy increased from 1% (22/2352) to 11% (1470/13,387), and the in-person cohort increased from 0% (0/4249) to 16% (2687/17,047). The changes observed from the pre-PHE to the PHE period were likely influenced by an interaction of patient, provider, setting, and intervention factors in these nonrandomized cohorts as usual-care approaches adjusted to the COVID-19 pandemic.

A large component of the differences between time periods was due to differences in the patient diagnosis recorded at each encounter. These increases in anxiety, dissociative, stress-related disorders, and substance use disorders are consistent with those found in population studies during the COVID-19 pandemic [16] and have been attributed to pandemic stressors impacting communities as a whole [16,17], as well as the potential psychiatric impacts associated with COVID-19 illness [18,19].

In responding to these changes in patient diagnostic category across time periods, the data indicate that behavioral health clinicians adjusted the treatment service type (in the form of CPT or HCPCS codes) for encounters within both treatment cohorts. Within both the telehealth and in-person cohorts, the percent of encounters involving HCPCS codes for alcohol, drug, and medication-assisted therapy increased substantially. Another change in service that was noted from the pre-PHE to the PHE period was that the telehealth cohort encounters involving 60-minute psychotherapy increased from 8% (190/2352) to 14% (1802/13,387), while in-person cohort encounters involving group therapy decreased from 12% (502/4249) to 4% (739/17,047). The decrease in group therapy is not surprising given that the PHE prompted providers and patients to consider treatment options that met the requirements for social distancing [2,21]. Previous studies have reported that a majority of mental health practitioners reported at least one practice adjustment during the PHE, such as providing additional therapeutic services [22] and that both the process and content of therapy often changed [23]. Changes in the service type likely resulted from a mix of other factors, including patient availability and preferences, changed provider preferences, and changed provider schedules or demands during the pandemic [24,25].

Analyses indicated that the provider type differed between time periods within the 2 treatment cohorts. Within both cohorts, the percent of encounters involving licensed or provisionally licensed professional counselors, clinical psychologists, and clinicians of unknown type increased from the pre-PHE period to the PHE period. The changes in clinician type providing services during the PHE may have resulted from changes by the Centers for Medicare and Medicaid Services and provisions of the US Coronavirus Aid, Relief, and Economic Security Act emergency policies that included authorization for multiple types of clinicians to offer telehealth services, allowance for clinicians to serve out-of-state patients, and improved provider payments for telehealth [1]. Another factor that may have been involved was the need for health systems to adapt to increased patient demand while their workforce needed to adjust to a changing work environment, including more providers delivering telehealth from their home-based rather than office-based settings [26-28].

Previous COVID-19 studies have described how various health systems and clinics-initiated telehealth services for the first time. But the literature on the pandemic has lacked information on the changes in existing behavioral health services regarding adjustments to provider type and services to meet the increased demand and changing patient needs triggered by the pandemic. These findings help to fill that gap by describing how experienced telehealth grantees adjusted their existing services to meet the changing needs presented by the pandemic. How service providers responded to this emergency situation and changing patient diagnoses involved adjustments in the types of clinicians providing treatment and the types of therapy services provided. We hypothesized that patients would show increases in certain psychiatric conditions during the pandemic, and the findings supported this hypothesis. We also hypothesized that treatment types by behavioral health providers would have adjusted accordingly to meet the needs of the changing patient population, and analyses supported that hypothesis along with describing changes in clinician type. To our knowledge, these changes for both telehealth and in-person behavioral health services have not been previously reported. In addition, much of the literature on pandemic changes focuses on urban settings, and this analysis contributes to the sparse literature on rural behavioral health care.

Looking forward, the conclusion of the PHE will provide an opportunity to examine changes as regulations and policies affecting behavioral telehealth either remain in place or return to prepandemic status. Understanding patterns in patient and treatment characteristics could benefit from future research that uses a national set of claims data to examine changes related to the pandemic. As data sets become available, Medicare claims data can be used to examine changing patterns in older populations, but employer-based and Medicaid claims data will be important for elucidating changes affecting younger populations. Also, future research is needed to explicate the interaction of patient, provider, setting, and intervention factors that influenced the patterns observed as a result of the COVID-19 pandemic.

Limitations

Limitations to the study include the use of convenience samples, which may have introduced multiple biases, including treatment selection biases. In addition, the percentages of unknown clinician types increased during the PHE, which raises uncertainty about changes in those characteristics. Likewise, data are not available on organization changes that might have happened before and after the pandemic including possible changes in health plan contracts. Furthermore, external validity may be limited given the focus on rural patient populations with chronic shortages of behavioral health providers and services. Moreover, the data were limited to a diverse group of grantees as opposed to the general population, and the results reflected averaged changes that likely varied among grantees and treatment sites. Factors beyond the pandemic, such as Medicaid expansion in some states [29], potentially impacted changes in coverage and use. When examining changes from pre-PHE to PHE time periods, it will be important to further differentiate these policy-related effects.

Conclusions

This study pooled data across 17 behavioral health care systems, all of which had established in-person and telehealth services. Comparisons from the pre-PHE period to the PHE period indicated that patient diagnoses, treatment service type, and clinician type all changed in both the telehealth cohort and the in-person cohort.

Acknowledgments

This study was supported by the Office for the Advancement of Telehealth (OAT), Health Resources and Services Administration (HRSA), and the US Department of Health and Human Services (HHS) to the Rural Telehealth Research Center under cooperative agreements (#UICRH29074 and #U3GRH40003). The Evidence-Based Tele-Behavioral Health Network Program provided funding for the delivery of telebehavioral health to these grantees, providing services with the following grants: Athol Hospital/Heywood Healthcare (G01RH321490), Avera Health (G01RH32150), Baptist Health Corbin Foundation (G01RH32151), CentraCare Health System (G01RH32152), Greater Oregon Behavioral Health, Inc (G01RH32153), Indiana Rural Health Association (G01RH32154), Lester E Cox Medical Center (G01RH32155), Primary Health Network (G01RH32156), Regents of the University of Minnesota (G01RH32157), Texas A & M University (G01RH32158), University of California, Davis (G01RH32159), University of Kansas Hospital (G01RH32160), University of Maryland (G01RH32161), and West Virginia University Research Corporation (G01RH32162). The Substance Abuse Treatment Telehealth Network Grant Program provided funding for the delivery of telesubstance abuse treatment to these grantees, providing services with the following grants: Avera Health (H1W-RH-31446), Union Hospital, Inc (H1W-RH-31447), and Westbrook Health Services, Inc (H1W-RH-31448). The information, conclusions, and opinions expressed are those of the authors, and no endorsement by OAT, HRSA, or HHS is intended or should be inferred.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient characteristics.
[DOCX File , 20 KB - mental_v10i1e47047_app1.docx]

References

1. Federal policy pertaining to COVID-19. Center for Connected Health Policy. URL: https://www.cchpca.org/federal/?category=covid-19&topic=originating-site [accessed 2023-01-15]
2. Hollander JE, Carr BG. Virtually perfect? Telemedicine for COVID-19. N Engl J Med 2020;382(18):1679-1681 [FREE Full text] [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
3. Demeke HB, Pao LZ, Clark H, Romero L, Neri A, Shah R, et al. Telehealth practice among health centers during the COVID-19 pandemic—United States, July 11-17, 2020. MMWR Morb Mortal Wkly Rep 2020;69(50):1902-1905 [FREE Full text] [doi: 10.15585/mmwr.mm6950a4] [Medline: 33332297]
4. Friedman AB, Gervasi S, Song H, Bond AM, Chen AT, Bergman A, et al. Telemedicine catches on: changes in the utilization of telemedicine services during the COVID-19 pandemic. Am J Manag Care 2022;28(1):e1-e6 [FREE Full text] [doi: 10.37765/ajmc.2022.88771] [Medline: 35049260]
5. Weiner JP, Bandeian S, Hatef E, Lans D, Liu A, Lemke KW. In-person and telehealth ambulatory contacts and costs in a large US insured cohort before and during the COVID-19 pandemic. JAMA Netw Open 2021;4(3):e212618 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2618] [Medline: 33755167]
6. Vogt EL, Welch BM, Bunnell BE, Barrera JF, Paige SR, Owens M, et al. Quantifying the impact of COVID-19 on telemedicine utilization: retrospective observational study. Interact J Med Res 2022;11(1):e29880 [FREE Full text] [doi: 10.2196/29880] [Medline: 34751158]

7. Cantor J, McBain RK, Kofner A, Hanson R, Stein BD, Yu H. Telehealth adoption by mental health and substance use disorder treatment facilities in the COVID-19 pandemic. Psychiatr Serv 2022;73(4):411-417 [FREE Full text] [doi: 10.1176/appi.ps.202100191] [Medline: 34407631]
8. Data brief: telehealth was critical for providing services to medicare beneficiaries during the first year of the COVID-19 pandemic OEI-02-20-00520. Office of Inspector General. 2022. URL: https://oig.hhs.gov/oei/reports/OEI-02-20-00520.asp [accessed 2023-01-15]
9. Mulvaney-Day N, Dean D, Miller K, Camacho-Cook J. Trends in use of telehealth for behavioral health care during the COVID-19 pandemic: considerations for payers and employers. Am J Health Promot 2022;36(7):1237-1241 [FREE Full text] [doi: 10.1177/08901171221112488e] [Medline: 36003014]
10. Uscher-Pines L, Thompson J, Taylor P, Dean K, Yuan T, Tong I, et al. Where virtual care was already a reality: experiences of a nationwide telehealth service provider during the COVID-19 pandemic. J Med Internet Res 2020;22(12):e22727 [FREE Full text] [doi: 10.2196/22727] [Medline: 33112761]
11. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
12. Zangani C, Ostinelli EG, Smith KA, Hong JSW, Macdonald O, Reen G, et al. Impact of the COVID-19 pandemic on the global delivery of mental health services and telemental health: systematic review. JMIR Ment Health 2022;9(8):e38600 [FREE Full text] [doi: 10.2196/38600] [Medline: 35994310]
13. Folk JB, Schiel MA, Oblath R, Feuer V, Sharma A, Khan S, et al. The transition of academic mental health clinics to telehealth during the COVID-19 pandemic. J Am Acad Child Adolesc Psychiatry 2022;61(2):277-290.e2 [FREE Full text] [doi: 10.1016/j.jaac.2021.06.003] [Medline: 34119633]
14. Appleton R, Williams J, Juan NVS, Needle JJ, Schlief M, Jordan H, et al. Implementation, adoption, and perceptions of telemental health during the COVID-19 pandemic: systematic review. J Med Internet Res 2021;23(12):e31746 [FREE Full text] [doi: 10.2196/31746] [Medline: 34709179]
15. Butzner M, Cuffee Y. Telehealth interventions and outcomes across rural communities in the United States: narrative review. J Med Internet Res 2021;23(8):e29575 [FREE Full text] [doi: 10.2196/29575] [Medline: 34435965]
16. Czeisler M, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic—United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
17. Lieberman JA, Olfson M. Meeting the mental health challenge of the COVID-19 pandemic. Psychiatric Times. 2020. URL: https://www.psychiatrictimes.com/view/meeting-mental-health-challenge-covid-19-pandemic [accessed 2023-01-15]
18. Taquet M, Sillett R, Zhu L, Mendel J, Camplisson I, Dercon Q, et al. Neurological and psychiatric risk trajectories after SARS-CoV-2 infection: an analysis of 2-year retrospective cohort studies including 1,284,437 patients. Lancet Psychiatry 2022;9(10):815-827 [FREE Full text] [doi: 10.1016/S2215-0366(22)00260-7] [Medline: 35987197]
19. Shafran R, Rachman S, Whittal M, Radomsky A, Coughtrey A. Fear and anxiety in COVID-19: preexisting anxiety disorders. Cogn Behav Pract 2021;28(4):459-467 [FREE Full text] [doi: 10.1016/j.cbpra.2021.03.003] [Medline: 33897249]
20. Ward MM, Fox K, Merchant K, Burgess A, Ullrich F, Parenteau L, et al. Measure and data element identification for the HRSA evidence-based tele-behavioral health network program and the HRSA substance abuse treatment telehealth network grant program. Rural Telehealth Research Center. 2020. URL: https://ruraltelehealth.org/research-and-policy-briefs.php [accessed 2023-05-01]
21. Oesterle TS, Kolla B, Risma CJ, Breitinger SA, Rakocevic DB, Loukianova LL, et al. Substance use disorders and telehealth in the COVID-19 pandemic era: a new outlook. Mayo Clin Proc 2020;95(12):2709-2718 [FREE Full text] [doi: 10.1016/j.mayocp.2020.10.011] [Medline: 33276843]
22. Reilly SE, Zane KL, McCuddy WT, Soulliard ZA, Scarisbrick DM, Miller LE, et al. Mental health practitioners' immediate practical response during the COVID-19 pandemic: observational questionnaire study. JMIR Ment Health 2020;7(9):e21237 [FREE Full text] [doi: 10.2196/21237] [Medline: 32931440]
23. Schriger SH, Klein MR, Last BS, Fernandez-Marcote S, Dallard N, Jones B, et al. Community mental health clinicians' perspectives on telehealth during the COVID-19 pandemic: mixed methods study. JMIR Pediatr Parent 2022;5(1):e29250 [FREE Full text] [doi: 10.2196/29250] [Medline: 35023839]
24. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Patient attitudes toward telepsychiatry during the COVID-19 pandemic: a nationwide, multisite survey. JMIR Ment Health 2020;7(12):e24761 [FREE Full text] [doi: 10.2196/24761] [Medline: 33302254]
25. Wilczewski H, Paige SR, Ong T, Soni H, Barrera JF, Welch BM, et al. Providers' perspectives on telemental health usage after the COVID-19 pandemic: retrospective analysis. JMIR Form Res 2022;6(11):e39634 [FREE Full text] [doi: 10.2196/39634] [Medline: 36322787]
26. Kisicki A, Becker S, Chaple M, Gustafson DH, Hartzler BJ, Jacobson N, et al. Behavioral healthcare organizations' experiences related to use of telehealth as a result of the COVID-19 pandemic: an exploratory study. BMC Health Serv Res 2022;22(1):775 [FREE Full text] [doi: 10.1186/s12913-022-08114-y] [Medline: 35698186]

27. Svistova J, Harris C, Fogarty B, Kulp C, Lee A. Use of telehealth amid the COVID-19 pandemic: experiences of mental health providers serving rural youth and elderly in Pennsylvania. Adm Policy Ment Health 2022;49(4):530-538 [FREE Full text] [doi: 10.1007/s10488-021-01181-z] [Medline: 34846613]
28. Zhu D, Paige SR, Slone H, Gutierrez A, Lutzky C, Hedriana H, et al. Exploring telemental health practice before, during, and after the COVID-19 pandemic. J Telemed Telecare 2021:1357633X211025943 [FREE Full text] [doi: 10.1177/1357633X211025943] [Medline: 34241545]
29. Status of state medicaid expansion decisions: interactive map. KFF State Health Facts. 2022. URL: https://www.kff.org/medicaid/issue-brief/status-of-state-medicaid-expansion-decisions-interactive-map/ [accessed 2023-01-15]

Abbreviations

CPT: current procedural terminology
HCPCS: Health Common Procedure Coding System
IRB: institutional review board
PHE: public health emergency
RUCA: rural-urban commuting area

Edited by J Torous; submitted 06.03.23; peer-reviewed by H Alshawaf, T Aslanidis, D Stubbings, D Hilty; comments to author 22.05.23; revised version received 23.07.23; accepted 28.07.23; published 18.09.23.

Please cite as:
Ward MM, Ullrich F, Bhagianadh D, Nelson EL, Marcin JP, Carter KD, Law KB, McCord C, Neufeld J, Merchant KAS
Telehealth and In-Person Behavioral Health Services in Rural Communities Before and During the COVID-19 Pandemic: Multisite Prospective Cohort Study
JMIR Ment Health 2023;10:e47047
URL: https://mental.jmir.org/2023/1/e47047
doi:10.2196/47047
PMID:37721793

Original Paper

Psychological Resilience Factors and Their Association With Weekly Stressor Reactivity During the COVID-19 Outbreak in Europe: Prospective Longitudinal Study

Sophie A Bögemann[1*], MSc; Lara M C Puhlmann[2,3*], PhD; Carolin Wackerhagen[4*], PhD; Matthias Zerban[5*], MSc; Antje Riepenhausen[4,6], PhD; Göran Köber[7,8], PhD; Kenneth S L Yuen[2,5], PhD; Shakoor Pooseh[8], PhD; Marta A Marciniak[9,10], PhD; Zala Reppmann[4], MSc; Aleksandra Uści ko[11], MSc; Jeroen Weermeijer[12], MSc; Dionne B Lenferink[1], MSc; Julian Mituniewicz[11], MSc; Natalia Robak[13], BSc; Nina C Donner[14], PhD; Merijn Mestdagh[15], PhD; Stijn Verdonck[15], PhD; Rolf van Dick[16], PhD; Birgit Kleim[9,10], PhD; Klaus Lieb[2,17], MD, PhD; Judith M C van Leeuwen[1], PhD; Dorota Kobylińska[11*], PhD; Inez Myin-Germeys[12*], PhD; Henrik Walter[4,6*], MD, PhD; Oliver Tüscher[2,17,18*], MD, PhD; Erno J Hermans[1*], PhD; Ilya M Veer[4,19*], PhD; Raffael Kalisch[2,5*], PhD

[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, Netherlands

[2]Leibniz Institute for Resilience Research (LIR), Mainz, Germany

[3]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[4]Research Division of Mind and Brain, Department of Psychiatry and Psychotherapy, Charité Campus Mitte, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[5]Neuroimaging Center (NIC), Focus Program Translational Neuroscience (FTN), Johannes Gutenberg University Medical Center, Mainz, Germany

[6]Berlin School of Mind and Brain, Faculty of Philosophy, Humboldt-Universität zu Berlin, Berlin, Germany

[7]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

[8]Freiburg Center for Data Analysis and Modelling, Institute of Physics, University of Freiburg, Freiburg, Germany

[9]Division of Experimental Psychopathology and Psychotherapy, Department of Psychology, University of Zurich, Zurich, Switzerland

[10]Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric University Hospital (PUK), University of Zurich, Zurich, Switzerland

[11]Faculty of Psychology, University of Warsaw, Warsaw, Poland

[12]Center for Contextual Psychiatry, Department of Neurosciences, KU Leuven, Leuven, Belgium

[13]College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Warsaw, Poland

[14]Concentris Research Management GmbH, Fürstenfeldbruck, Germany

[15]Research Group of Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium

[16]Institute of Psychology, Goethe University Frankfurt, Frankfurt am Main, Germany

[17]Department of Psychiatry and Psychotherapy, Johannes Gutenberg University Medical Center, Mainz, Germany

[18]Institute of Molecular Biology (IMB), Mainz, Germany

[19]Department of Developmental Psychology, University of Amsterdam, Amsterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Sophie A Bögemann, MSc
Donders Institute for Brain, Cognition and Behaviour
Radboud University Medical Center
Kapittelweg 29
Nijmegen, 6525 EN
Netherlands
Phone: 31 024 3614305
Email: Sophie.Bogemann@donders.ru.nl

Abstract

Background: Cross-sectional relationships between psychosocial resilience factors (RFs) and resilience, operationalized as the outcome of low mental health reactivity to stressor exposure (low "stressor reactivity" [SR]), were reported during the first wave of the COVID-19 pandemic in 2020.

Objective: Extending these findings, we here examined prospective relationships and weekly dynamics between the same RFs and SR in a longitudinal sample during the aftermath of the first wave in several European countries.

Methods: Over 5 weeks of app-based assessments, participants reported weekly stressor exposure, mental health problems, RFs, and demographic data in 1 of 6 different languages. As (partly) preregistered, hypotheses were tested cross-sectionally at baseline (N=558), and longitudinally (n=200), using mixed effects models and mediation analyses.

Results: RFs at baseline, including positive appraisal style (PAS), optimism (OPT), general self-efficacy (GSE), perceived good stress recovery (REC), and perceived social support (PSS), were negatively associated with SR scores, not only cross-sectionally (baseline SR scores; all *P*<.001) but also prospectively (average SR scores across subsequent weeks; positive appraisal (PA), *P*=.008; OPT, *P*<.001; GSE, *P*=.01; REC, *P*<.001; and PSS, *P*=.002). In both associations, PAS mediated the effects of PSS on SR (cross-sectionally: 95% CI –0.064 to –0.013; prospectively: 95% CI –0.074 to –0.0008). In the analyses of weekly RF-SR dynamics, the RFs PA of stressors generally and specifically related to the COVID-19 pandemic, and GSE were negatively associated with SR in a contemporaneous fashion (PA, *P*<.001; PAC,*P*=.03; and GSE, *P*<.001), but not in a lagged fashion (PA, *P*=.36; PAC, *P*=.52; and GSE, *P*=.06).

Conclusions: We identified psychological RFs that prospectively predict resilience and cofluctuate with weekly SR within individuals. These prospective results endorse that the previously reported RF-SR associations do not exclusively reflect mood congruency or other temporal bias effects. We further confirm the important role of PA in resilience.

(JMIR Ment Health 2023;10:e46518) doi:10.2196/46518

KEYWORDS

resilience; stressor reactivity; positive appraisal; pandemic; mental health; COVID-19

Introduction

Background

Outcome-based resilience refers to the maintenance or quick recovery of mental health despite exposure to adversity, presumably resulting from a dynamic process of adaptation [1]. While resilience has been primarily studied in the context of natural disasters, accidents, terror attacks, and other potentially traumatizing events [1-3], the outbreak of the COVID-19 pandemic in 2020 has brought up new types and levels of stressors that have impacted a vast majority of the global population. This is illustrated by the surge in stress-related mental disorders such as depression and anxiety during the pandemic [4]. In particular, people without mental health disorders before the pandemic exhibited significant increases in symptoms during the crisis compared with those who were already affected by a mental disorder [5]. Since 2021, the focus on the COVID-19 pandemic has shifted, and both media coverage and national policy responses have decreased substantially [6,7]. However, this study provides information on predictors, processes, and potential intervention targets for strategies to promote mental resilience, not only during the COVID-19 pandemic [1,8] but also in anticipation of increasingly frequent future global stressors [9].

Many studies worldwide have addressed questions of mental resilience during the COVID-19 pandemic via online surveys, conducted in China and other Asian countries [10-16], Iraq [17], Turkey [18,19], Israel [20], European countries [21-28], the United States [29-31], and Canada [32]. Increased levels of depressive symptoms and anxiety were frequently reported compared with population norms, while higher scores on trait resilience measures, behavioral coping (BC) strategies, and social support were cross-sectionally associated with lower symptoms of distress or better mental health. However, with the exception of our previous cross-sectional survey study "DynaCORE-C" (DynaMORE cross-sectional study on psychological resilience to the mental health consequences of the COVID-19 pandemic) [27], none of these studies considered individual-level stressor exposure, which is crucial for operationalizing resilience as the ability to maintain mental health despite exposure to such stressors [33,34].

In DynaCORE-C [27], we used a residualization approach [34-36], regressing internalizing mental health problems, retrospectively reported for a past 2-week time window, onto stressor exposure during that same time window. Using this method, individuals with a negative regression residual (a negative stressor reactivity [SR] score) can be seen as showing lower-than-expected symptom severity given their level of stressor exposure (ie, an indication of higher resilience), while individuals with positive residuals (positive SR score) show higher-than-expected stressor-related symptom severity (ie, an indication of lower resilience). This approach addresses the issue that individuals may well exhibit different degrees of mental health impairments in the COVID-19 pandemic; however, these differences may also be trivially explained by varying degrees of adversity experienced by the individuals rather than differences in their resilience capacities.

Positive Appraisal

Using this methodology, DynaCORE-C tested predictions put forth by the Positive Appraisal Style Theory of Resilience (PASTOR) [33]. According to PASTOR, individuals with a positive appraisal style (PAS) generally tend to set values for stressors, which they attribute to potential threats to their goals and needs, at levels that realistically reflect the threat. In some

cases, they may even slightly underestimate the threat on key appraisal dimensions such as threat magnitude or cost, threat probability, and coping potential. Positive appraisers typically avoid catastrophizing on the magnitude/cost dimension, pessimism on the probability dimension, and helplessness on the coping dimension. However, they also tend not to generate unrealistically positive (delusional) threat perceptions, which could lead to trivialization, blind optimism (OPT), or overconfidence. As a result, their average stress reactions tend to be optimally regulated, in the sense that positive appraisers are well adept at generating stress reactions when necessary while also avoiding the unnecessary expenditure of resources, such as overly strong, prolonged, or repeated stress responses. This gives them enough time for recovery, resource rebuilding, and exploration and limits deleterious allostatic load effects and resource depletion as much as possible.

The DynaCORE-C study found a self-report measure of PAS [27], along with the related constructs OPT and self-efficacy, to be positively associated with resilience (as approximated by a negative cross-sectional SR score). In addition to these measures of habitual appraisal styles, situational positive appraisal (PA), specifically related to the COVID-19 pandemic, was associated with resilience.

Another claim of PASTOR is that the effects of other social, biological, and psychological resilience factors (RFs) on outcome-based resilience are mediated by PAS, that is, RFs other than PAS are only beneficial for resilience to the extent that they shape someone's appraisal style toward the positive [33,37]. For instance, certain genetic or biological factors may render the brain circuits mediating PA and reappraisal processes more effective; spirituality may help find meaning in hardships; or trust in one's social networks may allow one to perceive many stressors as manageable.

In this regard, DynaCORE-C observed that the effects of perceived social support (PSS) were mediated by PAS [27]. Finally, DynaCORE-C found a weak cross-sectional association between BC style and resilience; additionally, it confirmed the well-known role of neuroticism (NEU) as a negative RF (ie, risk factor) [27].

These RF-SR associations from the cross-sectional DynaCORE-C study would be substantiated if one could show that (1) RFs also prospectively predict SR, ideally over an extended time window; and that (2) fluctuations in RFs are accompanied by fluctuations in SR, contemporaneously or prospectively (ie, with a time lag). Prospective associations, in particular, would help control for mood congruency or other state-dependent effects that may have exaggerated the previously reported cross-sectional associations [27].

Current Study

To achieve this, we conducted a longitudinal study (DynaCORE-L or DynaMORE longitudinal study on psychological resilience to the mental health consequences of the COVID-19 pandemic) with repeated weekly measures of above RFs and of stressor exposure and mental health (to repeatedly calculate SR) over 5 consecutive weeks (Figure 1).

With this approach, we addressed the following 5 sets of hypotheses (H):

First (H_1), we aimed to replicate the associations of RFs and SR found in DynaCORE-C [27] using the cross-sectional data assessed at baseline.

Second (H_2), we aimed to extend the cross-sectional DynaCORE-C findings [27] by exploring whether RFs at baseline prospectively predict resilience, as approximated by the average SR score over all follow-up time points.

Third (H_3), we investigated the relation between RFs and SR scores within individuals longitudinally across weekly time points, predicting contemporaneous cofluctuations.

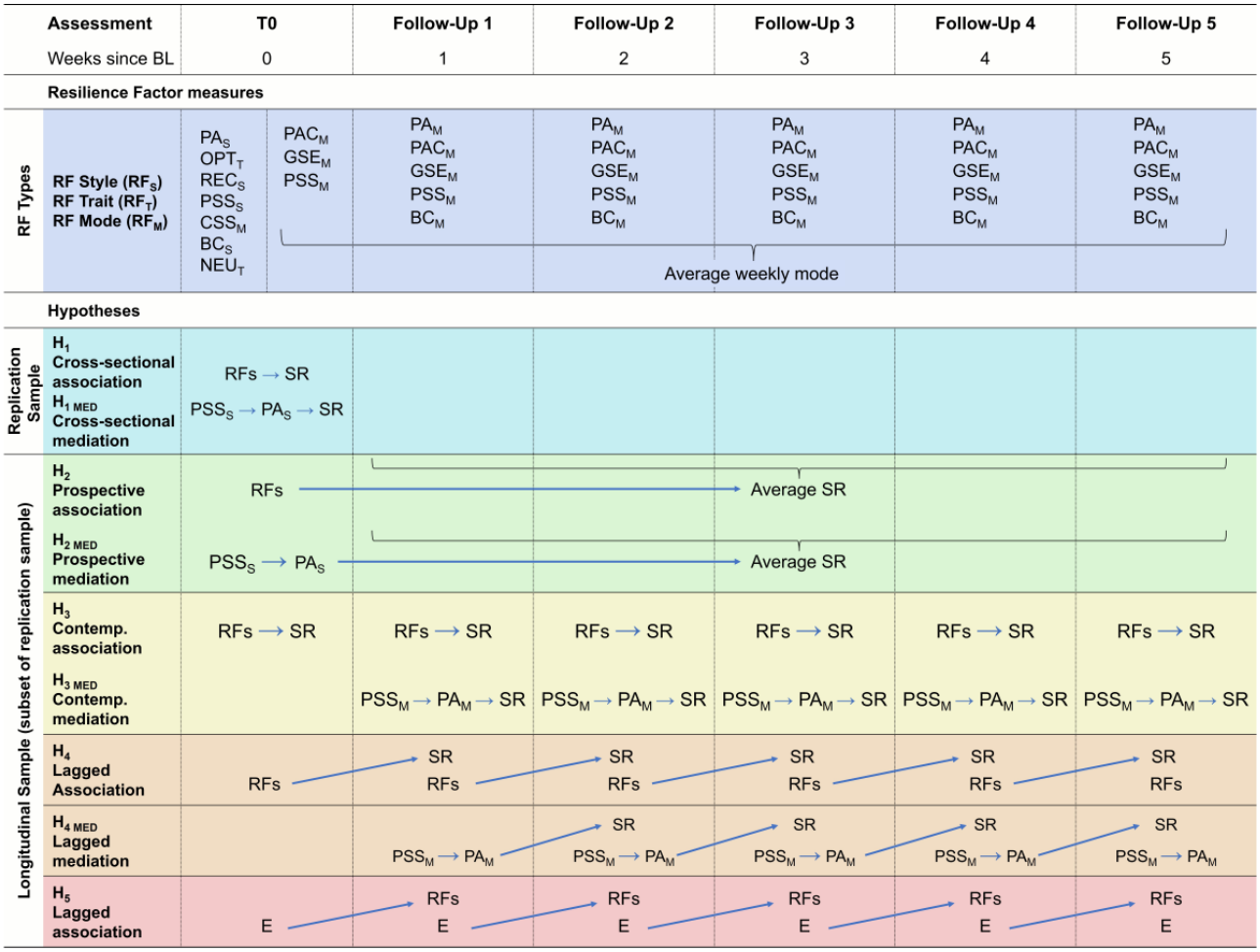
Fourth (H_4), in our primary hypothesis, we aimed to investigate the temporal dynamics of RFs and SR scores, namely, whether the use of RFs is prospectively associated with the SR score assessed 1 week later (lagged association).

For all analyses, we hypothesized negative associations between RFs and SR (except NEU). In line with PASTOR [33] and previous results [27], we further hypothesized that the statistical effect of PSS on SR is positively mediated by PA. The mediation hypothesis was tested for each type of association, that is, cross-sectional (H_{1_MED}), prospective (H_{2_MED}), contemporaneous (H_{3_MED}), and lagged (H_{4_MED}).

Fifth (H_5), and based on the consideration that the experience of stressors may compromise or, as in the phenomenon of stress inoculation [38-40], potentially also strengthen RFs, we longitudinally investigated stressor exposure–dependent fluctuations in the RFs measured in the subsequent week, hypothesizing that stressor exposure would be associated either negatively or positively with RFs in a time-lagged fashion.

All hypotheses, except H_2, were preregistered at the Center for Open Science (OSF) registries [41]. For simplification and better explanation of concepts, we changed the numbering of hypotheses relative to the preregistration.

Figure 1. Study design and hypotheses. To test the described hypotheses, the variables of interest were assessed at baseline (BL) and at 5 weekly follow-ups. The arrows indicate the hypothesized directions of statistical effects between the variables. At BL, resilience factors (RFs) were mostly assessed as general styles (subscript S) or traits (subscript T), while at the follow-ups, RFs were assessed as weekly modes (subscript M), that is, how frequent or extensively a certain RF was expressed during the preceding week. For each RF assessed as mode, an average weekly mode was also calculated, as the mean value across time points. Abbreviations: PA_S: positive appraisal style; OPT_T: optimism (trait); REC_S: perceived good stress recovery (style); PSS_S: perceived social support (style); CSS_M: perceived change in social support during the COVID-19 pandemic (mode); BC_S: behavioral coping style; NEU_T: neuroticism (trait); PAC_M: positive appraisal specifically of the COVID-19 pandemic (mode); GSE_M: general self-efficacy (mode); PSS_M: perceived social support (mode); PA_M: positive appraisal (mode); BC_M: behavioral coping (mode); SR: stressor reactivity.



Methods

Sample

Participants were recruited by snowball sampling via social media and mailing lists. The only inclusion criterion was a minimum age of 18 years. Data were collected across 6 time points per participant, comprising 1 baseline questionnaire and 5 weekly follow-ups (Figure 1). Data collection took place between April 17 and August 10, 2020.

Ethical Considerations

Participants were not financially reimbursed, but those who completed all assessments were included in a raffle to win an Amazon voucher worth €100 (US $90). Data collection was pseudonymous and informed consent was given electronically via the smartphone app m-Path [42]. The study was approved by the Ethics Committee of the State Medical Board of Rhineland-Palatinate, Mainz, Germany (2020-14967) and was conducted in accordance with the Declaration of Helsinki.

A total of 576 participants aged 18 years and above (mean age 31.7 years, SD 12.1 years, range 18-71 years, of which n=438 [76.1%] female) enrolled in the study, of which 210 participants (mean age 33.8 years, SD 13.3 years, range 18-68 years, of which n=160 [76.2%] female) completed at least four follow-up questionnaires. Follow-ups that were less than 5 or more than 9 days apart from the previous sampling time point were excluded from analysis, thus allowing a deviation of up to 2 days before and after the intended follow-up time point. Participants who answered less than 4 follow-up questionnaires or did not complete the baseline questionnaire were excluded from the longitudinal sample. We further excluded participants who reported demographic characteristics with exceptionally low frequencies compared with the rest of the sample, to prevent a statistically unreliable selection of covariates. After data cleaning, 558 participants were finally included in the cross-sectional (H_1) and 200 participants in the longitudinal (H_2-H_5) analyses.

Measured Variables

Overview

An overview of the items and inventories used for the measured variables is provided in Table S1 in Multimedia Appendix 1.

Demographic and Physical Health Variables

Demographic variables assessed at baseline included age and gender, as well as geographic, educational, and social variables. Health status variables were current or previous mental health diagnoses, as well as COVID-19 risk and infection status.

Resilience Factors

To be able to address the potentially dynamic associations of RFs with SR over time, RFs were assessed on 2 different timescales: typical characteristics (RF styles) and current modes (RF modes). At baseline, most RF questions asked about the participant's typical or usual behavior. They presumably reflect properties or qualities that are relatively durably associated with a person or constitute a typical way or tendency in which a person reacts to life experiences, but may still gradually change over time, for instance, through learning experiences and environmental changes. To demarcate these RFs from more trait-like RFs, we termed them "styles," in keeping with [33], and denoted them with the subscript S. Compared with traits, which are here denoted with the subscript T, styles are more likely to show adaptation over time and can be hypothesized as the basis for allostatic resilience processes [34]. Next to styles, at the weekly follow-ups, these same RFs were assessed as "modes" (denoted with the subscript M). With this new measurement approach, we assessed to what extent a particular RF was used or experienced in a given week. Complementary to RF style measures, RF mode measures may be more sensitive to changes in the strength of an RF, which would not become apparent from inquiring about typical or usual behavior. Thereby, repeated RF mode assessments allow for examining how an RF potentially is associated with SR in a shorter time frame.

RFs were PA [27,43,44], PA specifically of the COVID-19 pandemic (PAC) [27], OPT, general self-efficacy (GSE) [45], perceived good stress recovery (REC) [46], PSS [47], perceived change in social support during the COVID-19 pandemic (CSS) [27], and BC [27,43], complemented by NEU as a negative RF, or risk factor [48]. PA, PSS, and BC were assessed as both general styles (at baseline: PA_S, PSS_S, and BC_S, respectively) and weekly modes (at follow-ups: PA_M, PSS_M, and BC_M, respectively). PAC_M and GSE_M were assessed as weekly modes only (at both baseline and follow-ups). OPT_T, REC_S, CSS_M, and NEU_T were assessed as personality traits/general styles/weekly modes at baseline only (Figure 1).

Stressor Exposure

Participants reported the occurrence and severity of 11 general and 29 COVID-19 pandemic–specific stressors within the last 14 (baseline) or 7 days (follow-ups) on a 6-point scale ranging from 0 (did not happen) via 1 (not at all burdensome) to 5 (very burdensome). As in DynaCORE-C [27], E (ie, stressor exposure) was calculated as the total sum of all severity ratings.

Mental Health Problems

Internalizing symptoms were assessed for the past 14 days (baseline) or 7 days (follow-ups) using the 12-item General Health Questionnaire (GHQ-12) [49] total sum score.

Stressor Reactivity and Resilience

The SR score was computed as the residual of an individual's P score on the sample's E-P regression line [43]. E-P lines were fitted separately for the cross-sectional and longitudinal samples. For the cross-sectional analysis (H_1), the E-P regression line was fitted over all 558 participants who completed the baseline questionnaire (similar to DynaCORE-C [27]). For the longitudinal analyses (H_2-H_5), the E-P line was fitted over all 200 participants who were included in the longitudinal analysis and over all time points, using a mixed effects model with random slopes and intercepts for participants. To reduce bias in the SR score introduced by outliers, Mahalanobis distance [50] was used for outlier detection for the E-P distribution. Cases with a chi-square value corresponding to $P<.001$ were excluded from the analysis. The E-P regression line was then determined by the fixed effects estimates of the slope and intercept, providing an estimate of normative SR in the sample over the whole observation period. Adding a second-order polynomial term did not improve model fit either in the cross-sectional or in the longitudinal sample ($F_{1,555}=3.35$, $P=.07$ and $\chi_1^2=0.88$, $P=.35$, respectively, when comparing the model fit with and without the polynomial term). Subsequently, individual SR scores per time point were determined as residuals of individual P scores on the linear E-P line, by entering participants' P and E scores from the respective week into the normative E-P line equation. SR scores were calculated separately for the cross-sectional and longitudinal samples.

Covariate Selection

In all models, age, gender, and survey language were included as covariates. Further covariates were selected based on their estimated effect on SR, which was assessed using univariate regression analyses separately for the cross-sectional and longitudinal samples. Variables surviving a likelihood ratio test at $P<.2$ were included in statistical analyses. The key covariates selected in both samples were education, general health, previous or current mental health diagnosis, belonging to a risk group, and opinion about the authorities' measures to curtail the spread of the virus (for further details, see section 1.2 in Multimedia Appendix 1).

Statistical Analyses

The cross-sectional sample (N=558) was used to replicate the multiple regression and mediation results from the DynaCORE-C study [27] (H_1, H_{1_MED}) using the same analysis procedure (see section 1.3.1 in Multimedia Appendix 1). Separate multiple regression analyses were performed to assess the effects of each baseline RF style on the baseline SR score. Each model included the selected covariates (see Table S2 in Multimedia Appendix 1). Mediation analyses were conducted following the Baron and Kenny approach [51] and indirect paths were determined with the distribution-of-the-product method.

The prospective association between baseline RF styles and the average weekly SR score (H_2) as well as the corresponding mediation (H_{2_MED}) was calculated analogously, yet in the longitudinal sample (n=200).

All dynamic hypotheses (H_3-H_5) were tested in the longitudinal sample (n=200) by linear mixed model analyses (see sections 1.3.3-1.3.5 in Multimedia Appendix 1), using the lme4 package [52] in R (version 4.0.4; R Core Team). Each model included the selected covariates (see Table S3 in Multimedia Appendix 1) as well as the participant-level mean of the independent variable (for details, see section 1.2 in Multimedia Appendix 1). Random intercepts were assumed for each participant, and random slopes were fitted for the demeaned time-varying independent variable. To test model assumptions, visual checks of residual distributions were performed (see section 2.4.5 in Multimedia Appendix 1).

As preregistered and for consistency, an α level of P<.05, 2-tailed, was used for all analyses, including the directional tests. To correct for multiple testing, a Bonferroni correction was applied to the analyses addressing our primary hypotheses about the time-lagged effects of PA (H_4: PA_M and PAC_M). These hypotheses were considered significant when passing the adjusted α level (P_{corr}) <(.05/2)=.025. All reported β estimates are standardized.

Any time-lagged model that revealed significant associations was followed up with an analysis of the association between the independent variable and the change in the dependent variable. For example, in the hypothetical association between any RF (time t) and the lagged SR (t+1), the SR at time t would be added as an additional predictor to the model. To this end, the model would account for the variance shared with the previous measurement (t) of the dependent variable.

Finally, all analyses were repeated for participants in the top 2 tertiles (368/558, 65.9%, for the cross-sectional sample and 132/200, 66%, for the longitudinal sample) of stressor exposure (mean E counts over the observation period), to make sure that our results also apply when excluding participants with low stressor exposure.

Results

Sample Characteristics

Demographic characteristics of the cross-sectional baseline sample after exclusions (N=558) are provided in Table 1 and Tables S4-S6 in Multimedia Appendix 1. In this sample used for the cross-sectional replication analyses (hypothesis H_1), the most frequently reported stressors were COVID-19–related media coverage (547/558, 98%), not being able to carry out leisure activities (535/558, 95.8%), and loss of social contact (522/558, 93.5%). On average, the most severely rated stressors were the inability to attend the funeral of a loved one (mean severity 4.03), the death of a loved one (mean severity 3.87), and the inability to return to the country one lives (mean severity 3.65). See Table S7 in Multimedia Appendix 1 for frequencies and severity ratings of all stressors. Participants who reported a past or present psychiatric diagnosis had significantly higher SR scores (mean 0.29, SD 1.01) than those who did not (mean –0.15, SD 0.97, t_{556}=–4.92; P<.001).

Baseline characteristics of the longitudinal sample (n=200) used for all other analyses (hypotheses H_2-H_5) are provided in Table 2. Baseline characteristics of the baseline and longitudinal samples demonstrate notable similarities. Further details, including E, P, SR, and RFs per time point, are given in Tables S8-S10 in Multimedia Appendix 1. Frequencies and severity ratings of all stressors are provided in Table S11 in Multimedia Appendix 1.

Table 1. Characteristics of the cross-sectional sample, assessed at baseline.[a]

Characteristics	Values (N=558)
Gender, n (%)	
Male	128 (22.9)
Female	430 (77.1)
Age (years)	
Mean (SD)	31.6 (12.1)
Median (range)	27.0 (18.0-71.0)
Response language, n (%)	
Dutch	84 (15.1)
English	21 (3.8)
German	362 (64.9)
Hebrew	0 (0)
Italian	38 (6.8)
Polish	53 (9.5)
Education (years)	
Mean (SD)	17.5 (3.31)
Median (range)	17.0 (8.00-33.0)
Missing, n (%)	88 (15.8)
Relationship status, n (%)	
Married, in a domestic partnership, or in a civil union	120 (21.5)
In a steady relationship	211 (37.8)
Widowed	3 (0.5)
Divorced or separated	18 (3.2)
Single	196 (35.1)
Other	10 (1.8)
Good general health (self-report, 1-5)	
Mean (SD)	2.47 (0.991)
Median (range)	2.00 (1.00-5.00)
Diagnosed mental health condition (ever), n (%)	
No	374 (67.0)
Yes	184 (33.0)
Belong to a risk group, n (%)	
No	435 (78.0)
Yes	50 (9.0)
Not sure	73 (13.1)
Agreement with authorities' measures (self-report, 1-5)	
Mean (SD)	4.00 (0.985)
Median (range)	4.00 (1.00-5.00)

[a]The sample was used for the cross-sectional replication analyses (hypothesis H_1).

Table 2. Characteristics of the longitudinal sample, assessed at baseline.[a]

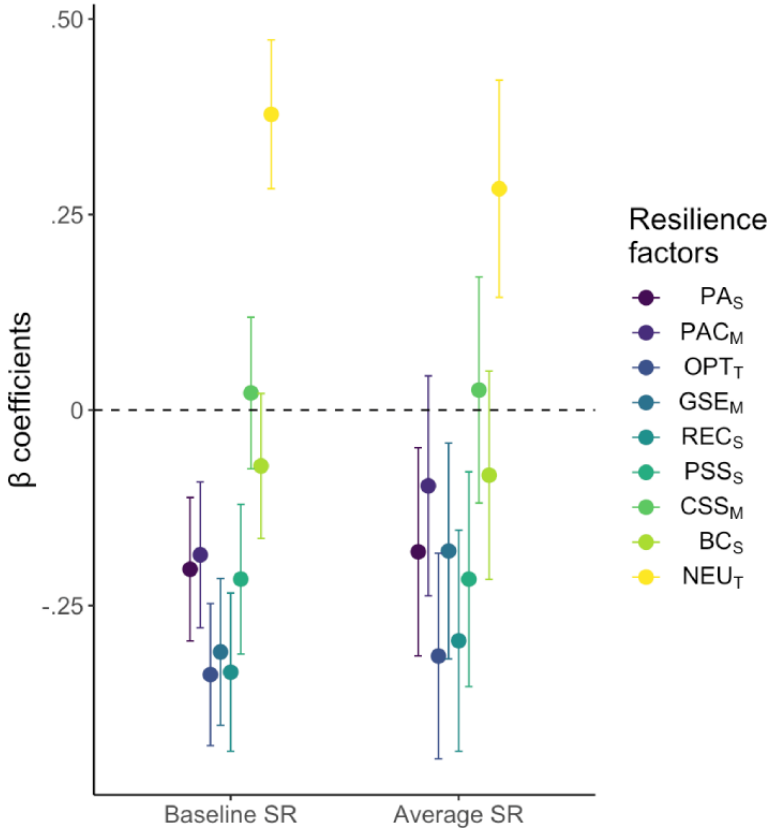
Characteristics	Values (N=200)
Gender, n (%)	
Male	43 (21.5)
Female	157 (78.5)
Age (years)	
Mean (SD)	33.9 (13.3)
Median (range)	28.0 (18.0-68.0)
Response language, n (%)	
Dutch	29 (14.5)
English	9 (4.5)
German	122 (61.0)
Hebrew	0 (0)
Italian	17 (8.5)
Polish	23 (11.5)
Education (years)	
Mean (SD)	17.9 (3.20)
Median (range)	18.0 (12.0-30.0)
Missing, n (%)	26 (13.0)
Relationship status, n (%)	
Married, in a domestic partnership or civil union	43 (21.5)
In a steady relationship	71 (35.5)
Widowed	3 (1.5)
Divorced or separated	10 (5.0)
Single	69 (34.5)
Other	4 (2.0)
Good general health (self-report, 1-5)	
Mean (SD)	2.42 (0.973)
Median (range)	2.00 (1.00-5.00)
Diagnosed mental health condition (ever), n (%)	
No	132 (66.0)
Yes	68 (34.0)
Belong to a risk group, n (%)	
No	152 (76.0)
Yes	20 (10.0)
Not sure	28 (14.0)
Agreement with authorities' measures (self-report, 1-5)	
Mean (SD)	4.04 (0.994)
Median (range)	4.00 (1.00-5.00)

[a]This sample is a subset of the cross-sectional sample and was used for longitudinal analyses (H_2-H_5).

Cross-Sectional RF-SR Associations and Mediations (H1 and H1_MED)

After controlling for the selected covariates, directed hypotheses were confirmed for most baseline RFs (negative cross-sectional association with baseline SR: PA_S, PAC_M, OPT_T, GSE_M, REC_S, PSS_S; positive association: NEU_T; all $P<.001$), but not for BC_S ($P=.13$) or CSS_M ($P=.66$; see Figure 2 and Table S12 in Multimedia Appendix 1). Additionally, we replicated the reported positive mediation of the effect of PSS_S on SR by PA_S [27] (indirect effect estimate: –0.035, 95% CI –0.064 to –0.013) at baseline.

Figure 2. Association of baseline resilience factors (RFs) with stressor reactivity (SR). The β coefficients show associations from separate multiple regression analyses for each baseline RF with baseline SR (testing hypothesis H1: cross-sectional association) and average SR across the weekly longitudinal follow-ups (testing H2: prospective prediction). Negative associations suggest that factors contribute to dampening mental health reactivity to stressor exposure, that is, they promote resilience. Error bars depict 95% CIs. Abbreviations: PA_S: positive appraisal style; PAC_M: positive appraisal specifically of the COVID-19 pandemic (mode); OPT_T: optimism (trait); GSE_M: general self-efficacy (mode); REC_S: perceived good stress recovery (style); PSS_S: perceived social support (style); CSS_M: perceived change in social support during the COVID-19 pandemic (mode); BC_S: behavioral coping style; NEU_T: neuroticism (trait).



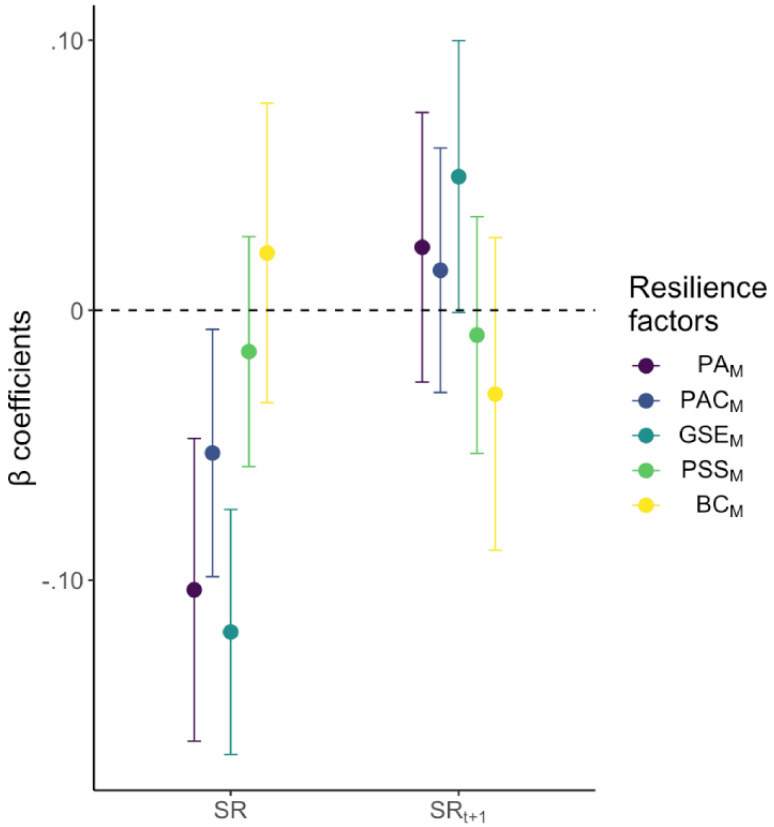
Prospective RF-SR Associations and Mediations (H2 and H2_MED)

After controlling for the selected covariates, prospective associations were found for most baseline RFs (negative: PA_S, $P=.008$; OPT_T, $P<.001$; GSE_M, $P=.01$; REC_S, $P<.001$; PSS_S, $P=.002$; positive: NEU_T, $P<.001$) with the average SR score across longitudinal follow-ups used as an approximation of outcome-based resilience in an extended time frame. No associations were found for PAC_M ($P=.18$), CSS_M ($P=.73$), and BC_S ($P=.22$; see Figure 2 and Table S13 in Multimedia Appendix 1). The mediation analyses suggest that baseline PA_S mediated the relationship between baseline PSS_S and future average SR (indirect effect estimate: –0.030, 95% CI –0.074 to –0.0008).

Contemporaneous RF-SR Associations and Mediations (H3 and H3_MED)

After controlling for the selected covariates and the respective average weekly RF mode (ie, participant-level mean), negative contemporaneous associations were found in the longitudinal data between the demeaned RFs (ie, within-participant mean centered; PA_M, $P<.001$; PAC_M, $P=.03$; and GSE_M, $P<.001$; see Figure 3 and Table S14 in Multimedia Appendix 1) and demeaned SR measured at the same weekly time points. No associations with SR were found for PSS_M ($P=.48$) and BC_M ($P=.45$). The hypothesized mediation effect of PSS_M on SR by PA_M was supported (indirect effect estimate –0.010, 95% CI –0.0225 to –0.0004).

Figure 3. Association of weekly measured resilience factors (RFs) with weekly stressor reactivity (SR). The β coefficients show associations from separate mixed effects analyses for each weekly measured RF mode with SR measured at the same time points (testing hypothesis H3: contemporaneous association) and SR measured 1 week later (t+1) in the longitudinal follow-ups (testing H4: lagged RF-SR association). Error bars depict 95% CIs. Abbreviations: PA_M: positive appraisal (mode); PAC_M: positive appraisal specifically of the COVID-19 pandemic (mode); GSE_M: general self-efficacy (mode); PSS_M: perceived social support (mode); BC_M: behavioral coping (mode).



In addition to the above described within-participant relationships, several between-participant relationships for the average weekly RF modes and SR were observed in the same models. Negative associations with SR were found for mean PA_M (P=.02), PAC_M (P=.03), GSE_M (P<.001), and PSS_M (P<.001; see Table S14 in Multimedia Appendix 1). No associations were found for mean BC_M (P=.78).

Intraclass correlation coefficients for each RF mode ranged between 0.64 and 0.85, indicating low within-participant variance across all time points. In line with the earlier reported cross-sectional relationships of RF styles and SR at baseline and the prospective prediction of average weekly SR by baseline RF styles, these results indicate that SR is negatively associated with relatively stable components within the RFs, even when they are assessed in a weekly mode format.

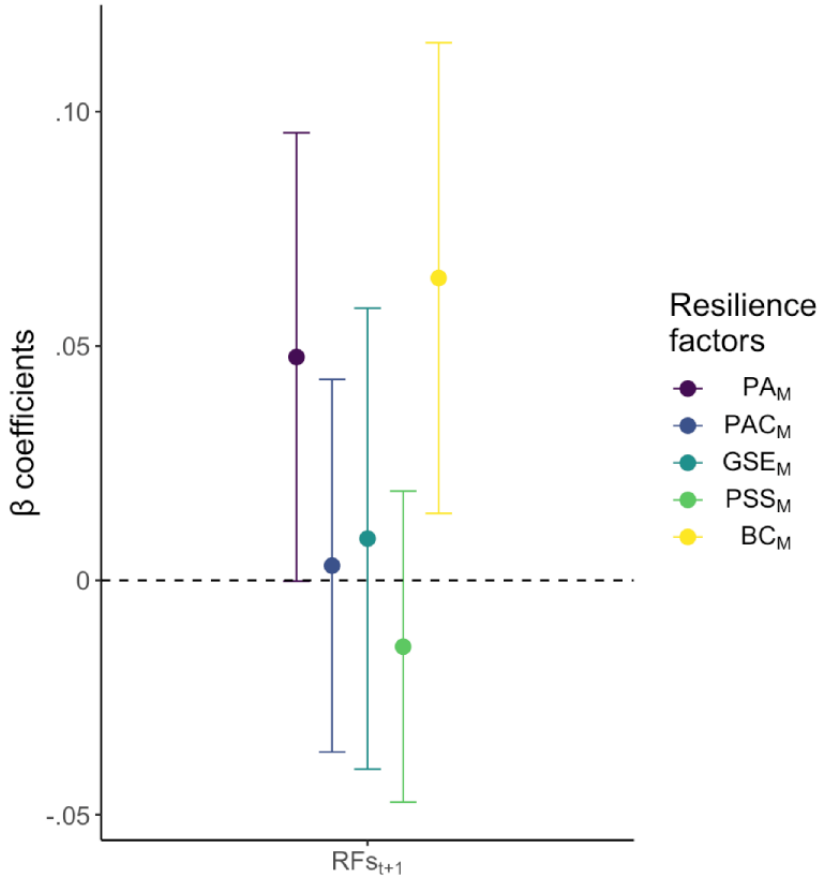
Lagged RF-SR Associations and Mediations (H4 and H4_MED)

None of the demeaned weekly RF modes were associated with demeaned SR 1 week later in the time-lagged analyses ($PA_{M, t-1}$, P=.36; $PAC_{M, t-1}$, P=.52; $GSE_{M, t-1}$, P=.06; $PSS_{M, t-1}$, P=.68; $BC_{M, t-1}$, P=.30; see Figure 3 and Table S15 in Multimedia Appendix 1). In addition, mediation analyses did not support the hypothesized effect of PSS_M on SR by PA_M (indirect effect estimate: –0.004, 95% CI –0.0024 to 0.0127).

Lagged E-RF Associations (H5)

After controlling for covariates and the average weekly stressor exposure E, a positive association was found for demeaned weekly stressor exposure (time t) and BC mode (BC_M) 1 week later (t+1; P=.01; see Figure 4 and Table S16 in Multimedia Appendix 1), suggesting that high stressor severity might have led to more active coping behavior. However, the association did not remain significant when adding the BC_M variable at time t to the model to investigate how BC_M evolves (P=.26; see Table S17 in Multimedia Appendix 1). This indicates that although stressor exposure is associated with BC 1 week later, this lagged association does not survive when considering autoregressive trends in BC. We therefore refrain from discussing this result further.

Figure 4. Association of weekly measured stressor exposure (E) with weekly lagged resilience factors (RFs). The β coefficients show associations from separate mixed effects analyses for each weekly measured E with RFs measured 1 week later (t+1) in the longitudinal follow-ups (testing H5: Lagged E-RF associations). Error bars depict 95% CIs. Abbreviations: PA_M: positive appraisal (mode); PAC_M: positive appraisal specifically of the COVID-19 pandemic (mode); GSE_M: general self-efficacy (mode); PSS_M: perceived social support (mode); BC_M: behavioral coping (mode).



No associations were found for demeaned weekly stressor exposure (time t) and the other RFs 1 week later (t+1; PA_M, P=.05; PAC_M, P=.88; GSE_M, P=.72; PSS_M, P=.41).

Repeating all analyses with the 66% of participants (368/558, 65.9%, for the cross-sectional sample and 132/200, 66%, for the longitudinal sample) with the highest degree of stressor exposure (mean E over the whole time) revealed the same pattern of results (see section 2.5 in Multimedia Appendix 1). This indicates that our observations also hold when a more stringent criterion for the presence of adversity is applied [34].

Discussion

Principal Findings

In this study, we investigated the dynamic relationships of RFs and SR as an indicator of outcome-based resilience measured weekly over 5 weeks during the COVID-19 pandemic. We find that RFs prospectively predict SR across a time frame of weeks and cofluctuate with SR contemporaneously, while week-to-week within-participant changes in the strength of RFs are not followed by lagged changes in SR scores. RFs exhibit pronounced temporal stability across weeks, such that a single baseline RF assessment of RF style sufficiently captures the between-participant RF variance that is relevant for an individual's future SR in the time frame tested here. We confirm an important role of PA in resilience to the pandemic, in line with the previous proposals made in the DynaCORE-C study [27].

The Temporal Relationship of Stressors, Resilience Factors, and Stressor Reactivity

The extant cross-sectional findings were replicated here for 7 out of 9 RFs (H_1). For BC style (BC_S), the estimated effect was similar in size and direction as in DynaCORE-C [27] but was only marginally significant, most likely due to the smaller sample size. For CSS_M, the apparent absence of an association with the SR score may as well be linked to the time of data collection during the pandemic. While the DynaCORE-C study [27] ran between March and April 2020, a phase with wider restrictions in most countries, the present data set was acquired between April and August of that year, when restrictions were partially lifted and COVID-19 pandemic–specific changes in social support might have become less pronounced. Notwithstanding these limitations, the presence of significant negative associations of most RF styles with the SR score at baseline even in a sample almost 30 times smaller (DynaCORE-C [27]: N=15,970) further supports that these RFs may serve as protective factors against stress-related mental health problems in pandemics and comparable crises.

Being cross-sectional, these replication analyses cannot rule out mood congruency or other state-dependent effects, and they only employ a snapshot measure of SR across the past 2 weeks.

In H_2, we investigated prospectively predictive associations of baseline RF styles with the average SR score across the 5 weeks of follow-ups. This time frame allows for detecting clinically relevant changes in mental health (most affective disorders have a time criterion of 2-4 weeks) and for relating them to temporally more extended stressor exposure, making the average SR score a more valid and robust approximation of resilience compared with the baseline snapshot. We observed significant negative associations for 6 out of 9 RF styles at baseline with the average SR score (PAC_M, CSS_M, and BC_S were not significant). All β estimates are fully standardized. A β of 0.25 for NEU (Figure 2), for example, thus means that scoring on baseline NEU 1 SD above the sample mean is associated with an increase in average SR by a quarter of an SD. These results not only support the validity of the hypothesized RFs but are also the first indication of their relative temporal stability, without which they would not be able to influence SR weeks later.

Further extending this, our contemporaneous within-participant analyses of the longitudinal data (H_3) revealed negative associations between weekly RF modes and SR scores for PA_M, PAC_M, and GSE_M, but not for PSS_M and BC_M. Fluctuations in weekly modes of the former RFs are thus related to fluctuations of SR within individuals, suggesting that these RF measures capture a state-like element that is associated with SR on a weekly basis. Nevertheless, all RF modes showed intraclass correlation coefficient values of 0.65 and higher, indicating less within- compared with between-participant variance. Moreover, participants' participant-level mean RF scores across time points were meaningfully related to their mean SR scores, indicating an important stable component, in addition to the fluctuating component expressed in the demeaned scores that were used for testing the contemporaneous effects.

Of note, the within-participant contemporaneous effects can still be explained based on mood congruency or other state dependency. Excluding such potential explanations motivated the analysis of time-lagged effects of RF modes on the SR score (H_4), to investigate whether changes in RF modes might prospectively predict changes in SR in the next week. This was not confirmed. A probable reason for this could be the timescale we used. Although ecological momentary assessment studies detected lagged relationships on timescales in the magnitude of several hours [53,54], prospective associations between psychological styles or traits and mental health measures are typically observed within months or years [55]. In combination with the aforementioned relative temporal stability of RFs during our study time window, this suggests that future studies should examine longer time intervals to detect meaningful adaptations in RFs and consequential effects on resilience outcomes.

Hypothesis 5 (H_5) sought to explore whether prior changes in stressor exposure could lead to potential adaptations in response to risk factors (RF). This inquiry was inspired by observations of stress inoculation effects in lifetime studies, where earlier encounters with moderate adversity are statistically linked to reduced stress reactivity and improved long-term mental health and psychosocial well-being [39,40]. No such effects could be observed on the timescale of this study, further emphasizing the need for temporally more extended observations.

Positive Appraisal

A major objective of this study was to test PASTOR. According to PASTOR, individuals exhibiting a PAS tend to appraise threats in a way that they assess threat magnitude or cost, threat probability, and their own coping potential realistically or slightly unrealistically positively. Thus, PAS encompasses constructs such as OPT (probability dimension) and GSE (coping dimension). However, it takes a broader perspective, acknowledging that, for instance, a person's habitual pessimism may be offset by their low catastrophizing tendencies or strong self-confidence, leading to potential positive effects. Insofar, OPT and GSE questionnaires, for instance, may be helpful but are potentially not sufficient for assessing PAS. In DynaCORE-C, the PAS questionnaire adopted a different approach to assess appraisal tendencies within the threat appraisal dimensions of PASTOR. Rather than measuring negative aspects, it aimed to gauge the cognitive processes or mental operations that individuals habitually employ in stressful situations to generate PA content [27]. Notably, the processes addressed by the scale include variations of cognitive positive reappraisal, such as trying to find positive aspects or potential good outcomes of a situation, to put the situation into perspective, to accept it, or to detach from it. This version of the PAS questionnaire is an early version of the Perceived Positive Appraisal Style Scale—process-focused (PASS-process) that was developed for the purpose of large-scale surveys during the COVID-19 pandemic and has since been optimized and validated in additional studies [56].

Both the original style variant of this questionnaire, as employed in the DynaCORE-C survey (PA_S in our paper), and our adaptation of it to a weekly mode assessment (PA_M) were consistently negatively associated with SR in this study, except in the time-lagged analysis (H_4). Further, our mediation analyses showed that PA_S positively mediated the association between PSS_S and SR in a cross-sectional (H_{1_MED}), prospective (H_{2_MED}), and contemporaneous fashion (H_3) [57]. Together with a recent study of high PA in individuals with favorable mental health reactions to the COVID-19 pandemic [57], these results can be taken as further support for PASTOR, at least within the pandemic context.

On a practical level, the employed PAS instrument was not superior to the OPT and GSE instruments in explaining variance in SR, as was already observed in the DynaCORE-C study [27]. This may be related to the challenges associated with self-reporting the utilization of mental processes, specifically the cognitive operations involved in positively appraising and reappraising challenging situations, as targeted by the instrument. It is often easier for individuals to report on the final appraisal outcomes generated by these processes, as seen in OPT and GSE, rather than the processes themselves. This may apply in particular when considering that many of the processes that produce PAs may occur at a nonconscious level [33]. We, therefore, propose that future work may consider simply relying on existing OPT and GSE instruments for assessing PA tendencies or also combine these dimensions into a single scale for PAS that focuses on appraisal contents rather than on generating processes. For the latter, see current work on the

development of the Perceived Positive Appraisal Style Scale—content-focused (PASS-content) [56].

Limitations

Although a strength of our study is that our resilience measure incorporated stressor exposure, thereby operationalizing resilience as an outcome of good mental health despite adversity, this approach requires modeling a sample-level E-P relationship, which for the longitudinal analyses was based on a relatively small convenience sample (n=200) across 6 time points only. The generalizability of our results may also be limited by biased demographics that resulted from the self-selection effect in our snowball-system recruitment approach. A gender bias with higher female participation rates is commonly observed in surveys and study trials [58-60]. According to the Social Exchange Theory, this could be attributed to a proposed inherent gender difference in social exchange decisions, with females purportedly placing a higher value on connective characteristics, and males supposedly preferring separative characteristics [59]. Another explanation might be that females are more likely to be interested in the topic of our survey because they are more often affected by stress-related mental disorders [61-63]. As participants were recruited through snowball sampling via social media and mailing lists, we did not have an influence on the gender balance in our sample, but we statistically controlled for the effects of gender on the SR score, making the results more generalizable across genders. However, to enhance personalized and gender-specific treatment strategies, gender differences in SR and its predictors should be investigated further.

General limitations of questionnaire studies must also be pointed out, which—in addition to being insensitive to mental operations and contents that are inaccessible to consciousness and not verbalizable—can suffer from subjectively biased and socially desirable reporting and issues related to semantic ambiguity of questions. It may therefore be useful to complement questionnaire measures of constructs such as PA or mental health with the more objective task–based or biological measures [33].

Conclusions

To conclude, we identified RFs—predictors of low SR as an indicator for resilience—during a global pandemic. These RFs can potentially be targeted to prevent negative mental health consequences of future pandemics and similar adverse events. Scales assessing facets of PA tendencies were among the most important RFs. Crucially, our results suggest that relationships between RFs and outcome-based resilience also exist not just as individual differences but also within participants. The finding that fluctuating components of RFs, termed RF modes, relate to concurrent differences in resilience may be of particular interest for interventions seeking immediate impact. More research on potential causality is, however, needed.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement Numbers 777084 (DynaMORE project) and 101016127 (RESPOND), Deutsche Forschungsgemeinschaft (DFG Grant CRC 1193, subprojects B01, C01, C04, Z03), BMBF NUM EViPan 01KX2021, and the State of Rhineland-Palatinate, Germany (MARP program, DRZ program, Leibniz Institute for Resilience Research). AR was supported by Studienstiftung des deutschen Volkes at the time of study conduct. We thank J. Timmer for critical comments; K. Berrevoets, H. Burger, A. Dimitrov, C. Rohr, T. Steppat, S. von Paleske, and E. Zimmermann for their help with study conduct; and N. Röthke for help with literature research.

Data Availability

The preregistration of the analysis plan is available at the OSF registries [41]. The survey questionnaire is available upon request. The anonymous data (raw and preprocessed) are available at OSF [64]. The code for data cleaning and analysis is available at OSF [65].

Authors' Contributions

SAB, LMCP, CW, MZ, AR, GK, KSLY, SP, JW, RvD, BK, KL, JMCvL, DK, IM-G, HW, OT, EJH, IMV, and RK conceptualized this study. SAB, LMCP, and MZ curated the data analyzed in this study. SAB, LMCP, MZ, SP, and GK performed formal analysis. AR, BK, KL, DK, IM-G, HW, OT, and RK were responsible for funding acquisition. SAB, LMCP, CW, MZ, AR, MAM, ZR, AU, DBL, JM, NR, and NCD performed the study investigation. SAB and JMCvL were responsible for project administration. MM and SV performed software analysis. SAB, LMCP, CW, MZ, and RK performed writing—original draft. All authors performed writing—review and editing.

Conflicts of Interest

RK has received advisory honoraria from JoyVentures, Herzlia, and Israel.

Multimedia Appendix 1
Supplemental information.
[DOC File , 977 KB - mental_v10i1e46518_app1.doc]

References

1. Kalisch R, Baker DG, Basten U, Boks MP, Bonanno GA, Brummelman E, et al. The resilience framework as a strategy to combat stress-related disorders. Nat Hum Behav 2017 Nov;1(11):784-790. [doi: 10.1038/s41562-017-0200-8] [Medline: 31024125]
2. Bonanno G, Romero S, Klein S. The temporal elements of psychological resilience: an integrative framework for the study of individuals, families, and communities. Psychological Inquiry 2015 Jun 11;26(2):139-169. [doi: 10.1080/1047840x.2015.992677]
3. Mancini AD, Bonanno GA. Predictors and parameters of resilience to loss: toward an individual differences model. J Pers 2009 Dec;77(6):1805-1832 [FREE Full text] [doi: 10.1111/j.1467-6494.2009.00601.x] [Medline: 19807863]
4. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psychiatr 2020;33(2):e100213 [FREE Full text] [doi: 10.1136/gpsych-2020-100213] [Medline: 32215365]
5. Pan K, Kok AAL, Eikelenboom M, Horsfall M, Jörg F, Luteijn RA, et al. The mental health impact of the COVID-19 pandemic on people with and without depressive, anxiety, or obsessive-compulsive disorders: a longitudinal study of three Dutch case-control cohorts. The Lancet Psychiatry 2021 Feb;8(2):121-129. [doi: 10.1016/s2215-0366(20)30491-0]
6. Pearman O, Boykoff M, Osborne-Gowey J, Aoyagi M, Ballantyne AG, Chandler P, et al. COVID-19 media coverage decreasing despite deepening crisis. Lancet Planet Health 2021 Jan;5(1):e6-e7 [FREE Full text] [doi: 10.1016/S2542-5196(20)30303-X] [Medline: 33421410]
7. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 2021 Apr 08;5(4):529-538. [doi: 10.1038/s41562-021-01079-8] [Medline: 33686204]
8. Vinkers CH, van Amelsvoort T, Bisson JI, Branchi I, Cryan JF, Domschke K, et al. Stress resilience during the coronavirus pandemic. Eur Neuropsychopharmacol 2020 Jun;35:12-16 [FREE Full text] [doi: 10.1016/j.euroneuro.2020.05.003] [Medline: 32446705]
9. Schäfer SK, Kunzler AM, Kalisch R, Tüscher O, Lieb K. Trajectories of resilience and mental distress to global major disruptions. Trends Cogn Sci 2022 Dec;26(12):1171-1189 [FREE Full text] [doi: 10.1016/j.tics.2022.09.017] [Medline: 36302711]
10. Hou W, Tong H, Liang L, Li T, Liu H, Ben-Ezra M, et al. Probable anxiety and components of psychological resilience amid COVID-19: a population-based study. J Affect Disord 2021 Mar 01;282:594-601 [FREE Full text] [doi: 10.1016/j.jad.2020.12.127] [Medline: 33445081]
11. Lahiri A, Jha SS, Acharya R, Dey A, Chakraborty A. Has loneliness and poor resilient coping influenced the magnitude of psychological distress among apparently healthy Indian adults during the lockdown? Evidence from a rapid online nation-wide cross-sectional survey. PLoS One 2021;16(1):e0245509 [FREE Full text] [doi: 10.1371/journal.pone.0245509] [Medline: 33444423]
12. Leng M, Wei L, Shi X, Cao G, Wei Y, Xu H, et al. Mental distress and influencing factors in nurses caring for patients with COVID-19. Nurs Crit Care 2021 Mar;26(2):94-101. [doi: 10.1111/nicc.12528] [Medline: 33448567]
13. Li F, Luo S, Mu W, Li Y, Ye L, Zheng X, et al. Effects of sources of social support and resilience on the mental health of different age groups during the COVID-19 pandemic. BMC Psychiatry 2021 Jan 07;21(1):16 [FREE Full text] [doi: 10.1186/s12888-020-03012-1] [Medline: 33413238]
14. Ou X, Chen Y, Liang Z, Wen S, Li S, Chen Y. Resilience of nurses in isolation wards during the COVID 19 pandemic: a cross-sectional study. Psychol Health Med 2021 Jan;26(1):98-106. [doi: 10.1080/13548506.2020.1861312] [Medline: 33305600]
15. Setiawati Y, Wahyuhadi J, Joestandari F, Maramis MM, Atika A. Anxiety and resilience of healthcare workers during COVID-19 pandemic in Indonesia. J Multidiscip Healthc 2021;14:1-8 [FREE Full text] [doi: 10.2147/JMDH.S276655] [Medline: 33442258]
16. Xiao X, Xiao J, Yao J, Chen Y, Saligan L, Reynolds NR, et al. The role of resilience and gender in relation to infectious-disease-specific health literacy and anxiety during the COVID-19 pandemic. Neuropsychiatr Dis Treat 2020;16:3011-3021 [FREE Full text] [doi: 10.2147/NDT.S277231] [Medline: 33324062]
17. Yıldırım M, Arslan G, Aziz IA. Why do people high in COVID-19 worry have more mental health disorders? The roles of resilience and meaning in life. Psychiatr Danub 2020;32(3-4):505-512 [FREE Full text] [doi: 10.24869/psyd.2020.505] [Medline: 33370760]
18. Satici S, Kayis A, Satici B, Griffiths M, Can G. Resilience, hope, and subjective happiness among the Turkish population: fear of COVID-19 as a mediator. Int J Ment Health Addict 2023;21(2):803-818 [FREE Full text] [doi: 10.1007/s11469-020-00443-5] [Medline: 33293904]
19. Yıldırım M, Arslan G, Özaslan A. Perceived risk and mental health problems among healthcare professionals during COVID-19 pandemic: exploring the mediating effects of resilience and coronavirus fear. Int J Ment Health Addict 2022;20(2):1035-1045 [FREE Full text] [doi: 10.1007/s11469-020-00424-8] [Medline: 33223977]

20. Levaot Y, Greene T, Palgi Y. The associations between media use, peritraumatic distress, anxiety and resilience during the COVID-19 pandemic. J Psychiatr Res 2022 Jan;145:334-338 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.11.018] [Medline: 33187689]
21. Anyan F, Hjemdal O, Ernstsen L, Havnen A. Change in physical activity during the coronavirus disease 2019 lockdown in Norway: the buffering effect of resilience on mental health. Front Psychol 2020;11:598481 [FREE Full text] [doi: 10.3389/fpsyg.2020.598481] [Medline: 33384645]
22. Cusinato M, Iannattone S, Spoto A, Poli M, Moretti C, Gatta M, et al. Stress, resilience, and well-being in Italian children and their parents during the COVID-19 pandemic. Int J Environ Res Public Health 2020 Nov 10;17(22):1-17 [FREE Full text] [doi: 10.3390/ijerph17228297] [Medline: 33182661]
23. Lenzo V, Quattropani MC, Musetti A, Zenesini C, Freda MF, Lemmo D, et al. Resilience contributes to low emotional impact of the COVID-19 outbreak among the general population in Italy. Front Psychol 2020;11:576485 [FREE Full text] [doi: 10.3389/fpsyg.2020.576485] [Medline: 33250818]
24. Marchini S, Zaurino E, Bouziotis J, Brondino N, Delvenne V, Delhaye M. Study of resilience and loneliness in youth (18-25 years old) during the COVID-19 pandemic lockdown measures. J Community Psychol 2021 Mar;49(2):468-480. [doi: 10.1002/jcop.22473] [Medline: 33169377]
25. Petzold MB, Bendau A, Plag J, Pyrkosch L, Mascarell Maricic L, Betzler F, et al. Risk, resilience, psychological distress, and anxiety at the beginning of the COVID-19 pandemic in Germany. Brain Behav 2020 Sep;10(9):e01745 [FREE Full text] [doi: 10.1002/brb3.1745] [Medline: 32633464]
26. Tsibidaki A. Anxiety, meaning in life, self-efficacy and resilience in families with one or more members with special educational needs and disability during COVID-19 pandemic in Greece. Res Dev Disabil 2021 Feb;109:103830 [FREE Full text] [doi: 10.1016/j.ridd.2020.103830] [Medline: 33370720]
27. Veer IM, Riepenhausen A, Zerban M, Wackerhagen C, Puhlmann LMC, Engen H, et al. Psycho-social factors associated with mental resilience in the Corona lockdown. Transl Psychiatry 2021 Jan 21;11(1):67 [FREE Full text] [doi: 10.1038/s41398-020-01150-4] [Medline: 33479211]
28. Zager Kocjan G, Kavčič T, Avsec A. Resilience matters: Explaining the association between personality and psychological functioning during the COVID-19 pandemic. Int J Clin Health Psychol 2021;21(1):100198 [FREE Full text] [doi: 10.1016/j.ijchp.2020.08.002] [Medline: 33363581]
29. Davidson B, Schmidt E, Mallar C, Mahmoud F, Rothenberg W, Hernandez J, et al. Risk and resilience of well-being in caregivers of young children in response to the COVID-19 pandemic. Transl Behav Med 2021 Mar 16;11(2):305-313 [FREE Full text] [doi: 10.1093/tbm/ibaa124] [Medline: 33236766]
30. Killgore WD, Taylor EC, Cloonan SA, Dailey NS. Psychological resilience during the COVID-19 lockdown. Psychiatry Res 2020 Sep;291:113216 [FREE Full text] [doi: 10.1016/j.psychres.2020.113216] [Medline: 32544705]
31. Riehm KE, Brenneke SG, Adams LB, Gilan D, Lieb K, Kunzler AM, et al. Association between psychological resilience and changes in mental distress during the COVID-19 pandemic. J Affect Disord 2021 Mar 01;282:381-385 [FREE Full text] [doi: 10.1016/j.jad.2020.12.071] [Medline: 33421866]
32. Coulombe S, Pacheco T, Cox E, Khalil C, Doucerain M, Auger E, et al. Risk and resilience factors during the COVID-19 pandemic: a snapshot of the experiences of Canadian workers early on in the crisis. Front Psychol 2020;11:580702 [FREE Full text] [doi: 10.3389/fpsyg.2020.580702] [Medline: 33343455]
33. Kalisch R, Müller MB, Tüscher O. A conceptual framework for the neurobiological study of resilience. Behav Brain Sci 2015;38:e92. [doi: 10.1017/S0140525X1400082X] [Medline: 25158686]
34. Kalisch R, Köber G, Binder H, Ahrens KF, Basten U, Chmitorz A, et al. The frequent stressor and mental health monitoring-paradigm: a proposal for the operationalization and measurement of resilience and the identification of resilience processes in longitudinal observational studies. Front Psychol 2021;12:710493 [FREE Full text] [doi: 10.3389/fpsyg.2021.710493] [Medline: 34539510]
35. Amstadter A, Myers J, Kendler K. Psychiatric resilience: longitudinal twin study. Br J Psychiatry 2014 Oct;205(4):275-280 [FREE Full text] [doi: 10.1192/bjp.bp.113.130906] [Medline: 24723629]
36. van Harmelen A, Kievit RA, Ioannidis K, Neufeld S, Jones PB, Bullmore E, et al. Adolescent friendships predict later resilient functioning across psychosocial domains in a healthy community cohort. Psychol Med 2017 Apr 11;47(13):2312-2322. [doi: 10.1017/s0033291717000836]
37. Kalisch R, Kampa M. Stressor appraisal as an explanation for the influence of extra-individual factors on psychological resilience. In: Ungar M, editor. Multisystemic Resilience - Adaptation and Transformation in Contexts of Change. Oxfordshire, UK: Oxford University Press; Aug 26, 2021:135-153.
38. Ayash S, Schmitt U, Lyons DM, Müller MB. Stress inoculation in mice induces global resilience. Transl Psychiatry 2020 Jun 19;10(1):1-8. [doi: 10.1038/s41398-020-00889-0]
39. Seery M, Holman E, Silver R. Whatever does not kill us: cumulative lifetime adversity, vulnerability, and resilience. J Pers Soc Psychol 2010 Dec;99(6):1025-1041. [doi: 10.1037/a0021344] [Medline: 20939649]
40. Seery MD, Leo RJ, Lupien SP, Kondrak CL, Almonte JL. An upside to adversity?: moderate cumulative lifetime adversity is associated with resilient responses in the face of controlled stressors. Psychol Sci 2013 Jul 01;24(7):1181-1189. [doi: 10.1177/0956797612469210] [Medline: 23673992]

41. Bögemann S, Puhlmann L, Wackerhagen C, Zerban M, Riepenhausen A, Köber G, et al. DynaCORE-L: The DynaMORE longitudinal study on psychological resilience to the mental health consequences of the Corona crisis. Center for Open Science. 2022 May 18. URL: https://osf.io/qm4ba [accessed 2023-02-09]
42. m-Path - Free tool for blended care by KU Leuven. m-Path. URL: https://m-path.io/landing/ [accessed 2023-02-09]
43. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
44. Garnefski N, Kraaij V. Cognitive emotion regulation questionnaire – development of a short 18-item version (CERQ-short). Personality and Individual Differences 2006 Oct;41(6):1045-1053. [doi: 10.1016/j.paid.2006.04.010]
45. Beierlein C, Kovaleva A, Kemper CJ, Rammstedt B. Allgemeine Selbstwirksamkeit Kurzskala (ASKU). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS) 2012;17:zis35. [doi: 10.6102/zis35]
46. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
47. Dunkel D, Antretter E, Fröhlich-Walser S, Haring C. [Evaluation of the short-form social support questionnaire (SOZU-K-22) in clinical and non-clinical samples]. Psychother Psychosom Med Psychol 2005 May;55(5):266-277. [doi: 10.1055/s-2004-834746] [Medline: 15875274]
48. Rammstedt B, John OP. Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality 2007 Feb;41(1):203-212. [doi: 10.1016/j.jrp.2006.02.001]
49. Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997 Jan 01;27(1):191-197. [doi: 10.1017/s0033291796004242] [Medline: 9122299]
50. Mahalanobis PC. On the generalized distance in statistics. Proceedings of the National Institute of Science of India 1936:49-55.
51. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 1986;51(6):1173-1182. [doi: 10.1037/0022-3514.51.6.1173]
52. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
53. Ben-Zeev D, Frounfelker R, Morris S, Corrigan P. Predictors of self-stigma in schizophrenia: new insights using mobile technologies. J Dual Diagn 2012 Oct;8(4):305-314 [FREE Full text] [doi: 10.1080/15504263.2012.723311] [Medline: 23459025]
54. Shiyko M, Naab P, Shiffman S, Li R. Modeling complexity of EMA data: time-varying lagged effects of negative affect on smoking urges for subgroups of nicotine addiction. Nicotine & Tobacco Research 2013 Aug 03;16(Suppl 2):S144-S150. [doi: 10.1093/ntr/ntt109]
55. Jeronimus B, Kotov R, Riese H, Ormel J. Neuroticism's prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history, but the adjusted association hardly decays with time: a meta-analysis on 59 longitudinal/prospective studies with 443 313 participants. Psychol Med 2016 Oct;46(14):2883-2906. [doi: 10.1017/S0033291716001653] [Medline: 27523506]
56. Petri-Romão P, Engen H, Rupanova A, Puhlmann L, Zerban M, Neumann R, et al. Self-report assessment of Positive Appraisal Style (PAS): development of a process-focused and a content-focused questionnaire for use in mental health and resilience research. PsyArXiv. Preprint posted online on February 8, 2023 :1 [FREE Full text] [doi: 10.31234/osf.io/fpw94]
57. Ahrens KF, Neumann RJ, Kollmann B, Brokelmann J, von Werthern NM, Malyshau A, et al. Impact of COVID-19 lockdown on mental health in Germany: longitudinal observation of different mental health trajectories and protective factors. Transl Psychiatry 2021 Jul 17;11(1):392 [FREE Full text] [doi: 10.1038/s41398-021-01508-2] [Medline: 34282129]
58. Porter SR, Umbach PD. Student survey response rates across institutions: why do they vary? Res High Educ 2006 Mar;47(2):229-247. [doi: 10.1007/s11162-005-8887-1]
59. Smith WG. Does gender influence online survey participation?: A record-linkage analysis of university faculty online survey response behavior. Education Resources Information Center. 2008. URL: https://scholarworks.sjsu.edu/elementary_ed_pub [accessed 2023-09-14]
60. Wu M, Zhao K, Fils-Aime F. Response rates of online surveys in published research: a meta-analysis. Computers in Human Behavior Reports 2022 Aug;7:100206. [doi: 10.1016/j.chbr.2022.100206]
61. Hallers-Haalboom E, Maas J, Kunst L, Bekker M. The role of sex and gender in anxiety disorders: being scared "like a girl"? In: Lanzenberger R, Kranz GS, Savic I, editors. Handbook of Clinical Neurology: Sex Differences in Neurology and Psychiatry. Amsterdam, The Netherlands: Elsevier BV; 2020:359-368.
62. Kuehner C. Why is depression more common among women than among men? The Lancet Psychiatry 2017 Feb;4(2):146-158 [FREE Full text] [doi: 10.1016/s2215-0366(16)30263-2]
63. Riecher-Rössler A. Sex and gender differences in mental disorders. The Lancet Psychiatry 2017 Jan;4(1):8-9. [doi: 10.1016/s2215-0366(16)30348-0]
64. DynaCORE-L: The DynaMORE longitudinal study on psychological resilience to the mental health consequences of the Corona crisis - Data. Center for Open Science. URL: https://osf.io/uw9ep/ [accessed 2023-02-09]

65. DynaCORE-L: The DynaMORE longitudinal study on psychological resilience to the mental health consequences of the Corona crisis - Scripts. Center for Open Science. 2023. URL: https://osf.io/s4byj/ [accessed 2023-02-09]

Abbreviations

BC: behavioral coping
CSS: perceived change in social support during the COVID-19 pandemic
DynaCORE-C: DynaMORE cross-sectional study on psychological resilience to the mental health consequences of the COVID-19 pandemic
DynaCORE-L: DynaMORE longitudinal study on psychological resilience to the mental health consequences of the COVID-19 pandemic
DynaMORE: dynamic modeling of resilience
E: stressor exposure
GHQ-12: 12-item General Health Questionnaire
GSE: general self-efficacy
H1-4_MED: mediation hypotheses 1-4
M: mode
NEU: neuroticism
OPT: optimism
OSF: Center for Open Science
P: mental health problems
PA: positive appraisal
PAC: PA specifically of the COVID-19 pandemic
PASS-content: Perceived Positive Appraisal Style Scale—content-focused
PASS-process: Perceived Positive Appraisal Style Scale—process-focused
PASTOR: Positive Appraisal Style Theory of Resilience
Pcorr: adjusted α level
PSS: perceived social support
REC: perceived good stress recovery
RF: resilience factor
S: style
SR: stressor reactivity
T: trait
t: time

Edited by J Torous; submitted 15.02.23; peer-reviewed by E Smail, H Yin; comments to author 19.06.23; revised version received 26.07.23; accepted 28.07.23; published 17.10.23.

Please cite as:
Bögemann SA, Puhlmann LMC, Wackerhagen C, Zerban M, Riepenhausen A, Köber G, Yuen KSL, Pooseh S, Marciniak MA, Reppmann Z, Uści ko A, Weermeijer J, Lenferink DB, Mituniewicz J, Robak N, Donner NC, Mestdagh M, Verdonck S, van Dick R, Kleim B, Lieb K, van Leeuwen JMC, Kobylińska D, Myin-Germeys I, Walter H, Tüscher O, Hermans EJ, Veer IM, Kalisch R
Psychological Resilience Factors and Their Association With Weekly Stressor Reactivity During the COVID-19 Outbreak in Europe: Prospective Longitudinal Study
JMIR Ment Health 2023;10:e46518
URL: https://mental.jmir.org/2023/1/e46518
doi:10.2196/46518
PMID:37847551

Original Paper

Efficacy of a Smartphone App in Enhancing Medication Adherence and Accuracy in Individuals With Schizophrenia During the COVID-19 Pandemic: Randomized Controlled Trial

Huan Hwa Chen[1], PhD; Hsin Tien Hsu[2,3], PhD; Pei Chao Lin[2,3,4,5], PhD; Chin-Yin Chen[6], MSN; Hsiu Fen Hsieh[2,3], PhD; Chih Hung Ko[2], MD, PhD

[1]School of Nursing, Chung Hwa University of Medical Technology, Tainan, Taiwan
[2]College of Nursing, Kaohsiung Medical University, Kaohsiung, Taiwan
[3]Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan
[4]Department of Biomechatronics Engineering, National Pingtung University of Science and Technology, Pingtung, Taiwan
[5]Center for Long-Term Care Research, Kaohsiung Medical University, Kaohsiung, Taiwan
[6]Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan

Corresponding Author:
Hsiu Fen Hsieh, PhD
College of Nursing, Kaohsiung Medical University
100 Shih-Chuan 1st Rd., Sanmin District
Kaohsiung, 80708
Taiwan
Phone: 886 7 3121101 ext 2626
Fax: 886 7 3218364
Email: hsiufen96@kmu.edu.tw

Abstract

Background: Poor medication adherence or inaccuracy in taking prescribed medications plays an important role in the recurrence or worsening of psychiatric symptoms in patients with schizophrenia, and the COVID-19 pandemic impacted their medication adherence with exacerbated symptoms or relapse. The use of mobile health services increased during the COVID-19 pandemic, and their role in improving mental health is becoming clearer.

Objective: This study aimed to explore the effectiveness of a smartphone app (MedAdhere) on medication adherence and accuracy among patients with schizophrenia and to measure their psychiatric symptoms and cognitive functions.

Methods: In this 12-week experimental study, participants were provided interventions with the MedAdhere app, and data were collected between June 2021 and September 2022. A total of 105 participants were randomly assigned to either the experimental or control groups. We used the Positive and Negative Syndrome Scale and Mini-Mental State Examination to measure the participants' psychiatric symptoms and cognitive functions. Generalized estimating equations were used for data analysis.

Results: A total of 94 participants met the inclusion criteria and completed the protocol, and the medication adherence rate of the experimental group was 94.72% (2785/2940) during the intervention. Psychotic symptoms (positive, negative, and general psychopathology symptoms) and cognitive functions (memory, language, and executive function) were significantly improved in the experimental group compared to the control group after the intervention.

Conclusions: The MedAdhere app effectively and significantly improved medication adherence and, thereby, the psychiatric symptoms of patients with schizophrenia. This artificial intelligence assisted app could be extended to all patients who need to be reminded to take medication on schedule.

Trial Registration: ClinicalTrials.gov NCT05892120; https://clinicaltrials.gov/study/NCT05892120

(JMIR Ment Health 2023;10:e50806) doi:10.2196/50806

KEYWORDS

cognitive functions; medication adherence; psychiatric symptoms; schizophrenia; smartphone app

Introduction

Overview

Schizophrenia is a mental disorder that imposes a heavy economic burden and causes severe functional disabilities, including cognitive impairment [1,2]. Its clinical manifestations are divided into positive and negative symptoms. Positive symptoms include delusions, hallucinations, and disorganized behavior, which are often accompanied by impaired cognitive and socio-occupational functions, while in contrast, negative symptoms manifest as apathy and decreased motivation, which gradually become apparent as the disorder progresses [3]. Many factors accelerate aging in patients with schizophrenia, making their life span shorter than that of individuals without schizophrenia, with impacts such as deterioration of life functioning, the presence of chronic illnesses, obesity, and an unhealthy lifestyle [4]. Male and female patients with schizophrenia have an average life expectancy of 61.48 years and 60.81 years, respectively [5].

Schizophrenia is a treatable disorder, and early intervention can prevent further deterioration. Antipsychotic drugs are the first choice for alleviating the psychiatric symptoms of these patients, and the goal of regular medication with the right drugs, dosages, and frequency is to prevent recurrence and overall functional deterioration [6,7]. Although it is common for patients with schizophrenia to not adhere to their prescriptions for a variety of reasons, including the stigma of mental illness, side effects of drugs, forgetting to take medicine, and lack of insight [8,9].

The medication adherence rate among patients diagnosed with schizophrenia for the first time is only 50% in the first year [4,5], and the rate of relapse within 5 years is up to 80% [10]. Similar studies have reported that 50% to 80% of patients with schizophrenia do not take their medications regularly [6,7] or take the wrong medications or dosages [8]. Nonadherence may lead to the deterioration of symptoms, a decline in daily life functions, and a poor prognosis; it may also result in hospitalization and the escalation of medical expenses [9,11].

The COVID-19 pandemic has affected social contact patterns in various situations, including in the field of medical services. Individuals with mental disorders exhibited higher levels of concern regarding COVID-19 and fear of getting infected [12], with the result that 64% of them had worsening of their symptoms, 39% were unable to receive treatment, and 38% missed their medications [13]. Subsequently, wider use of telehealth and digital tools was promoted, such as smartphone apps, in providing medical services for the reduction of person-to-person contact to lower the likelihood of COVID-19 transmission to patients [14]. Poor medication adherence was one of the major causes of the increasing relapse rate among patients with schizophrenia during the COVID-19 pandemic [15], and it is crucial to use remote monitoring to track oral medication adherence to lower relapse rates or the deterioration of psychiatric symptoms.

An increasing number of smartphone apps are being developed to care for and train patients with schizophrenia [16,17], which include apps for improving patients' psychiatric symptoms, medication adherence, and social and cognitive functions [17-19]. Most of the apps for medication adherence in patients with schizophrenia are designed to remind patients to take their medications on time [19,20]. A smartphone app named "MedAdhere" was developed by Chen et al [21] to observe the medication adherence of these patients with recognition of the appearance of the prescribed antipsychotic drugs as well as drug-taking behavior through the smartphone camera [21].

Study Aims

This study aimed to explore the effectiveness of intervention with the MedAdhere app on medication adherence and accuracy in patients with schizophrenia and to measure their psychiatric symptoms and cognitive functions during the period of the COVID-19 pandemic. We hypothesized that using the MedAdhere app would improve medication adherence, psychiatric symptoms, and cognitive function in patients with schizophrenia compared to the control group.

Methods

Study Design

The protocol was registered with ClinicalTrials.gov (NCT05892120) at the time of submission. We used an experimental design, using random sampling with each daycare center as a unit, and assigned patients in these centers to either the experimental or control group using a random number generator. This study is single-blinded, with research assistants being the blinded party. All participants received the usual care, while participants in the experimental group additionally downloaded and used the MedAdhere app on their personal smartphones during the 12-week study.

Participants and Setting

The sample size was calculated using G*Power (version 3.1.1; Axel Buchner) for the F_1 test and its repeated measures with an effect size of 0.25, a significance level of .05, and a power (1 – β) of 0.80; the estimated minimum sample size was 68 participants [22], with all participants being recruited from a psychiatric daycare center across 2 medical centers and 1 regional hospital.

There are various treatment modalities for patients with schizophrenia based on each patient's psychiatric condition, including acute psychiatric units, chronic wards, daycare centers, and community services, and they may be transferred from one treatment modality to another based on their condition. Patients in daycare were transferred from acute wards where they received initial treatment, and after a period of time, they were treated well with a relatively stabilized mental status under a suitable medication and dosage. The care in the daycare center mainly focuses on regular medication and psychosocial rehabilitation. Daycare for psychiatric patients can be effective in returning the patient to a normal life in the community or relieving the strain on a family [23]. The daycare center is an important psychiatric rehabilitation unit for patients with stable psychiatric symptoms who are suitable users of this app.

A total of 105 patients participated in this study, and they were randomly assigned to the experimental or control group,

although 11 patients in the experimental group did not complete this study; finally, 94 participants completed all processes of the study protocol (Multimedia Appendix 1).

The inclusion criteria were as follows: (1) patients diagnosed with schizophrenia by a psychiatrist according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* [3]; (2) aged between 20 and 65 years; (3) able to read traditional Chinese; (4) owned a smartphone; and (5) currently admitted to a psychiatric daycare center. The exclusion criteria were as follows: patients with intellectual disability or severe cognitive function impairment (Mini-Mental State Examination [MMSE] score <17) [24].

Study Protocol

Control Group

Medications for the control group are divided into 2 parts: medication in the daytime and nighttime. Daytime medication at the daycare center is assisted and supported by daily activities arranged by the staff. Nighttime medication is taken by the patient without any intervention.

Experimental Group

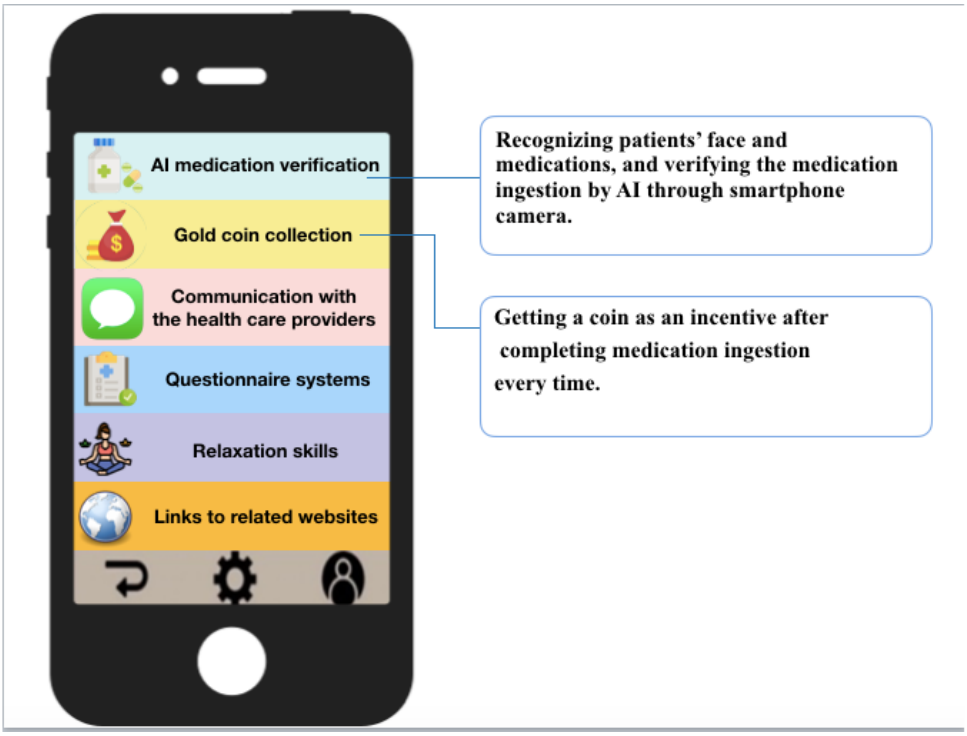
Medications for the experimental group are divided into 2 parts: medication in the daytime and nighttime. Daytime medication at the daycare center is assisted and supported by daily activities arranged by the staff. Intervention for nighttime medication is conducted with the MedAdhere app.

Before the intervention, researchers first explained the functions of the MedAdhere app to the patient and set the patient's antipsychotic medication details on their smartphones, including type, dosage, and frequency, and then completed the recognition of the patient's face and the appearance of all antipsychotics. Then, each patient was taught how to use the app step by step (Figure 1) by providing operation manual practical training. The intervention stage began, and we confirmed that the patient knew how to operate the app. In addition to the verification of medication status using the MedAdhere app, we used a bidirectional interaction approach where patients could ask researchers questions on the web through the Line app (NHN Japan; which is one of the related links in the MedAdhere app), and the team would send messages to or call the patient when any abnormalities were detected. The home page of the MedAdhere app is shown in Figure 2.

Figure 1 shows the steps of antipsychotic ingestion by patients recognized by the MedAdhere app, and any errors can be detected using this app. For example, if the participant took the medication at the wrong time or took the wrong drug or dosage, this MedAdhere app promptly detected such errors and uploaded them to the cloud, while the results of antipsychotic ingestion were also recorded by this app and uploaded to a cloud server. The MedAdhere app could also be used offline and automatically upload all records to the cloud server when connected to the internet. Researchers and assistants had the ability to access cloud data whenever necessary to observe the patient's medication status, in addition to 2 routine check-in times at 10 PM and 11 PM.

Figure 1. Steps of antipsychotic ingestion recognised by MedAdhere app.

Figure 2. Home page of the MedAdhere app. AI: artificial intelligence.



Data Collection

The data were collected between June 2021 and September 2022. Demographic characteristics, including age, gender, marital status, years of education, living status, and religious beliefs, were gathered. The symptoms and cognitive functions of all participants were measured using the Positive and Negative Syndrome Scale (PANSS) and MMSE at baseline and at the end of the intervention, respectively.

Instruments

MedAdhere App

The MedAdhere app is a medication management app available on Android platforms for patients' medication scheduling, reminders, tracking, medication adherence assessments and recognition of the patient's face and the appearance of antipsychotics. When the participant takes the medication at the wrong time or takes the wrong drug or dosage, this MedAdhere app can detect these errors, upload such errors to the cloud, and trigger an alarm for the participant [21].

Medication Adherence Rate

Medication adherence is defined as the proportion of the administered drug doses in relation to the prescribed doses taken by the patient within a specific time frame. Based on this definition, we calculated the patient's medication adherence rate of the control group by counting their pills left at the end of intervention, and the patient's medication adherence rate of the experimental group was calculated from the data in the cloud. In addition to antipsychotics, there was no antidepressant prescribed for any of our participants, but there were other psychotropic medications prescribed for our participants, and these adjunctive drugs were not included to be identified by our MedAdhere app.

PANSS Instrument

The PANSS is among the best-validated instruments for assessing positive, negative, and general psychopathological symptoms associated with patients with schizophrenia. It is a standardized clinical interview that rates the presence and severity of positive and negative symptoms as well as general psychopathological symptoms (such as anxiety and depression) within the past week for patients with schizophrenia. Of the 30 items on the scale, 7 are positive symptoms, 7 are negative symptoms, and 16 are general psychopathological symptoms. The symptom severity for each item is rated according to the anchoring points on a 7-point scale (1=absent and 7=extreme) that best describes the symptom [25]. The PANSS score is the sum of ratings across items, ranging from 7 to 49 for the PANSS and 16 to 112 for the general psychopathology scale. Higher scores indicate more severe symptoms [26]. The Cronbach α was .86 [27].

MMSE Instrument

Cognitive tests were conducted using the MMSE [24]. The MMSE is, in some way, the best-known and most widely used measure of cognition in clinical practice worldwide. The MMSE consists of 15 questions that assess the following 7 cognitive domains: orientation to time (5 points), place (5 points), memory registration and recall (3 points), attention and calculation (8 points), language (5 points), reasoning and problem-solving (3 points), and executive function (1 point). The total MMSE score ranges from 0 to 30, with higher scores indicating better cognitive function. The internal consistency of the MMSE method was represented by Cronbach α at the level of .78 [28].

Data Analysis

Descriptive analyses by chi-square and Mann-Whitney *U* test were performed using SPSS software (version 21.0; SPSS Inc). Descriptive analysis was based on participants' demographic

characteristics, such as age, gender, education, marital status, and religious beliefs. After controlling the confounding factor of age, the generalized estimating equation was used to assess psychotic symptoms (positive, negative, and general psychopathological symptoms) and cognitive functions (orientation, memory, attention and calculation, language, reasoning and problem-solving, and executive function) from baseline to the end of intervention in the 2 groups. All tests were 2-tailed, and $P<.05$ was considered significant.

Ethical Considerations

This study was approved by the institutional review board of the Kaohsiung Medical University Hospital: KMUHIRB-SV (I)–20200096. The principal investigator visited each psychiatric daycare center, explained the study's purpose and process to all participants, and inquired regarding their willingness to participate. Written informed consent was obtained from all participants, and they were informed that they could withdraw from the study at any time without providing a reason. Each participant was offered NT $400 (US $13-14) as compensation for their participation.

Results

Baseline Data

All participants had access to their private mobile networks (4G LTE). A total of 105 participants were randomly assigned to the experimental or control group, and 94 completed all the processes of our protocol. The demographic characteristics of the respondents are presented in Table 1.

Table 1. Demographic characteristics and baseline outcomes.

Characteristics	Control group (n=59)	Experimental group (n=35)	Chi-square test or Mann-Whitney *U* Test	*P* value
Age (years), mean (SD)	50.45 (11.4)	41.29 (10.5)	–1.587[a]	.11
Gender, n (%)			0.980[b]	.32
Male	28 (47.5)	20 (57.1)		
Female	31 (52.5)	15 (42.9)		
Education, n (%)			0.223[b]	.90
Junior high school	6 (10.2)	4 (11.5)		
Senior high school	37 (62.7)	20 (57.1)		
≥College	16 (27.1)	11 (31.4)		
Marital status, n (%)			0.826[b]	.66
Single	47 (79.7)	28 (80.0)		
Married	8 (13.6)	6 (17.1)		
Divorced	4 (6.7)	1 (2.9)		
Living status, n (%)			2.669[b]	.45
Alone	11 (18.6)	5 (14.3)		
Spouse	2 (3.14)	1 (2.8)		
Friends	2 (3.14)	0 (0)		
Parent or child	44 (74.6)	29 (82.9)		
Religious belief, n (%)			0.159[b]	.69
No	17 (28.8)	10 (28.6)		
Yes	42 (71.2)	25 (71.4)		
PANSS[c], mean (SD)				
Positive symptoms	46.36 (4.1)	49.41 (3.9)	–0.526[a]	.60
Negative symptoms	46.38 (3.6)	49.39 (3.7)	–0.519[a]	.60
General psychopathology symptoms	44.10 (6.2)	53.23 (7.0)	–1.572[a]	.11
MMSE[d], mean (SD)				
Orientation	46.08 (1.7)	49.89 (0.4)	–0.974[a]	.33
Attention and calculation	48.00 (1.7)	46.66 (1.6)	–0.247[a]	.81
Language	48.87 (0.6)	45.19 (0.8)	–0.889[a]	.37
Reasoning and problem-solving	47.53 (0.7)	47.46 (0.4)	–0.018[a]	.99
Memory	46.92 (0.7)	48.47 (0.7)	–0.329[a]	.74
Executive function	46.83 (0.4)	48.63 (0.3)	–0.517[a]	.61

[a]Chi-square test (*df*).

[b]Mann-Whitney *U* test.

[c]PANSS: Positive and Negative Syndrome Scale.

[d]MMSE: Mini-Mental State Examination.

Medication Adherence Rate

The medication adherence rates were 94.72% (2785/2940; total doses taken by all patients over 12 weeks / [35 person ×12 weeks × 7 days × doses] × 100%) and 64.43% (3193/4956; total doses taken by all patients over 12 weeks / [59 person ×12 weeks × 7 days × doses] × 100%) in the experimental group and the control group, respectively. The medication adherence rate of the experimental group was higher than the control group during the intervention (Table 2).

Table 2. Medication adherence rate.

Group	At 8 weeks, n/N (%)	At 12 weeks, n/N (%)
Control group (n=59)	2111/3304 (63.89)	3193/4956 (64.43)
Experimental group (n=35)	1764/1960 (90)	2785/2940 (94.72)

Changes in Psychiatric Symptoms and Cognitive Functions

The changes in psychiatric symptoms and cognitive functions in the 2 groups from baseline to the end of the intervention in 12 weeks are shown in Tables 3 and 4.

Table 3. Changes of psychiatric symptoms from baseline to the end of intervention.

Parameters	T2 vs T0	T1 vs T0	Age	EG[a] × T2 vs CG[b] × T2	EG × T1 vs CG × T1	EG × age
Positive symptoms						
β	–2.33	.92	–.02	–.84	–1.95	.15
Mean	10.54	13.79	__[c]	12.03	10.92	—
SE	1.7	1.49	—	0.26	0.23	—
P value	.17	.54	—	.33	.02[d]	.74
Negative symptoms						
β	–2.59	–.43	–.01	–.92	–1.99	0
Mean	8.63	10.79	—	10.3	9.23	—
SE	1.5	1.64	—	0.38	0.36	—
P value	.09	.79	—	.24	.007[e]	.71
General psychopathology symptoms						
β	–4.63	1.98	–3.16	–4.37	–.03	–4.63
Mean	20.9	27.51	—	21.93	21.16	—
SE	2.97	2.62	—	0.18	0.15	—
P value	.12	.45	—	.02[d]	<.001[f]	.65

[a]EG: experimental group.

[b]CG: control group.

[c]Not available.

[d]*P*<.05.

[e]*P*<.01.

[f]*P*<.001.

Table 4. Changes of cognitive functions from baseline to the end of the intervention.

Parameter	T2 vs T0	T1 vs T0	Age	EG[a] × T2 vs CG[b] × T2	EG × T1 vs CG × T1	EG × age
Orientation						
β	–.10	–.32	–2.5E	.25	.25	.00
Mean	9.68	9.46	__[c]	10.03	10.03	—
SE	0.63	0.63	0.01	0.26	0.26	0.01
P value	.87	.53	—	.35	.28	.65
Attention and calculation						
β	.72	–.04	.00	.19	.16	–.00
Mean	7.46	6.70	—	6.93	6.90	—
SE	0.73	0.73	0.02	0.02	0.02	0.02
P value	.32	.95	—	.62	.66	.61
Memory						
β	.30	–.26	.01	–.07	.31	–.02
Mean	2.64	2.08	—	2.27	2.65	—
SE	0.37	0.37	0.01	0.01	0.01	0.01
P value	.42	.28	—	.71	.03[d]	.02[d]
Language						
β	.19	.18	.01	.41	.28	.00
Mean	4.74	4.73	—	4.96	4.83	—
SE	0.34	0.34	0.01	0.01	0.01	0.01
P value	.57	.64	—	.16	.16	.86
Reasoning and problem-solving						
β	.14	–.28	.00	.18	–.01	–.01
Mean	2.79	2.37	—	2.83	2.65	—
SE	0.25	0.25	0.00	0.00	0.00	0.00
P value	.58	.30	—	.16	.97	.03[d]
Executive function						
β	–.45	–.18	–.00	.33	.18	.00
Mean	0.59	0.84	—	1.35	1.20	—
SE	0.22	0.22	0.00	0.00	0.00	0.00
P value	.04[d]	.40	—	<.001[e]	.08	.80

[a]EG: experimental group.

[b]CG: control group.

[c]Not available.

[d]*P*<.05.

[e]*P*<.001.

After controlling the confounding factor of age, we found significant improvements in positive, negative, and general psychopathology symptoms in the experimental group, with a decrease of 0.143 in the positive (*P*=.02), 0.137 in the negative (*P*=.007), and 0.013 in the general psychopathological symptoms (*P*<.001) following an 8-week intervention with the MedAdhere app (Table 3). Additionally, there were significant improvements in the cognitive domains of memory, language, and executive function in the experimental group, with an increase of 1.51 in language (*P*=.012) and 1.39 in executive function (*P*<.001) after intervention with the MedAdhere app for 12 weeks (T2; Table 4).

Discussion

Overview

The policy of the government in Taiwan during the COVID-19 period was to reduce interpersonal contact against the transmission of coronavirus. In addition to the medication adherence of patients, this app was designed to reduce interpersonal contacts during the COVID-19 pandemic and to reduce the burden on health care providers.

Main Findings

In this study, the MedAdhere app contributed to the medication adherence rate during the intervention, and this study also found that patients' positive and negative symptoms and general psychopathology symptoms significantly improved after the intervention, even though the experimental group had more severe general psychopathology symptoms than the control group at baseline. In addition, there were significant differences in cognitive functions between the 2 groups.

Medication Adherence

The estimated medication adherence rate was 50% among patients with schizophrenia who did not receive any intervention [29]. Another study reported an adherence rate of 65% with different interventions for the same purpose of enhancing medication adherence in patients [30]. A higher medication adherence rate essentially contributes to stabilizing the psychiatric symptoms of patients and lowering their readmission rates [31]. The medication adherence rate in the experimental group was 94.72% during the intervention using the MedAdhere app, higher than the control group's 64.43% in this study. Our results suggest that the utility of the MedAdhere app for artificial intelligence verification of medication ingestion behavior plays an important role in enhancing patients' medication adherence, and this is an easily implemented intervention [32].

Psychiatric Symptoms

Our results showed that intervention with the MedAdhere app could significantly improve patients' positive and negative symptoms in 12 weeks compared to the control group. This result was similar to that of a previous study showing that using the medication adherence app could improve positive and negative symptoms in patients with schizophrenia [33].

Negative symptoms should be treated by other rehabilitation modalities to improve pleasure, motivation, and cognitive skills relevant to apathy and anhedonia [34,35]. In addition to verifying medication ingestion behavior using the MedAdhere app, patients were able to ask questions and interact with researchers on the web during the intervention through the Line app, which is one of the related links on the MedAdhere app's home page. Interactions between researchers and patients can be regarded as a form of social skill training that has been documented to have medium effects on reducing negative symptoms in individuals with schizophrenia [36], and such interactions might be one of the reasons for alleviating our participants' negative symptoms of social withdrawal, blunted affect, emotional withdrawal, and a lack of spontaneity and flow of conversation.

Cognitive Functions

Cognitive functions declined in the control group, whereas the cognitive functions of the experimental group improved significantly in the domains of memory, language, and executive function after the 12-week intervention. This could be attributed to the use of the MedAdhere app with novel technology, which provides repeated multisensory stimuli and rewards after each medication ingestion, which in turn enhances their motivation to continue using the app.

Cognitive function in patients with schizophrenia can be improved through exercise [37], quality sleep [38], and symptomatic stability [39]. Our results may have been affected by the patient's sleep patterns [38] and symptomatic stability [39], both of which are considered to be improved by antipsychotics. Compared to conventional methods for medication adherence that require routine monitoring by family members, telephone visits by home health care providers, or monthly visits by clinicians, this innovative MedAdhere app effectively empowers patients' medication adherence and reduces the burden on health care providers. Such an artificial intelligence-assisted adherence-enhancing app can be extended to all patients who require regular medication but may forget to take their medication on time, for the benefits of better adherence and to prevent the intake of the wrong medication or duplicating doses.

The MedAdhere app requires minimal training for implementation. To ensure that patients can continue using this app, we initially designed the app interface to be user-friendly and easy to operate, with the added incentive of earning a digital gold coin upon completing medication ingestion as a reward. In addition, we have established alliances with clinical professionals and family members, providing the app and cloud access to health care teams and allowing them to keep monitoring patients' medication adherence. With the support of patients' families, each participant can keep using this MedAdhere app after the intervention.

Limitations

This study had some limitations. First, there might be selection bias because the MedAdhere app can only be installed on the Android platform. Patients who did not own a smartphone or whose smartphones were not Android system based were excluded from this study. Second, this was a single-blinded design with a 12-week study duration, and a double-blinded study design with long-term follow-up is indicated to reduce potential bias and confirm the effects of the MedAdhere app on improving psychiatric symptoms and cognitive functions in patients with schizophrenia. Third, we did not calculate how many times or the frequency of communication between participants and our team. In the future, it will be possible to calculate the number of bidirectional interactions. Fourth, the MedAdhere app has only been used in Taiwan thus far and can be translated into other languages for use in other countries in the future. Finally, there were other psychotropic medications prescribed for our participants, and these adjunctive drugs were not included nor identified by our MedAdhere app.

Conclusions

The novelty of this study is the development of the MedAdhere app using a smartphone's camera to recognize patients' faces, drugs, and medication-taking behavior. This accessible tool has the potential to significantly improve medication adherence in patients with schizophrenia. The additional value of this MedAdhere app is to assist associated health care providers or partners' families in confirming patients' medication status to reduce the risk of relapse or repeated hospitalization.

Acknowledgments

The authors thank all participants for their generous involvement and also thank the Ministry of Science and Technology of Taiwan for their support (MOST 110-2314-B-037-028). The article processing charge is supported by the Ministry of Science and Technology, Taiwan (MOST 111-2314-B-037-026).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study protocol.
[PDF File (Adobe PDF File), 299 KB - mental_v10i1e50806_app1.pdf]

Multimedia Appendix 2
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 1179 KB - mental_v10i1e50806_app2.pdf]

References

1. Acosta FJ, Hernández JL, Pereira J, Herrera J, Rodríguez CJ. Medication adherence in schizophrenia. World J Psychiatry 2012 Oct 22;2(5):74-82. [doi: 10.5498/wjp.v2.i5.74] [Medline: 24175171]
2. Semahegn A, Torpey K, Manu A, Assefa N, Tesfaye G, Ankomah A. Psychotropic medication non-adherence and associated factors among adult patients with major psychiatric disorders: a protocol for a systematic review. Syst Rev 2018 Jan 22;7(1):10-10 [FREE Full text] [doi: 10.1186/s13643-018-0676-y] [Medline: 29357926]
3. Tandon R, Gaebel W, Barch DM, Bustillo J, Gur RE, Heckers S, et al. Definition and description of schizophrenia in the DSM-5. Schizophr Res 2013 Oct;150(1):3-10. [doi: 10.1016/j.schres.2013.05.028] [Medline: 23800613]
4. Chafetz L, White MC, Collins-Bride G, Nickens J, Cooper BA. Predictors of physical functioning among adults with severe mental illness. Psychiatr Serv 2006 Feb 20;57(2):225-231. [doi: 10.1176/appi.ps.57.2.225] [Medline: 16452700]
5. Kirkpatrick B, Messias E, Harvey P, Fernandez-Egea E, Bowie C. Is schizophrenia a syndrome of accelerated aging? Schizophr Bull 2008 Aug 20;34(6):1024-1032 [FREE Full text] [doi: 10.1093/schbul/sbm140]
6. Alvarez-Jimenez M, Priede A, Hetrick SE, Bendall S, Killackey E, Parker AG, et al. Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies. Schizophr Res 2012 Aug;139(1-3):116-128. [doi: 10.1016/j.schres.2012.05.007] [Medline: 22658527]
7. Moritz S, Hünsche A, Lincoln T. Nonadherence to antipsychotics: the role of positive attitudes towards positive symptoms. Eur Neuropsychopharmacol 2014 Nov;24(11):1745-1752 [FREE Full text] [doi: 10.1016/j.euroneuro.2014.09.008]
8. Uhlmann C, Kaehler J, Harris MSH, Unser J, Arolt V, Lencer R. Negative impact of self-stigmatization on attitude toward medication adherence in patients with psychosis. J Psychiatr Pract 2014 Sep;20(5):405-410. [doi: 10.1097/01.pra.0000454787.75106.ae] [Medline: 25226204]
9. Trauer T, Vaddadi K, Soosai E. Issues in drug administration in a psychiatric rehabilitation unit. Int J Soc Psychiatry 1995;41(3):174-179. [doi: 10.1177/002076409504100302] [Medline: 8847198]
10. García S, Martínez-Cengotitabengoa M, López-Zurbano S, Zorrilla I, López P, Vieta E, et al. Adherence to antipsychotic medication in bipolar disorder and schizophrenic patients. J Clin Psychopharmacol 2016;36(4):355-371. [doi: 10.1097/jcp.0000000000000523]
11. Leucht S, Heres S. Epidemiology, clinical consequences, and psychosocial treatment of nonadherence in schizophrenia. J Clin Psychiatry 2006;67 Suppl 5:3-8. [Medline: 16822090]
12. Quittkat HL, Düsing R, Holtmann FJ, Buhlmann U, Svaldi J, Vocks S. Perceived impact of COVID-19 across different mental disorders: a study on disorder-specific symptoms, psychosocial stress and behavior. Front Psychol 2020 Nov 01;11(11):586246 [FREE Full text] [doi: 10.3389/fpsyg.2020.586246] [Medline: 33281685]
13. Costa M, Pavlo A, Reis G, Ponte K, Davidson L. COVID-19 concerns among persons with mental illness. Psychiatr Serv 2020 Nov 01;71(11):1188-1190 [FREE Full text] [doi: 10.1176/appi.ps.202000245] [Medline: 32878542]

14. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193-1133 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
15. Bassiony MM, Sehlo MG, Ibrahim EF, Zayed AE, Atwa SA. Assessment of compliance and relapse in patients with schizophrenia before and after the onset of COVID-19 pandemic. Int J Psychiatry Clin Pract 2023 Jun;27(2):127-133 [FREE Full text] [doi: 10.1080/13651501.2022.2124175] [Medline: 36149774]
16. Niendam TA, Tully LM, Iosif AM, Kumar D, Nye KE, Denton JC, et al. Enhancing early psychosis treatment using smartphone technology: A longitudinal feasibility and validity study. J Psychiatr Res 2018 Jan 28;96(2):239-246 [FREE Full text] [doi: 10.1016/j.jpsychires.2017.10.017] [Medline: 29126059]
17. Schlosser D, Campellone T, Kim D, Truong B, Vergani S, Ward C, et al. Feasibility of PRIME: a cognitive neuroscience-informed mobile app intervention to enhance motivated behavior and improve quality of life in recent onset schizophrenia. JMIR Res Protoc 2016 Apr 28;5(2):e77-113 [FREE Full text] [doi: 10.2196/resprot.5450] [Medline: 27125771]
18. Eisner E, Bucci S, Berry N, Emsley R, Barrowclough C, Drake RJ. Feasibility of using a smartphone app to assess early signs, basic symptoms and psychotic symptoms over six months: a preliminary report. Schizophr Res 2019 Jun;208(4):105-113 [FREE Full text] [doi: 10.1016/j.schres.2019.04.003] [Medline: 30979665]
19. Kreyenbuhl J, Record EJ, Himelhoch S, Charlotte M, Palmer-Bacon J, Dixon LB, et al. Development and feasibility testing of a smartphone intervention to improve adherence to antipsychotic medications. Clin Schizophr Relat Psychoses 2019;12(4):152-167 [FREE Full text] [doi: 10.3371/CSRP.KRRE.070816] [Medline: 27454213]
20. Kidd SA, Feldcamp L, Adler A, Kaleis L, Wang W, Vichnevetski K, et al. Feasibility and outcomes of a multi-function mobile health approach for the schizophrenia spectrum: App4Independence (A4i). PLoS One 2019 Jan;14(7):e0219491-e0219496 [FREE Full text] [doi: 10.1371/journal.pone.0219491] [Medline: 31306439]
21. Chen CY, Liang TY, Ko CH, Hsieh HF. The effectiveness of real-time identification of medication ingestion in schizophrenic patients in Taiwan. Kaohsiung J Med Sci 2023 Jan;39(1):95-96 [FREE Full text] [doi: 10.1002/kjm2.12637] [Medline: 36533567]
22. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160 [FREE Full text] [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
23. Ives GA. Psychiatric day care. Can Fam Physician 1975 Oct;21(10):61-63 [FREE Full text] [Medline: 20469245]
24. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198 [FREE Full text] [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
25. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987 Nov;13(2):261-276. [doi: 10.1093/schbul/13.2.261] [Medline: 3616518]
26. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? Schizophr Res 2005 Nov 15;79(2-3):231-238. [doi: 10.1016/j.schres.2005.04.008] [Medline: 15982856]
27. Montoya A, Valladares A, Lizán L, San L, Escobar R, Paz S. Validation of the Excited Component of the Positive and Negative Syndrome Scale (PANSS-EC) in a naturalistic sample of 278 patients with acute psychosis and agitation in a psychiatric emergency room. Health Qual Life Outcomes 2011 Mar 29;9(2-3):18 [FREE Full text] [doi: 10.1186/1477-7525-9-18] [Medline: 21447155]
28. Kabátová O, Puteková S, Martinková J, Súkenníková M. Analysis of psychometric features of the Mini-Mental State Examination and the Montreal Cognitive Assessment methods. CSWHI 2016 Aug 14;7(2):62-69 [FREE Full text] [doi: 10.22359/cswhi_7_2_08]
29. Lacro JP, Dunn LB, Dolder CR, Leckband SG, Jeste DV. Prevalence of and risk factors for medication nonadherence in patients with schizophrenia: a comprehensive review of recent literature. J Clin Psychiatry 2002 Oct 14;63(10):892-909 [FREE Full text] [doi: 10.4088/jcp.v63n1007] [Medline: 12416599]
30. Pijnenborg GHM, Withaar FK, Brouwer WH, Timmerman ME, van den Bosch RJ, Evans JJ. The efficacy of SMS text messages to compensate for the effects of cognitive impairments in schizophrenia. Br J Clin Psychol 2010 Jun;49(Pt 2):259-274. [doi: 10.1348/014466509X467828] [Medline: 19735607]
31. Nieuwenhuis-Mark RE. The death knoll for the MMSE: has it outlived its purpose? J Geriatr Psychiatry Neurol 2010 Sep;23(3):151-157. [doi: 10.1177/0891988710363714] [Medline: 20231732]
32. Tessier A, Dupuy M, Baylé FJ, Herse C, Lange AC, Vrijens B, et al. Brief interventions for improving adherence in schizophrenia: a pilot study using electronic medication event monitoring. Psychiatry Res 2020 Mar;285(3):112780-112787 [FREE Full text] [doi: 10.1016/j.psychres.2020.112780] [Medline: 31954541]
33. Cochran JM, Fang H, Sonnenberg JG, Cohen EA, Lindenmayer JP, Reuteman-Fowler JC. Healthcare provider engagement with a novel dashboard for tracking medication ingestion: impact on treatment decisions and clinical assessments for adults with schizophrenia. Neuropsychiatr Dis Treat 2022 Mar;18:1521-1534 [FREE Full text] [doi: 10.2147/NDT.S369123] [Medline: 35928793]

34. Correll CU. Using patient-centered assessment in schizophrenia care: defining recovery and discussing concerns and preferences. J Clin Psychiatry 2020 Apr 14;81(3):1521-1534 [FREE Full text] [doi: 10.4088/JCP.MS19053BR2C] [Medline: 32297720]
35. Favrod J, Nguyen A, Chaix J, Pellet J, Frobert L, Fankhauser C, et al. Improving pleasure and motivation in schizophrenia: a randomized controlled clinical trial. Psychother Psychosom 2019 Apr 14;88(2):84-95 [FREE Full text] [doi: 10.1159/000496479] [Medline: 30783071]
36. Granholm E, Harvey PD. Social skills training for negative symptoms of schizophrenia. Schizophr Bull 2018 Apr 06;44(3):472-474 [FREE Full text] [doi: 10.1093/schbul/sbx184] [Medline: 29315427]
37. Firth J, Stubbs B, Rosenbaum S, Vancampfort D, Malchow B, Schuch F, et al. Aerobic exercise improves cognitive functioning in people with schizophrenia: a systematic review and meta-analysis. Schizophr Bull 2017 May 01;43(3):546-556 [FREE Full text] [doi: 10.1093/schbul/sbw115] [Medline: 27521348]
38. Miller DD. Atypical antipsychotics: sleep, sedation, and efficacy. Prim Care Companion J Clin Psychiatry 2004;6(Suppl 2):3-7 [FREE Full text] [Medline: 16001094]
39. Ali D, Elserafi D, Soltan M, Eissa M, Zein H, Elshahawi H. Schizophrenic patients' cognitive functions in relation to their metabolic profile: a cross-sectional, comparative study on an Egyptian sample. Middle East Curr Psychiatry 2020 Oct 20;27(1):3-7 [FREE Full text] [doi: 10.1186/s43045-020-00053-w]

Abbreviations

MMSE: Mini-Mental State Examination
PANSS: Positive and Negative Syndrome Scale

Edited by J Torous; submitted 13.07.23; peer-reviewed by HF Hsieh, K Nye, W Wei; comments to author 24.08.23; revised version received 20.09.23; accepted 10.10.23; published 14.12.23.

Please cite as:
Chen HH, Hsu HT, Lin PC, Chen CY, Hsieh HF, Ko CH
Efficacy of a Smartphone App in Enhancing Medication Adherence and Accuracy in Individuals With Schizophrenia During the COVID-19 Pandemic: Randomized Controlled Trial
JMIR Ment Health 2023;10:e50806
URL: https://mental.jmir.org/2023/1/e50806
doi:10.2196/50806
PMID:38096017

Original Paper

Efficacy of an Electronic Cognitive Behavioral Therapy Program Delivered via the Online Psychotherapy Tool for Depression and Anxiety Related to the COVID-19 Pandemic: Pre-Post Pilot Study

Elnaz Moghimi[1*], PhD; Callum Stephenson[1*], MSc; Anika Agarwal[1*]; Niloofar Nikjoo[1*], MD; Niloufar Malakouti[1,2*], MD; Gina Layzell[3*], BSc; Anne O'Riordan[4*], BScOT; Jasleen Jagayat[3*], BSc; Amirhossein Shirazi[5*], MD, PhD; Gilmar Gutierrez[1*], MD; Ferwa Khan[1*]; Charmy Patel[1*], BSc, MPH; Megan Yang[1*], MD; Mohsen Omrani[1,5*], MD, PhD; Nazanin Alavi[1,3*], MD

[1]Department of Psychiatry, Faculty of Health Sciences, Queen's University, Kingston, ON, Canada

[2]School of Rehabilitation Therapy, Queen's University, Kingston, ON, Canada

[3]Centre for Neuroscience Studies, Queen's University, Kingston, ON, Canada

[4]Patient and Family Centered Care, Kingston Health Sciences Centre, Kingston, ON, Canada

[5]OPTT Inc, Toronto, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Nazanin Alavi, MD
Department of Psychiatry
Faculty of Health Sciences
Queen's University
166 Brock Street
Kingston, ON, K7L 5G23
Canada
Phone: 1 6135443310
Fax: 1 6135449666
Email: nazanin.alavitabari@kingstonhsc.ca

Abstract

Background: Lockdowns and social distancing resulting from the COVID-19 pandemic have worsened the population's mental health and made it more difficult for individuals to receive care. Electronic cognitive behavioral therapy (e-CBT) is a cost-effective and evidence-based treatment for anxiety and depression and can be accessed remotely.

Objective: The objective of the study was to investigate the efficacy of online psychotherapy tailored to depression and anxiety symptoms during the pandemic.

Methods: The pilot study used a pre-post design to evaluate the efficacy of a 9-week e-CBT program designed for individuals with depression and anxiety affected by the pandemic. Participants were adults (N=59) diagnosed with major depressive disorder and generalized anxiety disorder, whose mental health symptoms initiated or worsened during the COVID-19 pandemic. The online psychotherapy program focused on teaching coping, mindfulness, and problem-solving skills. Symptoms of anxiety and depression, resilience, and quality of life were assessed.

Results: Participants demonstrated significant improvements in symptoms of anxiety (P=.02) and depression (P=.03) after the intervention. Similar trends were observed in the intention-to-treat analysis. No significant differences were observed in resilience and quality-of-life measures. The sample comprised mostly females, making it challenging to discern the benefits of the intervention in males. Although a pre-post design is less rigorous than a controlled trial, this design was selected to observe changes in scores during a critical period.

Conclusions: e-CBT for COVID-19 is an effective and accessible treatment option. Improvements in clinical symptoms of anxiety and depression can be observed in individuals whose mental health is affected by the COVID-19 pandemic.

Trial Registration: ClinicalTrials.gov NCT04476667; https://clinicaltrials.gov/study/NCT04476667

International Registered Report Identifier (IRRID): RR2-10.2196/24913

(JMIR Ment Health 2023;10:e51102) doi:10.2196/51102

KEYWORDS

mental health; depression; anxiety; cognitive behavioral therapy; online; COVID-19; efficacy; electronic cognitive behavioral therapy; online psychotherapy tool; pandemic; evidence-based treatment

Introduction

The persistence of the novel COVID-19 pandemic significantly deteriorated the mental health of the global population [1]. Although lockdowns and restrictions are critical in reducing viral transmissions, these public health measures generate feelings of uncertainty, fear, isolation, and loneliness in citizens [2,3]. Compared to the pre-pandemic era, an increased prevalence of stress, anxiety, depression, and suicidal ideations has also been observed [4,5]. The psychosocial implications of the COVID-19 pandemic can result in worsened mental health symptoms in vulnerable persons [6]. For example, individuals who experienced mandatory COVID-19 quarantine were 5 times more likely to have suicidal and self-harm ideations than those who did not quarantine [7]. Specific to Canada, suicidal ideation in adults increased from 2.7% to 4.2% during the pandemic [8].

Moreover, 50% of Canadians reported a worsening in their mental health and 41% reported being anxious [9]. Since the onset of the COVID-19 pandemic, the number of Canadians experiencing high to extremely high anxiety quadrupled and those with high self-reported depression more than doubled [10]. Also of concern, study respondents reported a reduction in the quality and quantity of available mental health support systems [10]. Undoubtedly, alternative mental health services are needed to adequately address the adverse consequences of the pandemic, narrow the gap in mental health inequities, and reduce the pressure on an already burdened health care system [11].

Electronically delivered psychological interventions (e-psychotherapy) are easily scalable, cost-effective, and pragmatic methods of delivering necessary mental health care to a wide range of individuals [12]. Using the most evidence-based psychotherapy, electronic cognitive behavioral therapy (e-CBT) offers flexible and accessible mental health care to people affected by lockdowns and social distancing. The goal of cognitive behavioral therapy (CBT) is to enhance the awareness and modification of maladaptive cognitions and behaviors within a short period. Numerous studies have demonstrated the efficacy of e-CBT in a myriad of implementation styles. Self-help, guided, and asynchronous deliveries have all shown effectiveness in symptom management, with results showing comparable improvements in in-person CBT, medications, and control groups [13-17]. e-CBT has also been shown to improve resilience and quality of life, in addition to depressive symptom management [17]. The predesigned therapy content of e-CBT enables clinicians to disseminate core and standardized elements of the therapy at a considerably faster rate [18-22]. This distinguishing feature of e-CBT can contribute to a broader audience reach and shorten wait times [21].

The efficacy of e-psychotherapy has also been demonstrated in populations affected by the COVID-19 pandemic. In a sample of 670 Swedish adults, a 3-week self-guided e-CBT program reduced dysfunctional worry and anxiety related to the pandemic [23]. Additionally, distant-delivered CBT has shown promise in patients with posttraumatic stress disorder (PTSD) during the pandemic [24]. Stress-related insomnia was also reduced in a sample of 194 individuals who completed a 1-week self-guided e-CBT intervention during the pandemic [25]. Further, a largely self-guided e-CBT program consisting of 6 lessons was associated with significant reductions in anxiety and depression symptoms and psychological distress [26]. The same study observed a 504% increase in the number of monthly e-CBT course registrations during the COVID-19 period compared to the year prior [26].

While these studies have highlighted e-CBT's benefits in treating pandemic-related mental distress, the efficacy of the programs can vary depending on the degree of mental care provider engagement [27]. As therapist engagement positively influences the efficacy of e-CBT programs, a therapist-guided e-CBT program was postulated to render greater benefits than self-guided programs. Therapist-guided e-CBT can improve patient outcomes while simultaneously lowering costs and increasing care capacity [19,22].

The purpose of this study was to develop and administer an e-psychotherapy program for patients with major depressive disorder (MDD) and generalized anxiety disorder (GAD), affected by the COVID-19 pandemic [28]. The program focused on pandemic-related mental health concerns present in people with MDD and GAD. The 9-week therapist-guided e-psychotherapy was available through a secure, confidential, and cloud-based platform. Patient progress was tracked through the platform. The primary objective of the program was to significantly reduce stress and psychological distress in patients, from pre- to postintervention. Additionally, the effects of the program on quality of life and resilience were investigated.

Methods

Study Design

This pilot study had a pre-post single-arm design. Since the efficacy of the e-CBT intervention in mitigating GAD and MDD symptoms has already been established in previous trials [18-22], this study focused on the pre-post effects of participating in the e-CBT program during the COVID-19 pandemic. Moreover, since the trial was conducted at the beginning of the pandemic, it was important to observe the pre-post effects in a pilot study and subsequently use that data to inform a randomized controlled trial (RCT). The intervention was a 9-week care provider-guided e-psychotherapy program that addressed mental health problems related to the COVID-19 pandemic.

Ethical Considerations

All procedures were approved by the Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (6029910). The online psychotherapy tool (OPTT; OPTT Inc), complies with the Health Insurance Portability and Accountability Act, Personal Information Protection and Electronic Documents Act, and Service Organization Control-2. OPTT only collects anonymized metadata and uses encrypted data on its platform.

Participants

Participants (n=80) were recruited through referrals from Hotel Dieu Hospital and Providence Care Hospital outpatient clinics in Kingston, Ontario, Canada, and self-referrals from social media and web-based communities. Recruitment occurred from June 2020 to June 2021. Participants who provided informed consent were evaluated by a psychiatrist on the research team through a secure video appointment. During the appointments, the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) criteria were used to determine MDD or GAD diagnoses. To confirm the diagnosis, the Mini-International Neuropsychiatrist Interview (MINI), version 7.0.2 DSM-5, was also administered by a trained research assistant on the team. Of the participants assessed for eligibility, 14 did not meet the inclusion and exclusion criteria and 7 never began the program after enrollment. As a result, 59 participants commenced the study. Demographic information for all participants who began the program can be found in Table 1.

Table 1. Demographic information of participants who began the online program.

Participants	Total (n=59)	Completers (n=38)	Dropouts (n=21)
Age[a]			
n	54	35	19
Mean (SD)	32.26 (12.67)	37.37 (13.53)	35.47 (15.23)
Gender[b]**, n (%)**			
Woman	41 (69)	30 (79)	11 (52)
Man	11 (19)	5 (13)	6 (29)
Other	2 (3)	0 (0)	2 (10)
Did not indicate	5 (9)	3 (8)	2 (10)

[a]t_{52}=–0.471; *P*=.64.

[b]χ^2_3=6.76; *P*=.08.

Inclusion criteria included adults residing in Ontario, Canada, between the ages of 18 and 65 years; capacity to consent; ability to speak and read English; a primary diagnosis of GAD or MDD, with symptoms that started or worsened during the COVID-19 pandemic; and consistent and reliable access to the internet. Exclusion criteria included active psychosis, acute mania, severe alcohol or substance use disorder, active suicidal or homicidal ideations, and mental health problems that were secondary to a medical condition. To prevent potential confounds, participants were also excluded if they were receiving or had previously received any form of CBT within the past year.

Intervention

At the start of the program, participants were assigned to a care provider that was supervised by the lead psychiatrist on the team. All the care providers who had previous training in psychotherapy and before the study commencement were trained by a psychiatrist involved in the study. At a specified date during each of the 9 weeks, the care provider assigned a predesigned therapy module to the patient through OPTT, a secure cloud-based platform. The modules were also followed by homework assignments which were due on a specific day of the week. The average completion time of the weekly modules was approximately 40 minutes. The homework was then submitted through OPTT, and appropriate care provider feedback was provided to the patient. To maintain standardized care and efficiency, the care providers used predesigned session-specific feedback templates to respond to each homework submission (Alavi and Omrani [17]). The feedback focused on praising the participant's time and effort, summarizing material from the previous sessions, and reviewing, discussing, and evaluating the participant's homework. If participants had additional questions or concerns for their care providers, OPTT's secure chat option was available. A more comprehensive description of the treatment, care providers, and procedures was previously described in Alavi et al's [28] protocol.

The care-provider-guided e-psychotherapy program consisted of a combination of CBT, mindfulness therapy [29,30], and problem-solving–based therapy [31,32]. The modules were designed to be accessible on any device (ie, desktops, cellphones, and tablets) and compatible across multiple browsers. The sessions included multiple animations and examples to retain participant interest and engagement. The therapy's engaging modules were also customized to reflect common challenges faced by individuals with MDD and GAD during the COVID-19 pandemic. These modules were adapted from previous clinical trials, which used a similar approach to treating depression and anxiety [18-22].

The primary intent of the program was to teach individuals how to identify and change maladaptive thought patterns, behaviors, and emotions. The lessons focused on the effects of the pandemic on mood, the basics of CBT, deep breathing

techniques, body scan and meditation, the self-care kit, SMART (Specific, Measurable, Achievable, Realistic, and Timely) goals, thinking errors, the 5-part model [33], and thought records. The first 2 sessions addressed symptoms caused by the fear of illness and concerns about personal safety in the context of the pandemic. The remaining 7 sessions focused on building adaptive coping skills to address the uncertainties of the COVID-19 pandemic and related symptoms of depression and anxiety. A combination of CBT techniques, problem-solving techniques, and mindfulness practices was integrated into the program. Textbox 1 provides further detail on the content of each session.

Textbox 1. An overview of the content covered in each weekly session of the online program.

(1) Anxiety and depression during the COVID-19 pandemic

- Introduces psychotherapy structure and cognitive behavioral therapy concepts and discusses common mental health symptoms while setting expectations for the course. Also provides an overview of the COVID-19 pandemic and how it can affect mental health. SMART (Specific, Measurable, Achievable, Realistic, and Timely) goals are provided as an assignment.

(2) Cognitive behavioral therapy and deep breathing

- The 5-part model is introduced along with breathing techniques.

(3) Body scanning and meditation

- Different techniques for body scanning and meditation are discussed, along with healthy distractions to use as coping mechanisms.

(4) Self-care kit

- Techniques to calm the fight or flight response are discussed. Building a self-care kit with calming sensory items are introduced (see, taste, feel, and hear) along with sleep hygiene.

(5) Thinking errors

- The idea of thinking errors is introduced with mental filters, jumping to conclusions, overgeneralization, discounting positives, magnification and minimization, emotional reasoning, and more examples discussed. Tools to identify a thinking error and fix it are presented.

(6) Thought records and automatic thoughts

- Cognitive reappraisal strategies are discussed initially and the thought record tool to better help understand feelings are taught. Automatic thoughts and how to identify them are included.

(7) Evidence

- Ways to compile evidence that a negative automatic thought is unsupported are discussed, along with strategies to better balance thought patterns.

(8) Alternative and balanced thinking

- Producing alternative and balanced thoughts are further discussed with reflection and double standard tools being provided.

(9) Review

- A general overview of breathing techniques, body scanning, goal setting, self-care kit, testing thoughts, challenging thinking errors, thought records, and double standard techniques are discussed to summarize the content covered in the program.

The care providers in the study were trained research assistants who were under the supervision of the principal investigator, a clinician-scientist with expertise in e-CBT [18-22]. All homework feedback was reviewed by licensed therapists and the principal investigator before submission. The psychotherapy platform enabled care providers to schedule and assign predesigned modules, homework assignments, and symptomatology questionnaires. Feedback, checkups, questions, and concerns were communicated through the platform's chat feature. The chat feature was a text-based messaging system operating within the OPTT platform. All sessions were followed by homework assignments that were submitted through the platform, reviewed by the care provider, and followed by personalized written feedback within 3 days of submissions. Predesigned and session-specific feedback templates were used for care providers to write their personalized feedback. Feedback for each session was given weekly by the care provider with additional interactions occurring through the chat feature on OPTT; however, often the participant wished to contact them with a question, comment, or concern. The weekly feedback for homework sessions was structured based on predesigned templates that were then customized to fit each participant's responses, needs, and situation.

Outcome Measures

Data from all outcomes were collected at baseline, week 4, and after the last session at week 9. The severity of anxiety and depression symptoms were assessed using the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7) [34] and 9-item Patient Health Questionnaire (PHQ-9) [35], respectively. The 14-item resilience scale (RS-14) [36] was used to measure resilience levels. Scores range from 14 to 98 with scores less

than 65 indicating low resilience. Finally, scores from the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) [37] were used to measure the quality of life. Q-LES-Q scores range from 14 to 70. All the questionnaires were collected directly through OPTT. The GAD-7 assesses the severity of symptoms related to anxiety with each response ranging from 0 (not at all) to 3 (nearly every day) with scores ranging from 0 to 21 with a higher score representing more severe symptoms. The PHQ-9 assesses a client's depressive symptom severity as well as functional health. Each question was scored 0 (not at all) to 3 (nearly every day) with a total score of 0 to 27 with 27 being the most severe. The RS-14 evaluates emotional resilience with each question ranging from 1 (strongly disagree) to 7 (strongly agree). Scores range from 14 to 98 with a higher score equating to increased resilience. The Q-LES-Q assesses levels of enjoyment and satisfaction in daily functioning and life. Each question was rated from 1 (very poor) to 5 (very good) with scores ranging from 14 to 70, with a higher score indicating a better quality of life.

Statistical Analysis

A 1-way repeated measure ANOVA was conducted to evaluate the differences in clinical scores of study completers at baseline, midtreatment, and at the end of treatment. Independent samples *t* tests were used to compare pre- and posttreatment scores for all clinical outcomes. For this analysis, intention-to-treat analysis was used to include data from participants who did not complete the study in its entirety. Missing data were not imputed and were analyzed on a per-protocol basis. The significance level for all tests was set to α=.05. All statistical analyses were conducted using IBM SPSS Statistics for Mac (version 24; IBM Corp).

Results

Out of the eligible participants who commenced the study (N=59; Figure 1), 21 participants dropped out of the study (n=11 from weeks 1-3, n=7 from weeks 4-6, and n=3 at week 7), and 38 participants completed the study. Reasons for dropout were not disclosed. Most of the total sample identified as women (n=41, 69.49%). A total of 2 participants identified as other and both dropped out of the treatment at weeks 4 and 6, respectively. The average age of the sample was 32.26 (SD 12.67). No significant differences were observed at baseline for any demographic variables (Table 1) or scores of treatment completers and dropouts (Table 2) for participants who began the program. A significant difference was observed between the number of sessions completed by those who dropped out and those who finished the program (*P*<.001). On average, treatment dropouts completed 40.77% of the treatment before dropping out.

GAD-7 scores were reduced from 11.57 to 9.86 to 9.43 (start, mid, and end, respectively). A 1-way repeated measures ANOVA demonstrated improvements in GAD-7 scores (*P*=.02; Table 3). Bonferroni post hoc analysis demonstrated a significant reduction in GAD-7 scores at 0 weeks and 9 weeks (*P*=.02) but not at 0 and 4 weeks (*P*=.22) and 4 and 9 weeks (*P*=.99). Intention-to-treat analysis using unpaired sample *t* tests also demonstrated a significant lowing of scores from pre- to posttreatment (*P*=.02; Table 4).

PHQ-9 scores were reduced from 14.65 to 12.43 to 11.84. One-way repeated measures ANOVA demonstrated significant improvements in PHQ-9 scores (*P*=.03; Table 3). Bonferroni's post hoc analysis demonstrated a significant reduction in GAD-7 scores at 0 weeks and 9 weeks (*P*=.01) but not at 0 weeks and 4 weeks (*P*=.18) and 4 and 9 weeks (*P*=.99). Intention-to-treat analysis using unpaired sample *t* tests also demonstrated a significant reduction in scores from pre- to posttreatment (*P*=.02; Table 4). ANOVA and intention-to-treat analyses indicated no significant differences in Q-LES-Q or RS-14 scores over the 9 weeks (Figure 2).

Regarding noncompleters, 6 participants completed session 1, three completed session 2, two completed session 3, one completed session 4, two completed session 5, four completed session 6, and three completed session 7.

Figure 1. Flow diagram of study recruitment.

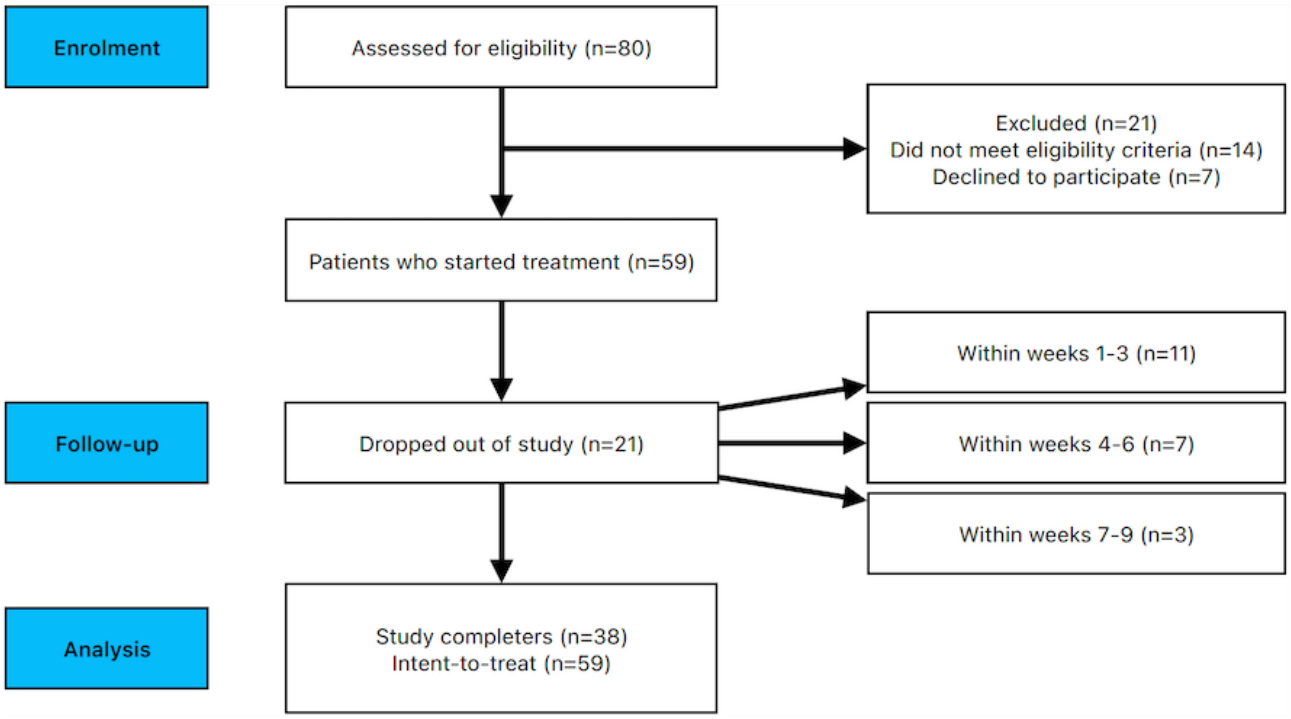


Table 2. Demographics and characteristics of the sample.

Characteristics	Total (n=59)	Completers (n=38)	Dropouts (n=21)	Analysis (completers vs dropouts)	
				t test (*df*)	*P* value
Sessions completed, mean (SD)	7.10 (2.92)	9 (0)	3.67 (2.35)	–10.39 (20)	≤.001
Baseline scores, mean (SD)					
GAD-7[a]	11.74 (5.15)	11.58 (5.07)	12.05 (5.40)	0.33 (57)	.74
PHQ-9[b]	14.95 (6.46)	14.53 (6.26)	15.71 (6.90)	0.50 (57)	.67
Q-LES-Q[c]	39.71 (8.89)	40.05 (8.42)	39.10 (9.86)	–0.39 (57)	.70
RS-14[d]	60.12 (15.70)	59.45 (15.11)	61.33 (17.03)	0.44 (57)	.66

[a]GAD-7: 7-item Generalized Anxiety Disorder Questionnaire.

[b]PHQ-9: 9-item Patient Health Questionnaire.

[c]Q-LES-Q: Quality of Life Enjoyment and Satisfaction Questionnaire.

[d]RS-14: 14-item resilience scale.

Table 3. Means, SDs, and ANOVA of primary outcomes at 3 time points within the 9-week trial.

Questionnaire	Baseline (0 weeks), mean (SD)	Midtreatment (4 weeks), mean (SD)	Posttreatment (9 weeks), mean (SD)	Repeated measures ANOVA (Time)		
				F test (*df*)	*P* value	ηp^2
GAD-7[a] (n=37)	11.57 (5.14)	9.86 (4.71)	9.43 (5.42)	4.00 (2, 36)	.02	0.10
PHQ-9[b] (n=37)	14.65 (6.30)	12.43 (7.74)	11.84 (7.41)	3.72 (2, 36)	.03	0.094
Q-LES-Q[c] (n=36)	40.17 (8.53)	41.64 (9.63)	42.11 (11.79)	0.670 (2, 35)	.51	0.242
RS-14[d] (n=37)	59.19 (15.23)	62.00 (17.29)	42.83 (19.06)	1.96 (2, 46)	.15	0.052

[a]GAD-7: 7-item Generalized Anxiety Disorder Questionnaire.

[b]PHQ-9: 9-item Patient Health Questionnaire.

[c]Q-LES-Q: Quality of Life Enjoyment and Satisfaction Questionnaire.

[d]RS-14: 14-item resilience scale.

Table 4. Means, SDs, and pre-post independent samples *t* test outcomes[a].

Questionnaire	Baseline (0 weeks)		Posttreatment (9 weeks)		Independent *t* test	
	n	Mean (SD)	n	Mean (SD)	*t* test (*df*)	*P* value (1-sided)
GAD-7[b]	59	11.75 (5.15)	37	9.43 (5.42)	2.100 (94)	.02
PHQ-9[c]	59	14.95 (6.46)	37	11.84 (7.41)	2.170 (94)	.02
Q-LES-Q[d]	59	39.71 (8.89)	36	42.11 (11.79)	–1.052 (59.18)	.15
RS-14[e]	59	60.12 (15.70)	37	62.00 (19.06)	–0.526 (94)	.30

[a]Equal variances are assumed, and all significance is listed as 1-tailed.

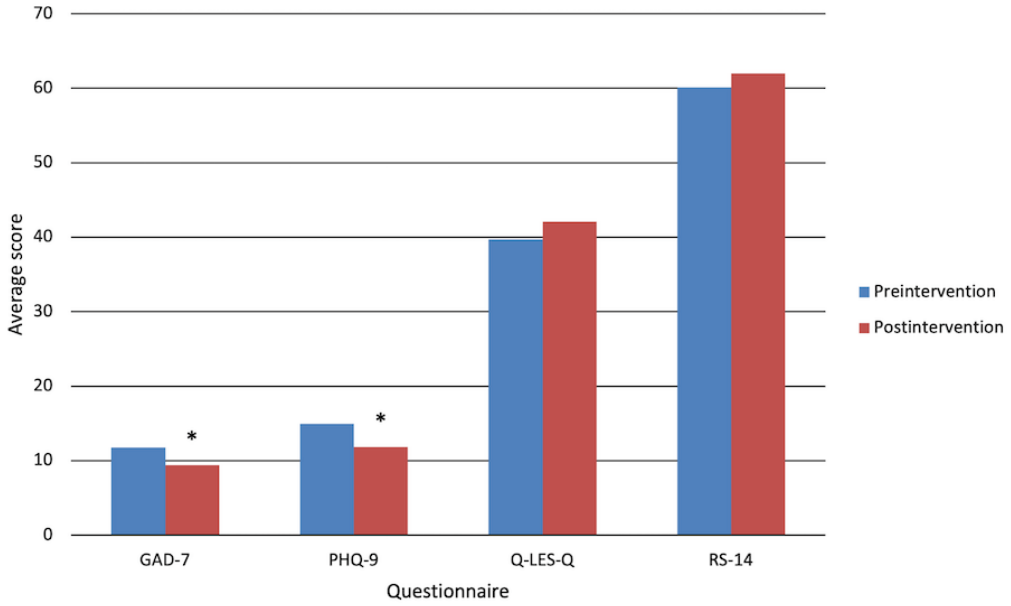
[b]GAD-7: 7-item Generalized Anxiety Disorder Questionnaire.

[c]PHQ-9: 9-item Patient Health Questionnaire.

[d]Q-LES-Q: Quality of Life Enjoyment and Satisfaction Questionnaire.

[e]RS-14: 14-item resilience scale.

Figure 2. Average intention-to-treat scores (n=38) for each questionnaire, pre- and postintervention. * denotes a significant difference at *P*<.05. GAD-7: Generalized Anxiety Disorder 7-Item Questionnaire; PHQ-9: Patient Health Questionnaire 9-Item; Q-LES-Q: Quality of Life Enjoyment and Satisfaction Questionnaire; RS-14: 14-item resilience scale.



Discussion

Principal Findings

This study aimed to investigate the efficacy of an e-CBT program focused on mental health challenges that were initiated or worsened during the COVID-19 pandemic. A single-arm pre-post study design was used to observe changes in clinical symptoms associated with an e-CBT program disseminated via OPTT at the beginning of the pandemic. The program focused specifically on managing anxiety and depression-related symptoms and navigating through challenges during this time. The findings indicated significant improvements in clinical symptoms after the e-CBT intervention.

Within program completers, anxiety symptoms were reduced by 18.50% and depression symptoms by 19.19%. Average baseline levels of GAD-7 scores indicated moderate anxiety, which was reduced to mild level posttreatment. While average PHQ-9 scores indicated moderate to severe levels of depression, the score moved from the higher end to the lower end of the range posttreatment. Symptoms and disorders related to anxiety and depression are 2 of the most common mental health concerns during the COVID-19 pandemic [5]. Prevalence rates during this period have been reported to be 31.90% for anxiety and 33.70% for depression [5]. Slightly greater reductions of 19.74% in anxiety symptoms and 20.80% in depression symptoms were observed in the intention-to-treat analysis. These findings suggest that treatment dropouts may have also demonstrated some degree of improvement in their clinical symptoms.

Dropout rates for this study were 35.59%, which is on par with rates found in other therapist-guided online interventions and lower than self-guided online interventions [38]. Therefore, the addition of therapists is suggested to reduce attrition rates of online interventions. Although a previous 12-week mobile app effectively improved mental health during the pandemic [39], this study supports the sufficiency of a 9-week intervention in mitigating clinical symptoms. The addition of a care provider

to this online program may have improved adherence and sped the trajectory of clinical improvement. Indeed, guided interventions not only demonstrate higher efficacy than their unguided counterparts but can significantly increase completion rates of online interventions targeting depression and anxiety disorders [40]. This study contributes to the body of evidence by suggesting that the time needed for observable clinical improvements may lie between 4 and 9 weeks.

Although improvements in clinical symptoms were observed, there were no significant differences in quality of life and resilience scores before and after the intervention. It is noted that resilience scores can be a screening tool for quality of life, which may partly explain why both scores did not significantly differ pre- and postintervention [36]. Although it requires further investigation, it is postulated that environmental changes may have contributed to a stagnation of the scores [41]. The study was conducted toward the beginning of the pandemic when individuals were just beginning to become acclimated to lockdowns, quarantines, and social distancing laws. There is mounting data that highlight declining mental health as the pandemic persisted [42-44]. At the same time, social isolation, loneliness, financial anxiety, the unknown nature of the virus, collapsed health care systems, high death rates, and other factors inevitably contributed to a worsening of quality of life [45-47]. It is noted that the nonsignificant improvements in quality of life and resilience scores postintervention may have been due to a relatively small sample size. However, program efficacy may be enhanced with the addition of more unique strategies and skills that specifically target these factors during the pandemic.

Additionally, it is possible that changes in quality of life could occur later after program completion, as they may take time to manifest. Future work should incorporate a long-term follow-up to investigate these possible changes. The efficacy of other situation-specific digital psychotherapeutics related to population-wide crises (ie, pandemics and natural disasters) should continue to be developed and further investigated.

Limitations

Although the study possessed many strengths, it was not without limitations. Instead of an RCT, a pre-post study design was used for several reasons. First, the efficacy of online psychotherapy programs has already been demonstrated in previous trials [18-22]. Second, the study was implemented as a pilot that aimed to explore the efficacy of the e-CBT program over a critical period. This project was used as a pilot and proof of concept to inform a future and larger-scale RCT. It was necessary to implement a design that factored in preintervention scores that coincided with the beginning stages of the pandemic.

Nevertheless, previous data support attributing the improvement symptoms to the online psychotherapy intervention. The second limitation pertained to the gender dispersion within the sample. The sample was relatively homogenous, with 69.49% of participants identifying as women. Differences in clinical outcomes vary due to sex-based differences in immunological response and gender-based differences in behavior and comorbidities [48]. As a result, males tend to experience greater severity and fatality for COVID-19 infections than females [48]. Since the intervention was designed for individuals with mild to moderate clinical symptoms, fewer men may have met the inclusion criteria, thereby resulting in a gender-imbalanced sample. However, it is possible that the program was not as accessible to men, and a qualitative investigation into this should be implemented in the future. Additionally, there is a possible confounder of geographic location affecting symptom severity. As government lockdown restrictions were not always universal across the province, living in a different area of Ontario could have resulted in heightened or lessened restrictions, altering the stressors in that participant's life.

Conclusions

In conclusion, the study demonstrated significant improvements in clinical symptoms of anxiety and depression after the use of a 9-week COVID-19–specific e-CBT program. Taken together, e-CBT provides individuals limited by location, time, and cost access to evidence-based and effective therapies. The evidence strongly suggests that online psychotherapy can supplement this care model. Although no changes in quality of life or resilience were reported, these findings may be due to the persistent environmental challenges outside the normative levels observed prepandemic. While the efficacy of e-CBT has been observed across various populations, it is warranted for future studies to investigate the role of gender in treatment availability and help-seeking.

Conflicts of Interest

AS and MO have ownership stakes in OPTT, the online psychotherapy platform used in the study. NA and MO are cofounders of OPTT. NA also has shares in the company.

References

1. Kumar A, Nayar KR. COVID 19 and its mental health consequences. J Ment Health 2021;30(1):1-2 [FREE Full text] [doi: 10.1080/09638237.2020.1757052] [Medline: 32339041]
2. McGinty EE, Presskreischer R, Han H, Barry CL. Psychological distress and loneliness reported by US adults in 2018 and April 2020. JAMA 2020;324(1):93-94 [FREE Full text] [doi: 10.1001/jama.2020.9740] [Medline: 32492088]
3. Minihan E, Gavin B, Kelly BD, McNicholas F. COVID-19, mental health and psychological first aid. Ir J Psychol Med 2020;37(4):259-263 [FREE Full text] [doi: 10.1017/ipm.2020.41] [Medline: 32404221]

4. Daly Z, Slemon A, Richardson CG, Salway T, McAuliffe C, Gadermann AM, et al. Associations between periods of COVID-19 quarantine and mental health in Canada. Psychiatry Res 2021;295:113631 [FREE Full text] [doi: 10.1016/j.psychres.2020.113631] [Medline: 33310417]
5. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
6. Hao F, Tan W, Jiang L, Zhang L, Zhao X, Zou Y, et al. Do psychiatric patients experience more psychiatric symptoms during COVID-19 pandemic and lockdown? A case-control study with service and research implications for immunopsychiatry. Brain Behav Immun 2020;87:100-106 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.069] [Medline: 32353518]
7. Xin M, Luo S, She R, Yu Y, Li L, Wang S, et al. Negative cognitive and psychological correlates of mandatory quarantine during the initial COVID-19 outbreak in China. Am Psychol 2020;75(5):607-617 [FREE Full text] [doi: 10.1037/amp0000692] [Medline: 32673008]
8. Liu L, Pollock NJ, Contreras G, Tonmyr L, Thompson W. Prevalence of suicidal ideation among adults in Canada: results of the second survey on COVID-19 and mental health. Health Rep 2022;33(5):13-21 [FREE Full text] [doi: 10.25318/82-003-x202200500002-eng] [Medline: 35587219]
9. Worry, gratitude and boredom: as COVID-19 affects mental, financial health, who fares better; who is worse? Angus Reid Institute. 2020. URL: https://angusreid.org/wp-content/uploads/2020/04/2020.04.27_COVID-mental-health.pdf [accessed 2023-12-12]
10. Dozois DJA. Anxiety and depression in Canada during the COVID-19 pandemic: a national survey. Can Psychol 2021;62(1):136-142 [FREE Full text] [doi: 10.1037/cap0000251]
11. Jenkins EK, McAuliffe C, Hirani S, Richardson C, Thomson KC, McGuinness L, et al. A portrait of the early and differential mental health impacts of the COVID-19 pandemic in Canada: findings from the first wave of a nationally representative cross-sectional survey. Prev Med 2021;145:106333 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106333] [Medline: 33509605]
12. Twomey C, O'Reilly G, Byrne M. Effectiveness of cognitive behavioural therapy for anxiety and depression in primary care: a meta-analysis. Fam Pract 2015;32(1):3-15. [doi: 10.1093/fampra/cmu060] [Medline: 25248976]
13. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
14. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychol Med 2014;44(15):3137-3150. [doi: 10.1017/S0033291714000245] [Medline: 25065947]
15. Saddichha S, Al-Desouki M, Lamia A, Linden IA, Krausz M. Online interventions for depression and anxiety—a systematic review. Health Psychol Behav Med 2014;2(1):841-881 [FREE Full text] [doi: 10.1080/21642850.2014.945934] [Medline: 25750823]
16. Sztein DM, Koransky CE, Fegan L, Himelhoch S. Efficacy of cognitive behavioural therapy delivered over the internet for depressive symptoms: a systematic review and meta-analysis. J Telemed Telecare 2018;24(8):527-539. [doi: 10.1177/1357633X17717402] [Medline: 28696153]
17. Alavi N, Omrani M. Online Cognitive Behavioral Therapy: An E-Mental Health Approach to Depression and Anxiety. Cham, Switzerland: Springer Nature Switzerland AG; 2018.
18. Alavi N, Hirji A. The efficacy of PowerPoint-based CBT delivered through email: breaking the barriers to treatment for generalized anxiety disorder. J Psychiatr Pract 2020;26(2):89-100. [doi: 10.1097/PRA.0000000000000455] [Medline: 32134882]
19. Alavi N, Hirji A, Sutton C, Naeem F. Online CBT is effective in overcoming cultural and language barriers in patients with depression. J Psychiatr Pract 2016;22(1):2-8. [doi: 10.1097/PRA.0000000000000119] [Medline: 26813483]
20. Alavi N, Stefanoff M, Hirji A, Khalid-Khan S. Cognitive behavioural therapy through PowerPoint: efficacy in an adolescent clinical population with depression and anxiety. Int J Pediatr 2018;2018:1396216 [FREE Full text] [doi: 10.1155/2018/1396216] [Medline: 30532790]
21. Alavi N, Stephenson C, Omrani M, Gerritsen C, Martin MS, Knyahnytskyi A, et al. Delivering an online cognitive behavioral therapy program to address mental health challenges faced by correctional workers and other public safety personnel: protocol for a mixed methods study. JMIR Res Protoc 2021;10(7):e30845 [FREE Full text] [doi: 10.2196/30845] [Medline: 34088656]
22. Alavi N, Stephenson C, Rivera M. Effectiveness of delivering dialectical behavioral therapy techniques by email in patients with borderline personality disorder: nonrandomized controlled trial. JMIR Ment Health 2021;8(4):e27308 [FREE Full text] [doi: 10.2196/27308] [Medline: 33835936]
23. Wahlund T, Mataix-Cols D, Lauri KO, de Schipper E, Ljótsson B, Aspvall K, et al. Brief online cognitive behavioural intervention for dysfunctional worry related to the COVID-19 pandemic: a randomised controlled trial. Psychother Psychosom 2021;90(3):191-199 [FREE Full text] [doi: 10.1159/000512843] [Medline: 33212440]

24. Lenferink LIM, Meyerbröker K, Boelen PA. PTSD treatment in times of COVID-19: a systematic review of the effects of online EMDR. Psychiatry Res 2020;293:113438 [FREE Full text] [doi: 10.1016/j.psychres.2020.113438] [Medline: 32905864]
25. Zhang C, Yang L, Liu S, Xu Y, Zheng H, Zhang B. One-week self-guided internet cognitive behavioral treatments for insomnia in adults with situational insomnia during the COVID-19 outbreak. Front Neurosci 2020;14:622749 [FREE Full text] [doi: 10.3389/fnins.2020.622749] [Medline: 33551732]
26. Mahoney A, Li I, Haskelberg H, Millard M, Newby JM. The uptake and effectiveness of online cognitive behaviour therapy for symptoms of anxiety and depression during COVID-19. J Affect Disord 2021;292:197-203 [FREE Full text] [doi: 10.1016/j.jad.2021.05.116] [Medline: 34126311]
27. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012;12(7):861-869; quiz 870 [FREE Full text] [doi: 10.1586/ern.12.63] [Medline: 22853793]
28. Alavi N, Yang M, Stephenson C, Nikjoo N, Malakouti N, Layzell G, et al. Using the online psychotherapy tool to address mental health problems in the context of the COVID-19 pandemic: protocol for an electronically delivered cognitive behavioral therapy program. JMIR Res Protoc 2020;9(12):e24913 [FREE Full text] [doi: 10.2196/24913] [Medline: 33290245]
29. Hofmann SG, Gómez AF. Mindfulness-based interventions for anxiety and depression. Psychiatr Clin North Am 2017;40(4):739-749 [FREE Full text] [doi: 10.1016/j.psc.2017.08.008] [Medline: 29080597]
30. Parsons CE, Crane C, Parsons LJ, Fjorback LO, Kuyken W. Home practice in mindfulness-based cognitive therapy and mindfulness-based stress reduction: a systematic review and meta-analysis of participants' mindfulness practice and its association with outcomes. Behav Res Ther 2017;95:29-41 [FREE Full text] [doi: 10.1016/j.brat.2017.05.004] [Medline: 28527330]
31. D'Zurilla TJ, Neuz AM. Problem-solving therapy. In: Dobson KS, editor. Handbook of Cognitive-Behavioral Therapies, Third Edition. New York: Guilford Press; 2009:197-225.
32. Kleiboer A, Donker T, Seekles W, van Straten A, Riper H, Cuijpers P. A randomized controlled trial on the role of support in internet-based problem solving therapy for depression and anxiety. Behav Res Ther 2015;72:63-71. [doi: 10.1016/j.brat.2015.06.013] [Medline: 26188373]
33. Greenberger D, Padesky CA. Mind Over Mood: Change How You Feel by Changing the Way You Think, Second Edition. New York: Guilford Publications; 2015.
34. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder screener (GAD-7) in the general population. Med Care 2008;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
35. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019;365:l1476 [FREE Full text] [doi: 10.1136/bmj.l1476] [Medline: 30967483]
36. Miroševič Š, Klemenc-Ketiš Z, Selič P. The 14-item resilience scale as a potential screening tool for depression/anxiety and quality of life assessment: a systematic review of current research. Fam Pract 2019;36(3):262-268 [FREE Full text] [doi: 10.1093/fampra/cmy081] [Medline: 30165643]
37. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. Psychopharmacol Bull 1993;29(2):321-326. [Medline: 8290681]
38. Taylor CB, Graham AK, Flatt RE, Waldherr K, Fitzsimmons-Craft EE. Current state of scientific evidence on internet-based interventions for the treatment of depression, anxiety, eating disorders and substance abuse: an overview of systematic reviews and meta-analyses. Eur J Public Health 2021;31(Suppl 1):i3-i10 [FREE Full text] [doi: 10.1093/eurpub/ckz208] [Medline: 32918448]
39. Summers C, Wu P, Taylor AJG. Supporting mental health during the COVID-19 pandemic using a digital behavior change intervention: an open-label, single-arm, pre-post intervention study. JMIR Form Res 2021;5(10):e31273 [FREE Full text] [doi: 10.2196/31273] [Medline: 34459740]
40. Musiat P, Johnson C, Atkinson M, Wilksch S, Wade T. Impact of guidance on intervention adherence in computerised interventions for mental health problems: a meta-analysis. Psychol Med 2022;52(2):229-240. [doi: 10.1017/S0033291721004621] [Medline: 34802474]
41. Wong LP, Alias H, Fuzi AAM, Omar IS, Nor AM, Tan MP, et al. Escalating progression of mental health disorders during the COVID-19 pandemic: evidence from a nationwide survey. PLoS One 2021;16(3):e0248916 [FREE Full text] [doi: 10.1371/journal.pone.0248916] [Medline: 33765039]
42. Aknin LB, De Neve JE, Dunn EW, Fancourt DE, Goldberg E, Helliwell JF, et al. Mental health during the first year of the COVID-19 pandemic: a review and recommendations for moving forward. Perspect Psychol Sci 2022;17(4):915-936 [FREE Full text] [doi: 10.1177/17456916211029964] [Medline: 35044275]
43. Robinson E, Sutin AR, Daly M, Jones A. A systematic review and meta-analysis of longitudinal cohort studies comparing mental health before versus during the COVID-19 pandemic in 2020. J Affect Disord 2022;296:567-576 [FREE Full text] [doi: 10.1016/j.jad.2021.09.098] [Medline: 34600966]

44. Talevi D, Socci V, Carai M, Carnaghi G, Faleri S, Trebbi E, et al. Mental health outcomes of the COVID-19 pandemic. Riv Psichiatr 2020;55(3):137-144 [FREE Full text]
45. Nandasena HMRKG, Pathirathna ML, Atapattu AMMP, Prasanga PTS. Quality of life of COVID 19 patients after discharge: systematic review. PLoS One 2022;17(2):e0263941 [FREE Full text] [doi: 10.1371/journal.pone.0263941] [Medline: 35171956]
46. Panayiotou G, Panteli M, Leonidou C. Coping with the invisible enemy: the role of emotion regulation and awareness in quality of life during the COVID-19 pandemic. J Context Behav Sci 2021;19:17-27 [FREE Full text] [doi: 10.1016/j.jcbs.2020.11.002]
47. Kasar KS, Karaman E. Life in lockdown: social isolation, loneliness and quality of life in the elderly during the COVID-19 pandemic: a scoping review. Geriatr Nurs 2021;42(5):1222-1229 [FREE Full text] [doi: 10.1016/j.gerinurse.2021.03.010] [Medline: 33824008]
48. Mukherjee S, Pahan K. Is COVID-19 gender-sensitive? J Neuroimmune Pharmacol 2021;16(1):38-47 [FREE Full text] [doi: 10.1007/s11481-020-09974-z] [Medline: 33405098]

Abbreviations

CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
e-CBT: Electronic cognitive behavioral therapy
e-psychotherapy: Electronically delivered psychological interventions
GAD: generalized anxiety disorder
GAD-7: 7-item Generalized Anxiety Disorder Questionnaire
MDD: major depressive disorder
MINI: Mini-International Neuropsychiatrist Interview
OPTT: online psychotherapy tool
PHQ-9: 9-item Patient Health Questionnaire
PTSD: posttraumatic stress disorder
Q-LES-Q: Quality of Life Enjoyment and Satisfaction Questionnaire
RCT: randomized controlled trial
RS-14: 14-item resilience scale
SMART: Specific, Measurable, Achievable, Realistic, and Timely

Edited by J Torous; submitted 20.07.23; peer-reviewed by A AL-Asadi, N Esfandiari; comments to author 17.11.23; revised version received 17.11.23; accepted 21.11.23; published 25.12.23.

Please cite as:
Moghimi E, Stephenson C, Agarwal A, Nikjoo N, Malakouti N, Layzell G, O'Riordan A, Jagayat J, Shirazi A, Gutierrez G, Khan F, Patel C, Yang M, Omrani M, Alavi N
Efficacy of an Electronic Cognitive Behavioral Therapy Program Delivered via the Online Psychotherapy Tool for Depression and Anxiety Related to the COVID-19 Pandemic: Pre-Post Pilot Study
JMIR Ment Health 2023;10:e51102
URL: https://mental.jmir.org/2023/1/e51102
doi:10.2196/51102
PMID:37993984

Original Paper

Assessing Detection of Children With Suicide-Related Emergencies: Evaluation and Development of Computable Phenotyping Approaches

Juliet Beni Edgcomb[1,2], MD, PhD; Chi-hong Tseng[3], PhD; Mengtong Pan[3], BA, BS; Alexandra Klomhaus[3], PhD; Bonnie T Zima[1,2], MPH, MD

[1]Mental Health Informatics and Data Science (MINDS) Hub, Center for Community Health, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, United States

[2]Department of Psychiatry, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States

[3]Department of Medicine Statistics Core, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Juliet Beni Edgcomb, MD, PhD
Mental Health Informatics and Data Science (MINDS) Hub, Center for Community Health
Semel Institute for Neuroscience and Human Behavior
University of California Los Angeles
760 Westwood Plaza
Los Angeles, CA, 90095
United States
Phone: 1 310 794 8278
Email: jedgcomb@mednet.ucla.edu

Abstract

Background: Although suicide is a leading cause of death among children, the optimal approach for using health care data sets to detect suicide-related emergencies among children is not known.

Objective: This study aimed to assess the performance of suicide-related International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes and suicide-related chief complaint in detecting self-injurious thoughts and behaviors (SITB) among children compared with clinician chart review. The study also aimed to examine variations in performance by child sociodemographics and type of self-injury, as well as develop machine learning models trained on codified health record data (features) and clinician chart review (gold standard) and test model detection performance.

Methods: A gold standard classification of suicide-related emergencies was determined through clinician manual review of clinical notes from 600 emergency department visits between 2015 and 2019 by children aged 10 to 17 years. Visits classified with nonfatal suicide attempt or intentional self-harm using the Centers for Disease Control and Prevention surveillance case definition list of ICD-10-CM codes and suicide-related chief complaint were compared with the gold standard classification. Machine learning classifiers (least absolute shrinkage and selection operator–penalized logistic regression and random forest) were then trained and tested using codified health record data (eg, child sociodemographics, medications, disposition, and laboratory testing) and the gold standard classification. The accuracy, sensitivity, and specificity of each detection approach and relative importance of features were examined.

Results: SITB accounted for 47.3% (284/600) of the visits. Suicide-related diagnostic codes missed nearly one-third (82/284, 28.9%) and suicide-related chief complaints missed more than half (153/284, 53.9%) of the children presenting to emergency departments with SITB. Sensitivity was significantly lower for male children than for female children (0.69, 95% CI 0.61-0.77 vs 0.84, 95% CI 0.78-0.90, respectively) and for preteens compared with adolescents (0.66, 95% CI 0.54-0.78 vs 0.86, 95% CI 0.80-0.92, respectively). Specificity was significantly lower for detecting preparatory acts (0.68, 95% CI 0.64-0.72) and attempts (0.67, 95% CI 0.63-0.71) than for detecting ideation (0.79, 95% CI 0.75-0.82). Machine learning–based models significantly improved the sensitivity of detection compared with suicide-related codes and chief complaint alone. Models considering all 84 features performed similarly to models considering only mental health–related ICD-10-CM codes and chief complaints (34 features) and models considering non–ICD-10-CM code indicators and mental health–related chief complaints (53 features).

Conclusions: The capacity to detect children with SITB may be strengthened by applying a machine learning–based approach to codified health record data. To improve integration between clinical research informatics and child mental health care, future

research is needed to evaluate the potential benefits of implementing detection approaches at the point of care and identifying precise targets for suicide prevention interventions in children.

(JMIR Ment Health 2023;10:e47084) doi:10.2196/47084

KEYWORDS

child mental health; suicide; self-harm; machine learning; phenotyping

Introduction

Background

In the United States, suicide is the second leading cause of death among children aged 10 to 14 years, and 1 in 13 children attempts suicide before adulthood [1,2]. Emergency departments are often the first point of access to mental health care for children at risk for suicide, and >1.12 million pediatric emergency department visits each year are suicide related [3-5]. Emergency department visits for self-harm among children tripled between 2007 and 2016 [6], and visits for suicide attempts further increased during the pandemic, particularly among girls and older children [7]. The concurrent rapid growth of health informatics has brought promise that comprehensive clinical data from health records can be used to detect care for suicide-related emergencies in a timely and accurate manner [8-10]. However, the optimal approach to detecting childhood-onset self-injurious thoughts and behaviors (SITB) using health record data remains unknown.

Medical records provide an expanding repository of clinical and phenotypic data to enable low-cost population-based studies on a large scale [11] and inform targeted point-of-care interventions [12]. The discovery of individuals with specific health conditions from within health record data sets historically relied on laborious and time-intensive manual chart review [13]. In recent years, algorithms to classify child psychiatric disorders and adverse childhood experiences have demonstrated the capacity to distinguish cases from noncases using semiautomated approaches to structured codified data (eg, demographics, diagnostic codes, and medications) and text mining with natural language processing [14,15]. Phenotype algorithms currently exist for many childhood-onset mental health conditions, including pediatric depression [16], anxiety [17], developmental language disorder [18], attention-deficit/hyperactivity disorder [15], and autism [19], as well as general pediatric conditions such as Crohn's disease [20], sepsis [21], leukemia and lymphoma [22], and pulmonary hypertension [23].

Nevertheless, little is known about whether the detection of suicide-related emergency department visits using medical record data can be improved through the development and application of phenotype algorithms. Children experience heterogenous manifestations of suicidal thoughts and behaviors across the developmental continuum, and the codified health data elements that differentiate children with SITB from those without are not well characterized. Most surveillance applications exclude children or combine children with adults [24-26]. Trade-offs in current approaches to detecting SITB in children are likely but remain unmeasured. For example, whether suicide-related International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), codes and suicide-related chief complaints are sufficiently sensitive and specific in detecting SITB in childhood. Machine learning–based approaches have supported the generation of other clinical phenotypes informative for predicting prognosis, enhancing clinical monitoring, detecting comorbid developmental conditions, and selecting effective treatments [27]. However, the relative benefits of using these approaches are not known for childhood-onset SITB. A recent study distinguishing children with suicidal thoughts and behaviors from those without used data from the Adolescent Brain Cognitive Development study and identified factors difficult to capture using health records: prodromal psychosis, family conflict, depression severity, and impulsivity [28]. Although there is increasing recognition of disparities in predicting suicide events using health records [29], variation in the accuracy of detection of SITB across pediatric population strata (sex, age, race, and ethnicity) remains scarcely described. Knowing which children with SITB are missed by existing approaches could inform efforts to improve detection in an equitable manner and mitigate inequity in the targeted identification of suicide precursors.

Objectives

To address the aforementioned gaps, the study objectives were to (1) compare the detection performance of suicide-related ICD-10-CM codes and chief complaint with that of clinician manual chart review, (2) examine variations in the detection performance by child sociodemographics and type of SITB (suicidal thoughts, preparatory acts, suicide attempt, and nonsuicidal self-injury), and (3) sequentially train and test a series of phenotype algorithms (machine learning classifiers) to detect SITB using codified health record data of varying complexity.

Methods

Design

This was a cross-sectional observational study of emergency department visits by children aged 10 to 17 years. The primary outcome was the classification of the presence or absence of SITB at the emergency department visit. The classification performance of codified medical record data (structured data elements) was compared with that of expert classification by clinician manual chart review of medical records. Algorithmic detection considering three sets of structured data elements was compared with detection considering suicide-related ICD-10-CM codes and suicide-related chief complaint alone (comparator) and chart review (gold standard): (1) mental health–related codes and mental health–related chief complaints, (2) suicide-related codes and non–ICD-10-CM code data elements (ie, other sociodemographic and clinical characteristics of the child), and (3) all structured data elements.

The study followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement guidelines.

Ethics Approval

The study was approved by the University of California Los Angeles institutional review board (20-001512).

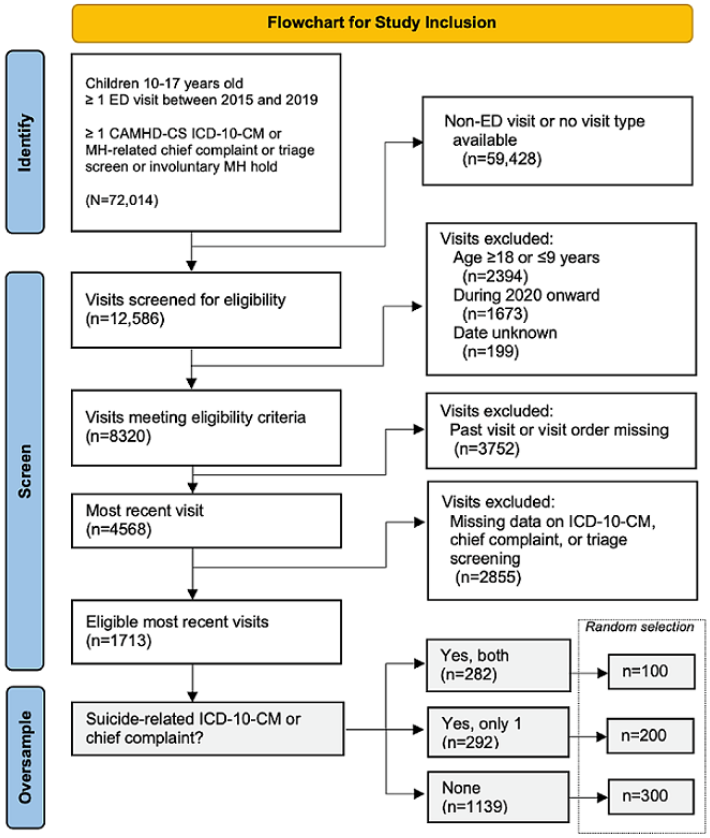
Data Source

The data source was a large university hospital health system comprising 4 hospitals (1 pediatric, 2 medical, and 1 psychiatric) across 2 sites (a tertiary academic medical center and a community hospital). For each child meeting the inclusion criteria, all emergency department medical records were delivered to the study team by the Clinical and Translational Science Institute from the Integrated Clinical and Research Data Repository, a large-scale clinical data warehouse that supports data analyses and extractions for research. The academic medical center site is a primary teaching hospital in Los Angeles, California. This site includes a colocated affiliated children's hospital and an independently accredited psychiatric hospital with 3 inpatient child psychiatric units serving children with mental illnesses and developmental disabilities. The academic medical center is staffed 24/7 with child and adolescent psychiatrists and general psychiatrists. The community hospital is affiliated with a 25-bed general inpatient pediatric ward. At the community hospital site, children with acute psychiatric complaints are seen by emergency department physicians and licensed clinical social workers.

Sampling

The flowchart of study inclusion is presented in Figure 1. A series of selection rules were applied to yield a sample feasible for chart review (n=600) and consistent with judicious oversampling informative cases [30]. Visits were restricted to the most recent mental health–related emergency department visit by each child, occurring between October 1, 2015, and October 1, 2019, and defined as emergency department encounters associated with (1) one or more diagnostic code as defined by the Child and Adolescent Mental Health Disorders Classification System (CAMHD-CS) [31]; (2) a mental health–related chief complaint; (3) a positive response to the triage screening question; "Does this patient have a primary psychiatric complaint or suspicion of psychiatric illness?"; or (4) an involuntary mental health detainment order. The final sample was intentionally structured to approximate an equal distribution of 50% cases and 50% noncases. Consequently, from the pool of children who met the inclusion criteria (n=1713), we randomly selected (1) a total of 35.4% (100/282) of children who had both a suicide-related code and a chief complaint, (2) a total of 68.5% (200/292) of children with either a suicide-related code or a chief complaint, and (3) a total of 26.3% (300/1139) of children with neither a suicide-related ICD-10-CM code nor suicide-related a chief complaint. Given the rigorous sampling strategy, a statistical comparison was conducted between the eligible children and those included in the study, and the results are presented in Multimedia Appendix 1. The only significant difference observed was a marginally higher representation of Hispanic or Latinx children in the final sample (27% compared with 24% in the sample of eligible children; P=.02).

Figure 1. Flowchart of study inclusion. CAMHD-CS: Child and Adolescent Mental Health Disorders Classification System; ED: emergency department; ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification; MH: mental health.



Study Variable Construction

Sociodemographics included child age, natal sex, race, and ethnicity. These variables were self- or parent-reported at the point of care. Socioeconomic disadvantage was assessed using the Area Deprivation Index (ADI) [32]. The Federal Information Processing System (FIPS) code of each child's home address was linked to the ADI with decile ranked at the state and national levels. The only variable for which missing values were present was ADI (missing for 56/600, 9% children), and missing values for ADI were imputed through corresponding medians. Additional structured data indicators were considered (eg, gender identity, family history, and language) but omitted owing to sparsity.

Clinical characteristics included diagnostic or billing codes, chief complaint, orders (medications, laboratory tests, and involuntary hold status), site (academic medical center vs community hospital), and prior care use. All mental health–related diagnostic or billing codes (ICD-10-CM) from the emergency department visit were categorized using the CAMHD-CS [31]. The presence of an ICD-10-CM code for SITB was determined by the presence of one or more codes from the Centers for Disease Control and Prevention (CDC) surveillance case definition list [24] and associated with the emergency department visit. Of note, the codes used to assign the CAMHD-CS category of suicide or self-injury align exactly with the CDC code list. The chief complaint for SITB was determined by the selection of *suicidal* or *suicide attempt* by nursing triage upon the child's arrival at the emergency department. Laboratory tests were restricted to those ordered and collected during the emergency department visit and included those related to overdose (serum acetaminophen, salicylates, benzodiazepines, and tricyclics), urine drug screen results, and serum alcohol. All psychotropic medications (n=97) received during the visit were consolidated using the Anatomical Therapeutic Chemical classification system into 8 categories (antidepressants, antiepileptics, antihistamines, antipsychotics, anxiolytics, hypnotics and sedatives, lithium, and psychostimulants). Additional clinical characteristics were encounter year, site, emergency department disposition, provider sex, as well as the child's number of prior 90-, 180-, and 365-day emergency department visits and general medical and psychiatric hospitalizations. A full list of sociodemographic and clinical characteristics and definitions are included in Multimedia Appendix 2.

Manual Chart Abstraction

All clinical notes from each emergency department visit were extracted and provided to the study team verbatim. The notes included physician history and physical examinations, progress notes, social work notes, and nursing notes.

Classification by the manual review of records was adapted from the Columbia Classification Algorithm of Suicide Assessment (C-CASA) [33]. The C-CASA is a system for categorizing suicide-related behavior that takes into account research-based definitions of suicidality and has been applied to the classification of emergency presentations for children [34]. The criteria for defining a suicide attempt include both self-harm and intent to die [33]. Including intent in the definition of suicide helps distinguish between those who engage in

self-harm with the intent to die and those who do so for other reasons. The C-CASA has 8 categories that differentiate among suicidal behavior, nonsuicidal behavior, and behavior that is potentially suicidal.

Consistent with operationalized guidelines for C-CASA, if >1 category was present, the abstractor coded the visit as consistent with the most severe category: suicide death, nonfatal attempt, preparatory behavior, suicidal ideation, self-injurious behavior intent unknown, not enough information, and self-injurious behavior without suicidal intent [33]. To capture cases with combined nonsuicidal self-injurious behavior and suicidal thoughts or behaviors, the classification system was adapted to specify the presence and type of self-injurious behavior (intent unknown or no suicidal intent) in a secondary classification field.

Classifications were compared and differed in only 0.3% (2/600) of the cases for presence or absence of SITB. Classifications differed in 2.5% (15/600) of the cases for type of SITB. In the second stage, a second board-certified child psychiatrist (BZ) and child psychiatric nurse practitioner (KC), also blinded, separately reviewed all discordant cases. Of the discordant cases for which concordance was not reached (4/600, 0.6%), consensus discussion yielded a final classification.

Analyses

Rule-Based Classification

Contingency matrixes were constructed comparing classification with suicide-related ICD-10-CM code and suicide-related chief complaint (comparator) and manual chart review (gold standard). The sensitivity, specificity, and accuracy were calculated, with 95% CIs computed using Clopper-Pearson CIs. Variations in performance by demographics were examined by subsetting the sample by demographic characteristics (eg, male children). Variations in detection performance for type of SITB (eg, suicidal ideation) were examined by comparing classification using structured data elements with classification of type upon manual chart review.

Machine Learning–Based Classification

Fit metrics were measured via 10-fold cross-validation. For each fold, a machine learning model was trained with structured data elements (features) and the manual chart review (gold standard) for each child in a training set. Next, this model was used to classify the presence or absence of SITB (predicted outcome) of each child in a test set, and this predicted outcome was compared with the manual chart review (gold standard) to yield fit metrics. CIs for fit metrics were calculated by examining the variations in fit metrics across the test sets.

Three sets of structured data elements were compared, representing varying levels of complexity of codified health record data: (1) mental health–related ICD-10-CM codes and mental health–related chief complaints (34 features); (2) suicide-related ICD-10-CM codes and all child sociodemographics and clinical characteristics, excluding mental health–related ICD-10-CM codes (53 features); and (3) all structured data elements (84 features). The first set was chosen to evaluate classification performance using mental health–related ICD-10-CM codes and chief complaints to detect cases. The second set was used to determine the relative importance of considering other, non–ICD-10-CM–based structured data elements (ie, how well detection can be performed without mental health comorbidity codes). Variables in the first and second sets are mutually exclusive, except for suicide-related ICD-10-CM codes and chief complaints, which are included in both sets. The third set was used to comprehensively evaluate the structured data elements that might support the detection of cases and to test the optimization of detection using a broad set of codified data.

Two classifier types were compared: least absolute shrinkage and selection operator (LASSO)–penalized logistic regression (hereinafter referred to as LASSO) and random forest. LASSO was selected to perform variable selection and yield a parsimonious model involving only a subset of variables relevant to the classification task [35]. Random forest was selected to stratify the predictor space and produce a consensus prediction using an ensemble of decision trees [36]. LASSO and random forest were selected because both are well documented in the informatics literature and widely used for phenotyping applications [37]. Fit metrics were compared using McNemar chi-square tests. The classifiers were anticipated to have predictive ability, with accuracy ranging from 70% to 95%. Given the study sample size, the margin of error was estimated to be <4%.

Feature engineering was conducted using R statistical software (version 4.2.0; R Foundation for Statistical Computing), and the models were implemented using Python (version 3.12; Python Software Foundation) with *scikit-learn* (version 1.2.2) toolboxes *sklearn.linear_model.lasso*, *sklearn.ensemble.RandomForestClassifier*, and *sklearn.metrics*. Hyperparameters were set to default and were as follows: LASSO-penalized logistic regression (L1 penalty, *liblinear* solver, and regularization score 1.0) and random forest (100 trees, bootstrap samples, Gini impurity for tree split quality, and no balancing or class weights). The random forest was run with out-of-bag samples to estimate generalization error. A set seed was used to ensure replicability. The code is available from the authors upon request.

Sampling Probability Adjustment

As the study population was a stratified random subsample of the total population, we compared rule-based classification fit metrics, both with and without the adjustment for sampling probability. The adjustment was performed by considering the subsample as a stratified 2-phase sample and applying inverse probability weighting. Further detail on this method is described by Katki et al [30].

Results

Sample Characteristics

Child sociodemographics and clinical characteristics are presented in Table 1. Additional sample characteristics are described in Multimedia Appendices 3 and 4.

Table 1. Sample characteristics (n=600).

	Values, n (%)
Sex	
Male	276 (46)
Female	324 (54)
Age group (years)	
10-12.9	115 (19.2)
13-15.9	215 (35.8)
16-17.9	270 (45)
Race	
American Indian or Alaska Native	2 (0.3)
Asian	35 (5.8)
Black or African American	61 (10.2)
Native Hawaiian or other Pacific Islander	0 (0)
White	323 (53.8)
Other[a]	127 (21.1)
Ethnicity	
Hispanic or Latinx	161 (26.8)
Not Hispanic or Latinx	390 (65)
Other[a]	3 (0.5)
State ADI[b] decile	
1-3	333 (55.5)
4-6	119 (19.8)
7-10	92 (15.3)
Missing	56 (9.3)
Site	
Academic medical center	455 (75.8)
Community hospital	145 (24.2)
Disposition	
Discharged without hospitalization[c]	322 (53.7)
General medical hospitalization	106 (17.7)
Psychiatric hospitalization	
Within health system	134 (22.3)
Transferred outside health system	38 (6.3)
Legal status	
72-hour hold (involuntary)	123 (20.5)
Voluntary	477 (79.5)
Chief complaint	
Psychiatric (including suicide related)	370 (61.7)
Suicide related	131 (21.8)
Other	227 (37.8)
Top 10 diagnostic code groups[d]	
Depressive disorders	221 (36.8)

	Values, n (%)
Suicide or self-injury	203 (33.8)
Anxiety disorders	181 (30.2)
Attention deficit hyperactivity disorder	105 (17.5)
Substance-related and addictive disorders	80 (13.3)
Mental health symptom	76 (12.7)
Autism spectrum disorder	59 (9.8)
Disruptive, impulse control, and conduct disorders	35 (5.8)
Obsessive-compulsive and related disorders	33 (5.5)
Trauma and stressor-related disorders	32 (5.3)
Bipolar and related disorders	24 (4)

[a]Multiple races, not available, other, patient refused, or unknown.

[b]ADI: Area Deprivation Index.

[c]Eloped (4/322, 1.2%), left without being seen (2/322, 0.6%), left against medical advice (2/322, 0.6%), inpatient rehabilitation facility (3/422, 0.9%), law enforcement (1/322, 0.3%), skilled nursing (1/322, 0.3%), and expired (3/322, 0.9%).

[d]Ten most prevalent Child and Adolescent Mental Health Disorders Classification System (CAMHD-CS) diagnostic code groups, in order of prevalence in study sample.

Performance of Rule-Based Classification

The detection performance of suicide-related ICD-10-CM codes and chief complaints compared with that of manual chart review is presented in Table 2. Manual chart review labeled 47.3% (284/600) of the visits as consistent with SITB (gold standard positive). Classification using suicide-related codes alone resulted in 85 false negatives with sensitivity 0.70, specificity 0.99, and accuracy 0.85. Classification using suicide-related chief complaint alone resulted in 155 false negatives with sensitivity 0.45, specificity 0.99, and accuracy 0.74. The highest misclassification was observed if a suicide-related code and a suicide-related chief complaint were necessary to classify the visit as SITB positive (sensitivity 0.38, specificity 1.00, and accuracy 0.71). The lowest misclassification rate was observed if either a suicide-related code or a suicide-related chief complaint classified the visit as SITB positive (sensitivity 0.77, specificity 0.98, and accuracy 0.89). The sensitivity of suicide-related codes and suicide-related chief complaints (either affirmed) was significantly lower among male children (0.69, 95% CI 0.61-0.77) than among female children (0.84, 95% CI 0.78-0.90). Sensitivity was also significantly lower in detecting cases of SITB among those aged 10 to 12 years (0.66, 95% CI 0.54-0.78) than among those aged 13 to 15 years (0.86, 95% CI 0.80-0.92). Differences in fit metrics by race and ethnicity did not reach statistical significance. There were no substantial differences between adjusted and unadjusted fit metrics, and sampling probability–adjusted estimates are included in Multimedia Appendix 5.

Detection performance by type of SITB is presented in Table 3. The sensitivity of detection did not differ by type. For suicide-related codes and suicide-related chief complaints (either affirmed), the specificity of detection was significantly lower for preparatory acts (0.68, 95% CI 0.64-0.72) and suicide attempts (0.67, 95% CI 0.63-0.71) than for suicidal ideation (0.79, 95% CI 0.75-0.82). There were no substantial differences between adjusted and unadjusted fit metrics, and sampling probability–adjusted estimates are included in Multimedia Appendix 6.

Table 2. Performance of International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), code (as defined by the Centers for Disease Control and Prevention case surveillance definition list) and suicide-related chief complaint in detecting cases of self-injurious thoughts and behaviors compared with that of manual chart abstraction: total sample and stratified by natal sex, age group, race, and ethnicity.

Sample and classification	True positive, n (%)	False positive, n (%)	False negative, n (%)	True negative, n (%)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
All (n=600)							
ICD-10-CM	199 (33.2)	4 (0.7)	85 (14.2)	312 (52)	0.70 (0.65-0.75)	0.99 (0.98-1.00)	0.85 (0.82-0.88)
CC[a,b]	129 (21.5)	2 (0.3)	155 (25.8)	314 (52.3)	0.45 (0.40-0.51)	0.99 (0.98-1.00)	0.74 (0.70-0.77)
ICD-10-CM or CC[c]	220 (36.7)	5 (0.8)	64 (10.7)	311 (51.8)	0.77 (0.73-0.82)	0.98 (0.97-1.00)	0.89 (0.86-0.91)
ICD-10-CM and CC[d]	108 (18)	1 (0.2)	176 (29.3)	315 (52.5)	0.38 (0.32-0.44)	1.00 (0.99-1.00)	0.71 (0.67-0.74)
Sex							
Male (n=276)							
ICD-10-CM	78 (28.3)	0 (0)	45 (16.3)	153 (55.4)	0.63 (0.55-0.72)	1.00 (1.00-1.00)	0.84 (0.79-0.88)
CC	52 (18.8)	0 (0)	71 (25.7)	153 (55.4)	0.42 (0.34-0.51)	1.00 (1.00-1.00)	0.74 (0.69-0.79)
ICD-10-CM or CC	85 (30.8)	0 (0)	38 (13.8)	153 (55.4)	0.69 (0.61-0.77)	1.00 (1.00-1.00)	0.86 (0.82-0.90)
ICD-10-CM and CC	45 (16.3)	0 (0)	78 (28.3)	153 (55.4)	0.37 (0.28-0.45)	1.00 (1.00-1.00)	0.72 (0.66-0.77)
Female (n=324)							
ICD-10-CM	121 (37.3)	4 (1.2)	40 (12.3)	159 (49.1)	0.75 (0.68-0.82)	0.98 (0.95-1.00)	0.86 (0.83-0.90)
CC	77 (23.8)	2 (0.6)	84 (25.9)	161 (49.7)	0.48 (0.40-0.56)	0.99 (0.97-1.00)	0.73 (0.69-0.78)
ICD-10-CM or CC	135 (41.7)	5 (1.5)	26 (8)	158 (48.8)	0.84 (0.78-0.90)	0.97 (0.94-1.00)	0.90 (0.87-0.94)
ICD-10-CM and CC	63 (19.4)	1 (0.3)	98 (30.2)	162 (50)	0.39 (0.32-0.47)	0.99 (0.98-1.00)	0.69 (0.64-0.74)
Age group (years)							
10-12.9 (n=115)							
ICD-10-CM	33 (28.7)	0 (0)	26 (22.6)	56 (48.7)	0.56 (0.43-0.69)	1.00 (1.00-1.00)	0.77 (0.70-0.85)
CC	25 (21.7)	0 (0)	34 (29.6)	56 (48.7)	0.42 (0.30-0.55)	1.00 (1.00-1.00)	0.70 (0.62-0.79)
ICD-10-CM or CC	39 (33.9)	0 (0)	20 (17.4)	56 (48.7)	0.66 (0.54-0.78)	1.00 (1.00-1.00)	0.83 (0.76-0.90)
ICD-10-CM and CC	19 (16.5)	0 (0)	40 (34.8)	56 (48.7)	0.32 (0.20-0.44)	1.00 (1.00-1.00)	0.65 (0.57-0.74)
13-15.9 (n=215)							
ICD-10-CM	90 (41.9)	1 (0.5)	26 (12.1)	98 (45.6)	0.78 (0.70-0.85)	0.99 (0.97-1.00)	0.87 (0.83-0.92)
CC	57 (26.5)	1 (0.5)	59 (27.4)	98 (45.6)	0.49 (0.40-0.58)	0.99 (0.97-1.00)	0.72 (0.66-0.78)
ICD-10-CM or CC	100 (46.5)	2 (0.9)	16 (7.4)	97 (45.1)	0.86 (0.80-0.92)	0.98 (0.95-1.00)	0.92 (0.88-0.95)
ICD-10-CM and CC	47 (21.9)	0 (0)	69 (32.1)	99 (46)	0.41 (0.32-0.49)	1.00 (1.00-1.00)	0.68 (0.62-0.74)
16-17.9 (n=270)							
ICD-10-CM	76 (28.1)	3 (1.11)	33 (1.5)	158 (58.5)	0.70 (0.61-0.78)	0.98 (0.96-1.00)	0.87 (0.83-0.91)
CC	47 (17.4)	1 (0.4)	62 (23)	160 (59.3)	0.43 (0.34-0.52)	0.99 (0.98-1.00)	0.77 (0.72-0.82)
ICD-10-CM or CC	81 (30)	3 (1.11)	28 (10.4)	158 (58.5)	0.74 (0.66-0.83)	0.98 (0.96-1.00)	0.89 (0.85-0.92)
ICD and CC	42 (15.6)	1 (0.4)	67 (24.8)	160 (59.3)	0.39 (0.29-0.48)	0.99 (0.98-1.00)	0.75 (0.70-0.80)
Race and ethnicity							
Asian, non-Hispanic (n=35)							
ICD-10-CM	12 (34.3)	0 (0)	3 (8.6)	20 (57.1)	0.8 (0.6-1.00)	1.00 (1.00-1.00)	0.82 (0.91-1.00)
CC	5 (14.3)	0 (0)	10 (28.6)	20 (57.1)	0.33 (0.09-0.57)	1.00 (1.00-1.00)	0.56 (0.71-0.86)
ICD-10-CM or CC	13 (37.1)	0 (0)	2 (5.7)	20 (57.1)	0.87 (0.69-1.00)	1.00 (1.00-1.00)	0.87 (0.94-1.00)
ICD-10-CM and CC	4 (11.4)	0 (0)	11 (31.4)	20 (57.1)	0.27 (0.04-0.49)	1.00 (1.00-1.00)	0.69 (0.53-0.84)
Black, non-Hispanic (n=61)							

Sample and classification	True positive, n (%)	False positive, n (%)	False negative, n (%)	True negative, n (%)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
ICD-10-CM	21 (34.4)	0 (0)	10 (16.4)	30 (49.2)	0.68 (0.51-0.84)	1.00 (1.00-1.00)	0.84 (0.74-0.93)
CC	12 (19.7)	0 (0)	19 (31.1)	30 (49.2)	0.39 (0.22-0.56)	1.00 (1.00-1.00)	0.57 (0.69-0.80)
ICD-10-CM or CC	23 (37.7)	0 (0)	8 (13.1)	30 (49.2)	0.74 (0.59-0.90)	1.00 (1.00-1.00)	0.78 (0.87-0.95)
ICD-10-CM and CC	10 (16.4)	0 (0)	21 (34.4)	30 (49.2)	0.32 (0.16-0.49)	1.00 (1.00-1.00)	0.66 (0.54-0.77)
Hispanic or Latinx (n=161)							
ICD-10-CM	44 (27.3)	1 (0.6)	25 (15.5)	91 (56.5)	0.64 (0.52-0.75)	0.99 (0.97-1.00)	0.84 (0.78-0.90)
CC	32 (19.9)	1 (0.6)	37 (23)	91 (56.5)	0.46 (0.35-0.58)	0.99 (0.97-1.00)	0.76 (0.70-0.83)
ICD-10-CM or CC	51 (31.7)	2 (1.2)	18 (11.2)	90 (55.9)	0.74 (0.64-0.84)	0.98 (0.95-1.00)	0.82 (0.88-0.93)
ICD-10-CM and CC	25 (15.5)	0 (0)	44 (27.3)	92 (57.1)	0.36 (0.25-0.48)	1.00 (1.00-1.00)	0.66 (0.73-0.80)
White, non-Hispanic (n=285)							
ICD-10-CM	106 (37.2)	2 (0.7)	42 (14.7)	135 (47.4)	0.72 (0.64-0.79)	0.99 (0.97-1.00)	0.85 (0.80-0.89)
CC	68 (23.9)	1 (0.4)	80 (28.1)	136 (47.7)	0.46 (0.38-0.54)	0.99 (0.98-1.00)	0.72 (0.66-0.77)
ICD-10-CM or CC	116 (40.7)	2 (0.7)	32 (11.2)	135 (47.4)	0.78 (0.72-0.85)	0.99 (0.97-1.00)	0.88 (0.84-0.92)
ICD-10-CM and CC	58 (20.4)	1 (0.4)	90 (31.6)	136 (47.7)	0.39 (0.31-0.47)	0.99 (0.98-1.00)	0.68 (0.63-0.73)
Other[e] (n=58)							
ICD-10-CM	16 (27.6)	1 (1.7)	5 (8.6)	36 (62.1)	0.76 (0.58-0.94)	0.97 (0.92-1.03)	0.82 (0.90-0.97)
CC	12 (20.7)	0 (0)	9 (15.5)	37 (63.8)	0.57 (0.36-0.78)	1.00 (1.00-1.00)	0.75 (0.84-0.94)
ICD-10-CM or CC	17 (29.3)	1 (1.7)	4 (6.9)	36 (62.1)	0.81 (0.64-0.98)	0.97 (0.92-1.03)	0.84 (0.91-0.99)
ICD-10-CM and CC	11 (19)	0 (0)	10 (17.2)	37 (63.8)	0.52 (0.31-0.74)	1.00 (1.00-1.00)	0.73 (0.83-0.92)

[a]CC: chief complaint.

[b]CC refers to suicide-related chief complaints.

[c]Cases classified as self-injurious thoughts and behaviors if either a suicide-related ICD-10-CM code or a suicide-related CC was present (either affirmed).

[d]Cases classified as self-injurious thoughts and behaviors if both a suicide-related ICD-10-CM code and a suicide-related CC were present (both affirmed).

[e]American Indian or Alaska Native, Native Hawaiian or Pacific Islander, multiple races, not available, other, patient refused, and unknown.

Table 3. Performance of International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), code (as defined by the Centers for Disease Control and Prevention case surveillance definition list) and suicide-related chief complaint in detecting cases of self-injurious thoughts and behaviors compared with manual chart abstraction: stratified by Columbia Classification Algorithm of Suicide Assessment categorization (n=600).

Categorization and classification	True positive, n (%)	False positive, n (%)	False negative, n (%)	True negative, n (%)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
Suicidal ideation							
ICD-10-CM: broad[a]	105 (17.5)	98 (16.3)	39 (6.5)	358 (59.7)	0.73 (0.66-0.80)	0.79 (0.75-0.82)	0.77 (0.74-0.81)
ICD-10-CM: strict[b]	104 (17.3)	87 (14.5)	40 (6.7)	369 (61.5)	0.72 (0.65-0.80)	0.81 (0.77-0.85)	0.76 (0.79-0.82)
CC[c,d]	73 (12.2)	58 (9.7)	71 (11.8)	398 (66.3)	0.51 (0.43-0.59)	0.87 (0.84-0.90)	0.75 (0.79-0.82)
ICD-10-CM or CC[e]	118 (19.7)	97 (16.2)	26 (4.3)	359 (59.8)	0.82 (0.76-0.88)	0.79 (0.75-0.82)	0.80 (0.76-0.83)
ICD-10-CM and CC[f]	60 (10)	49 (8.2)	84 (14)	407 (67.8)	0.42 (0.34-0.50)	0.89 (0.86-0.92)	0.78 (0.75-0.81)
Preparatory acts							
ICD-10-CM	35 (5.8)	168 (28)	10 (1.7)	387 (64.5)	0.78 (0.66-0.90)	0.70 (0.66-0.74)	0.70 (0.67-0.74)
CC	23 (3.8)	108 (18)	22 (3.7)	447 (74.5)	0.51 (0.37-0.66)	0.77 (0.81-0.84)	0.78 (0.75-0.82)
ICD-10-CM or CC	36 (6)	179 (29.8)	9 (1.5)	376 (62.7)	0.80 (0.68-0.92)	0.64 (0.68-0.72)	0.69 (0.65-0.72)
ICD-10-CM and CC	20 (3.3)	89 (14.8)	33 (5.5)	458 (76.3)	0.38 (0.25-0.51)	0.81 (0.87-0.84)	0.80 (0.76-0.83)
Suicide attempt							
ICD-10-CM	42 (7)	161 (26.8)	11 (1.8)	386 (64.3)	0.79 (0.68-0.90)	0.71 (0.67-0.74)	0.71 (0.68-0.75)
CC	22 (3.7)	109 (18.2)	31 (5.2)	438 (73)	0.42 (0.28-0.55)	0.80 (0.77-0.83)	0.77 (0.73-0.80)
ICD-10-CM or CC	44 (7.3)	181 (30.2)	9 (1.5)	366 (61)	0.83 (0.73-0.93)	0.67 (0.63-0.71)	0.67 (0.63-0.71)
ICD-10-CM and CC	20 (3.3)	89 (14.8)	33 (5.5)	458 (76.3)	0.38 (0.25-0.51)	0.84 (0.81-0.87)	0.80 (0.77-0.83)
Nonsuicidal self-injurious behavior							
ICD-10-CM	74 (12.3)	129 (21.5)	35 (5.8)	362 (60.3)	0.68 (0.59-0.77)	0.74 (0.70-0.78)	0.73 (0.69-0.76)
CC	47 (7.8)	84 (14)	62 (10.3)	407 (67.8)	0.43 (0.34-0.52)	0.83 (0.80-0.86)	0.76 (0.72-0.79)
ICD-10-CM or CC	81 (13.5)	134 (22.3)	28 (4.7)	357 (59.5)	0.74 (0.66-0.83)	0.73 (0.69-0.77)	0.69 (0.73-0.77)
ICD-10-CM and CC	37 (6.2)	72 (12)	71 (11.8)	420 (70)	0.34 (0.25-0.43)	0.82 (0.85-0.88)	0.73 (0.76-0.80)

[a]The entire Centers for Disease Control and Prevention case surveillance definition International Classification of Diseases, Tenth Revision, Clinical Modification, code list was used.

[b]Only the ICD-10-CM code for suicidal ideation (R45.81) was used.

[c]CC: chief complaint.

[d]CC refers to suicide-related chief complaints

[d]Cases classified as self-injurious thoughts and behaviors if either a suicide-related ICD-10-CM code or a suicide-related CC was present (either affirmed).

[e]Cases classified as self-injurious thoughts and behaviors if both a suicide-related ICD-10-CM code and a suicide-related CC were present (both affirmed).

Performance of Machine Learning–Based Classification

Fit metrics by classifier type and considered features are presented in Table 4. The LASSO and random forest classifiers performed similarly. Classification using only suicide-related codes and suicide-related chief complaints was less sensitive (0.77) and more specific (0.98) than classification using machine learning–based classification (sensitivity 0.84-0.86 and specificity 0.91-0.95). McNemar chi-square tests are presented in Multimedia Appendix 7.

The feature importances of models containing all structured data elements are presented in Figure 2, in descending order of importance, with the top predictors, including ICD-10-CM code for suicide or self-injury, mental health–related chief complaint, suicide-related chief complaint, and ICD-10-CM code for depressive disorders. Some features were identified as similarly important by both LASSO and random forest models (eg, ICD-10-CM code for depressive disorders and ICD-10-CM code for anxiety disorders), whereas other features were identified as important only in 1 model (eg, LASSO: ICD-10-CM code for trauma- and stressor-related disorders and random forest: age and national ADI).

There were significant differences in model performances by number and types of considered features. The sensitivity of detection of the machine learning models that considered all structured data elements was significantly higher than the sensitivity of detection using only suicide-related ICD-10-CM

code and suicide-related chief complaint (LASSO: χ^2_1=20.2, *P*<.001 and random forest: χ^2_1=21.6, *P*<.001). However, the detection sensitivity of the models considering all structured data elements (84 features) was not significantly different from the sensitivity of the models considering a smaller number of features (25 features and 53 features): both models considering mental health–related diagnostic codes and chief complaints (LASSO: χ^2_1=0.3, *P*=.59 and random forest: χ^2_1=0.7, *P*=.39) and the models considering structured data elements other than diagnostic codes (LASSO: χ^2_1=0.6, *P*=.44 and random forest: χ^2_1=0.4, *P*=.51) did not significantly differ in sensitivity from the models considering all data elements. Fit metrics and per-fold feature importances are reported in Multimedia Appendix 8.

Table 4. Comparison of classifier performance of rule-based and machine learning classifiers (n=600), with machine learning classifier threshold set at 0.5.

Classifier and classification	Features	True positive, n (%)	False positive, n (%)	False negative, n (%)	True negative, n (%)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)
Rule-based								
ICD-10-CM[a] or CC[b,c]	2	220 (36.7)	5 (0.8)	64 (10.7)	311 (51.8)	0.77 (0.73-0.82)	0.98 (0.97-1.00)	0.89 (0.86-0.91)
LASSO[d]								
Model 1[e]	34	240 (40)	28 (4.7)	44 (7.3)	288 (48)	0.85 (0.80-0.89)	0.91 (0.88-0.95)	0.88 (0.85-0.91)
Model 2[f]	53	239 (39.8)	30 (5)	45 (7.5)	286 (47.7)	0.84 (0.79-0.89)	0.91 (0.87-0.94)	0.87 (0.85-0.90)
Model 3[g]	84	242 (40.3)	29 (4.8)	42 (7)	287 (47.8)	0.86 (0.81-0.90)	0.91 (0.88-0.94)	0.88 (0.86-0.97)
Random forest								
Model 1	34	241 (40.2)	28 (4.7)	43 (7.2)	288 (48)	0.85 (0.80-0.89)	0.91 (0.88-0.94)	0.88 (0.86-0.91)
Model 2	53	242 (40.3)	39 (6.5)	42 (4)	277 (46.1)	0.85 (0.81-0.89)	0.88 (0.85-0.92)	0.86 (0.84-0.89)
Model 3	84	243 (40.5)	26 (4.3)	41 (6.8)	290 (48.3)	0.86 (0.81-0.90)	0.92 (0.88-0.95)	0.88 (0.86-0.91)

[a]ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification.

[b]CC: chief complaint.

[c]CC refers to suicide-related chief complaints.

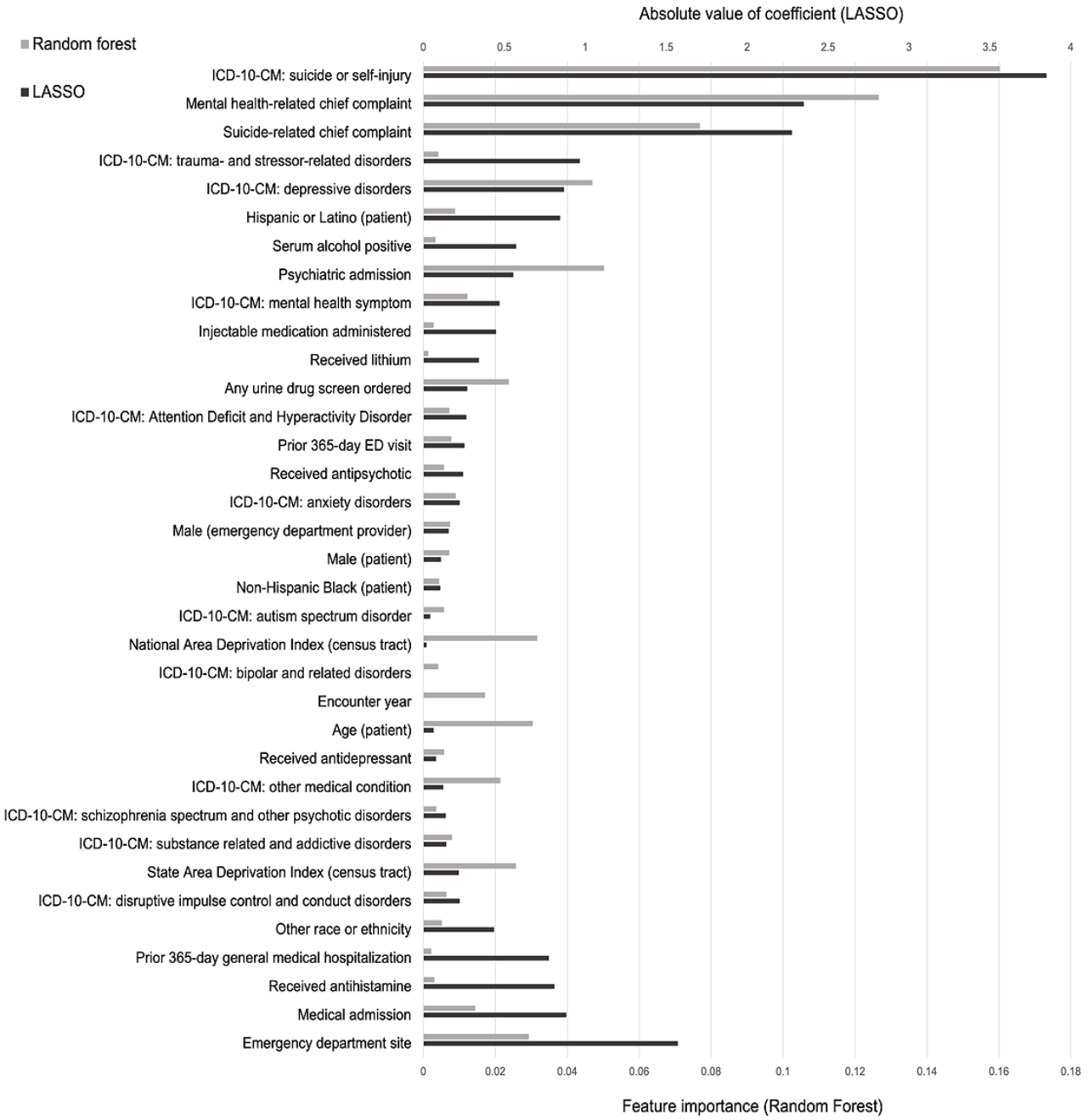
[d]LASSO: least absolute shrinkage and selection operator–penalized logistic regression.

[e]Model 1 considered all mental health–related ICD-10-CM codes organized by Child and Adolescent Mental Health Disorders Classification System categories as well as suicide-related CCs and mental health–related CCs.

[f]Model 2 considered suicide-related ICD-10-CM codes and all data elements (eg, child sociodemographics, emergency department disposition, involuntary hold status, medications, and laboratory tests) except mental health–related ICD-10-CM codes.

[g]Model 3 considered all structured data elements.

Figure 2. Feature importances for the classification of children's emergency department (ED) visits for self-injurious thoughts and behaviors. The diagram depicts features (y-axis) and the absolute value of the feature importance for least absolute shrinkage and selection operator (LASSO)–penalized logistic regression (top x-axis, dark gray) and random forest (bottom x-axis, light gray). Features with nonzero feature importance are displayed and ranked in descending order such that the topmost features are those with high positive predictive performance, and the bottommost features are those with high negative predictive importance, whereas features in the middle are of the lowest importance. ADHD: attention-deficit/hyperactivity disorder; ADI: Area Deprivation Index; CC: chief complaint; ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification.



Discussion

Principal Findings

Overall, our findings suggest that suicide-related ICD-10-CM codes and chief complaints substantially underdetect suicide-related emergency department visits and that the capacity to detect varies by sex and age group. When stratified by child demographics, suicide-related codes and chief complaints miss more male children and younger children than female children and adolescents. By contrast, machine learning–based models applied to codified health data were more sensitive in detecting suicide-related emergencies than suicide-related codes and chief complaints. When comparing machine learning–based models across health data sets with varying numbers of potential features, we found minimal differences in detection performances among models trained on all features versus those trained on mental health–related codes and chief complaints alone versus those trained on suicide-related codes and non–ICD-10-CM code–based features (eg, medications and laboratory testing). Thus, the results suggest that machine learning–based models may strengthen the sensitivity of detection of childhood-onset SITB, even when considering a focused set of potential indicators.

In this sample, nearly one-third (82/284, 28.9%) of the children presenting for suicide-related emergency care were missed by suicide-related ICD-10-CM codes, and more than half (153/284, 53.9%) of the children were missed by suicide-related chief complaints alone. Although accurate and timely detection of suicide-related emergency visits among children aligns with suicide prevention efforts by supporting tracking and rapid response to epidemiologic shifts at a population scale [38], the results of this study suggest that suicide-related codes and chief complaints alone are likely insufficient in detecting cases and potentially introduce bias regarding which children are correctly detected. The CDC National Syndromic Surveillance Program has prioritized surveillance to provide timely trend information and support public health response [26]. Using multistate public health agency reports that vary in mandates to report emergency department use for suicidal behavior, the CDC Emergency Department Surveillance of Nonfatal Suicide-Related Outcomes collects near–real-time data on nonfatal suicide-related outcomes [25]. This surveillance enabled the discovery of the rise in suicide-related emergency department visits among female adolescents aged 12 to 17 years by 50.6% during the COVID-19 pandemic [5] and provides weekly reports surveilling suicidal ideation and behavior in the state of Washington via the Rapid Health Information Network [39]. Although the surveillance of SITB is a key tool in suicide prevention, the findings of this study challenge the highly prevalent use of diagnostic codes and chief complaints as a preliminary screening tool to search for potential cases of childhood-onset SITB in clinical data sets [40-42].

This study's findings add to previously described concerns regarding the validity of suicide risk prediction models relying solely on ICD-10-CM codes to screen for the outcome of interest and discover potential antecedents [43]. The significantly poorer sensitivity of suicide-related codes and chief complaints in detecting SITB among male children and preteens and the trend (without statistical significance) toward poorer sensitivity among Black and Hispanic or Latinx children (sensitivity 0.74 vs 0.78-0.87) also raise concern that children misclassified by traditional indicators are not missed at random. The variable detection of SITB by child sociodemographics may result in biased estimates of child mental health service use and accentuate disparities; for example, bias may be introduced by unintentional omission of these children from suicide risk prediction algorithms relying on suicide-related codes and chief complaints to screen for cases. This finding builds on concern that clinical suicide risk prediction models reflect inequities in health care based on race and ethnicity and other aspects of patient identity [29].

More severe behaviors (preparatory acts and suicide attempt) were most accurately detected by requiring both suicide-related codes and chief complaints to be affirmed, whereas suicidal ideation was most accurately detected if only 1 of these (code or complaint) was required to consider the case affirmed. This is perhaps because the receipt of 2 suicide-related codified data elements may be a proxy for severity, with children with more severe behaviors receiving both data elements. The accuracy of the detection of nonsuicidal self-injurious behavior was poor compared with other SITB types, which suggests that separate phenotype definitions for types of SITB (eg, separate definitions for suicidal ideation vs preparatory acts vs suicide attempt) may produce more accurate classification than combining all SITB into a single category.

The optimal choice of detection approach may also depend on the specific use case; for instance, the results of this study suggest that suicide-related codes and chief complaints are sufficient when high specificity is important, such as flagging previous suicide-related emergencies in a patient chart. The finding that suicide-related codes and chief complaints have good specificity parallels a recent systematic assessment of self-harm coding under the ICD-10-CM in adults, which suggested that 90% of the events coded as self-harm had documentation of self-harm intent in the clinical notes [44]. In the case of a chart flag, the reduction in specificity could render a machine learning–based approach not only inconvenient but also potentially detrimental if false positives are increased. By contrast, a machine learning–based approach is more effective when maximizing sensitivity is essential, and some reduction to specificity is allowable, such as when screening data sets for potential cases. As each model generates a continuous probability of class assignment, the probability threshold may be changed depending on the use case. In uses where a high sensitivity is important to detecting all cases (eg, to not miss preteens presenting for suicide-related visits), the capacity to vary the probability threshold of classification may allow more flexibility and improved detection. These findings fit within other recent proof-of-concept applications of machine learning to classify adolescent suicidal behavior using health records, such as detection within a sample of 73 hospitalized adolescents in 1 community health system in the United States [42], a stepwise rule-based natural language processing approach evaluated on a cohort of 500 adolescents with autism spectrum disorder [8], and detection within a sample of 200 adolescents aged 11 to 17 years in contact with Child and Adolescent Mental Health Services in the United Kingdom [45].

In addition, the findings suggest that although machine learning–based approaches to detection are potentially advantageous in improving sensitivity, it may not be necessary to have access to a highly comprehensive set of data elements to meaningfully improve the sensitivity of detection. Smaller sets of mental health-related data elements, both ICD-10-CM–code based and non–ICD-10-CM–code based, performed similarly to more comprehensive data elements in the detection task. This finding aligns with work involving the phenotyping of suicidal thoughts and behaviors using discharge summaries from intensive care unit admissions in the Medical Information Mart for Intensive Care III (MIMIC-III) database and demonstrating promise of using elastic net penalized regression to detect SITB with as few as 11 features [46].

This study has several limitations. The sample is limited to a single health system in urban Los Angeles and may not generalize to less-resourced settings. The sample was also restricted to oversample case positives and individuals without missingness. Despite adjustment for selective stratification, the study sample size remained insufficient to develop and test separate machine learning–based models by sociodemographic characteristic to explore potential bias with a machine

learning–based approach. The triage screening question, "Does this patient have a primary psychiatric complaint or suspicion of psychiatric illness?" was used to determine the study sample, but triage screening questions related to suicide (eg, Columbia Suicide Severity Rating Scale items) were subsequently not selected for inclusion from the classification models because questions were asked using flow sheet cascades with a high degree of nonrandom missingness of individual items. Although commonly regarded as a gold standard for classification, the manual chart review is a silver standard for truth because it ultimately depends on information documented in clinical notes that contain the biases and idiosyncrasies of the clinical documenter and may imperfectly reflect the reality of the clinical scenario. Chief complaints were documented through an electronic health record speed button and thus may not generalize to less-structured text descriptions of the presenting problem.

Conclusions

Taken together, this study adds to existing efforts made toward developing clinical phenotypes of pediatric health conditions. Going forward, future research is needed to refine the detection of SITB across different health systems and populations, elucidate the potential advantage of including point-of-care universal suicide screening tools into phenotype detection algorithms, and determine whether including indicators of suicide-related behavior from clinical text improves detection. To achieve better integration between clinical research informatics and child mental health care, further work is needed to test the implementation of detection approaches at the point of care and assess the potential benefits of the precise identification of targets for suicide prevention interventions in children.

Acknowledgments

The authors are grateful to the University of California, Los Angeles, Clinical and Translational Science Institute Biomedical Informatics Program, especially Amanda Do, Masters of Public Health, and Theona Tacorda, Master of Science, for delivering the data used in this study. The authors would also like to acknowledge Jonathan Heldt, Doctor of Medicine, for guidance on identification of structured data elements from the electronic health record; Kristen Choi, Doctor of Philosophy, for her contribution to coding discordant cases; and Chrislie Ponce, Bachelor of Arts, Bachelor of Science, and Liliana Perez, Bachelor of Science, for assistance with chart abstraction. Funding was provided by the American Foundation for Suicide Prevention, the Harvey T and Maude C Sorensen Foundation, the Brain and Behavior Research Foundation, the Thrasher Research Fund, and the National Institute of Mental Health (K23-MH130745-01). The research was also supported by a grant from the National Center for Advancing Translational Sciences of the National Institutes of Health (UL1TR001881).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of sample-eligible children with missingness and those without missingness of diagnostic code, chief complaint, and triage screening.
[DOCX File , 18 KB - mental_v10i1e47084_app1.docx]

Multimedia Appendix 2
Study variables (structured data elements).
[DOCX File , 18 KB - mental_v10i1e47084_app2.docx]

Multimedia Appendix 3
Additional sample characteristics.
[DOCX File , 18 KB - mental_v10i1e47084_app3.docx]

Multimedia Appendix 4
Laboratory testing and medications.
[DOCX File , 33 KB - mental_v10i1e47084_app4.docx]

Multimedia Appendix 5
Sampling probability–adjusted performance of International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), code and suicide-related chief complaint in detecting cases of self-injurious thoughts and behaviors compared with that of manual chart abstraction: total sample and stratified by natal sex, age group, race, and ethnicity.
[DOCX File , 31 KB - mental_v10i1e47084_app5.docx]

Multimedia Appendix 6
Sampling probability–adjusted performance of International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), code and suicide-related chief complaint in detecting cases of self-injurious thoughts and behaviors compared with that of manual chart abstraction: stratified by Columbia Classification Algorithm of Suicide Assessment categorization.
[DOCX File , 20 KB - mental_v10i1e47084_app6.docx]

Multimedia Appendix 7
Comparison of classifier performance using the McNemar chi-square test.
[DOCX File , 13 KB - mental_v10i1e47084_app7.docx]

Multimedia Appendix 8
Fit metrics and per-fold feature importances of least absolute shrinkage and selection operator (LASSO) and random forest classifiers.
[DOCX File , 98 KB - mental_v10i1e47084_app8.docx]

References

1. Centers for Disease Control and Prevention. 2014. URL: https://www.cdc.gov/injury/wisqars/index.html [accessed 2023-04-01]
2. Stone D, Holland K, Bartholow B, Crosby A, Davis S, Wilkins N. Preventing suicide: a technical package of policies, programs, and practice. Centers for Disease Control and Prevention. 2017. URL: https://stacks.cdc.gov/view/cdc/44275 [accessed 2022-12-19]
3. Larkin GL, Smith RP, Beautrais AL. Trends in US emergency department visits for suicide attempts, 1992-2001. Crisis 2008;29(2):73-80. [doi: 10.1027/0227-5910.29.2.73] [Medline: 18664232]
4. Ballard ED, Cwik M, Van Eck K, Goldstein M, Alfes C, Wilson ME, et al. Identification of at-risk youth by suicide screening in a pediatric emergency department. Prev Sci 2017 Feb;18(2):174-182 [FREE Full text] [doi: 10.1007/s11121-016-0717-5] [Medline: 27678381]
5. Yard E, Radhakrishnan L, Ballesteros MF, Sheppard M, Gates A, Stein Z, et al. Emergency department visits for suspected suicide attempts among persons aged 12-25 years before and during the COVID-19 pandemic - United States, January 2019-May 2021. MMWR Morb Mortal Wkly Rep 2021 Jun 18;70(24):888-894 [FREE Full text] [doi: 10.15585/mmwr.mm7024e1] [Medline: 34138833]
6. Lo CB, Bridge JA, Shi J, Ludwig L, Stanley RM. Children's mental health emergency department visits: 2007-2016. Pediatrics 2020 Jun;145(6):e20191536. [doi: 10.1542/peds.2019-1536] [Medline: 32393605]
7. Madigan S, Korczak DJ, Vaillancourt T, Racine N, Hopkins WG, Pador P, et al. Comparison of paediatric emergency department visits for attempted suicide, self-harm, and suicidal ideation before and during the COVID-19 pandemic: a systematic review and meta-analysis. Lancet Psychiatry 2023 May;10(5):342-351 [FREE Full text] [doi: 10.1016/S2215-0366(23)00036-6] [Medline: 36907199]
8. Downs J, Velupillai S, George G, Holden R, Kikoler M, Dean H, et al. Detection of suicidality in adolescents with autism spectrum disorders: developing a natural language processing approach for use in electronic health records. AMIA Annu Symp Proc 2018 Apr 16;2017:641-649 [FREE Full text] [Medline: 29854129]
9. Zhong QY, Mittal LP, Nathan MD, Brown KM, Knudson González D, Cai T, et al. Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: towards a solution to the complex classification problem. Eur J Epidemiol 2019 Feb;34(2):153-162 [FREE Full text] [doi: 10.1007/s10654-018-0470-0] [Medline: 30535584]
10. Bertoia ML, Spalding WM, Bulik CM, Yee KS, Zhou L, Cheng H, et al. Identification of patients with suicidal ideation or attempt in electronic health record data. Pharmacoepidemiol Drug Saf 2019;28(S2):162 [FREE Full text]
11. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc 2014 Mar;21(2):221-230 [FREE Full text] [doi: 10.1136/amiajnl-2013-001935] [Medline: 24201027]
12. Enticott J, Johnson A, Teede H. Learning health systems using data to drive healthcare improvement and impact: a systematic review. BMC Health Serv Res 2021 Mar 05;21(1):200 [FREE Full text] [doi: 10.1186/s12913-021-06215-8] [Medline: 33663508]
13. Zima BT, Hurlburt MS, Knapp P, Ladd H, Tang L, Duan N, et al. Quality of publicly-funded outpatient specialty mental health care for common childhood psychiatric disorders in California. J Am Acad Child Adolesc Psychiatry 2005 Feb;44(2):130-144. [doi: 10.1097/00004583-200502000-00005] [Medline: 15689726]
14. Bejan CA, Angiolillo J, Conway D, Nash R, Shirey-Rice JK, Lipworth L, et al. Mining 100 million notes to find homelessness and adverse childhood experiences: 2 case studies of rare and severe social determinants of health in electronic health records. J Am Med Inform Assoc 2018 Jan 01;25(1):61-71. [doi: 10.1093/jamia/ocx059] [Medline: 29016793]
15. Connolly J. ADHD phenotype algorithm. National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd004968.1 [accessed 2022-11-30]

16. Depression algorithm. Phenotype Knowledgebase (PheKB). 2018. URL: https://phekb.org/phenotype/1095 [accessed 2023-04-26]
17. Connolly J. Anxiety algorithm. Phenotype Knowledgebase (PheKB). 2018. URL: https://phekb.org/phenotype/1105 [accessed 2023-04-26]
18. Walters Jr CE, Nitin R, Margulis K, Boorom O, Gustavson DE, Bush CT, et al. Automated phenotyping tool for identifying developmental language disorder cases in health systems data (APT-DLD): a new research algorithm for deployment in large-scale electronic health record systems. J Speech Lang Hear Res 2020 Sep 15;63(9):3019-3035 [FREE Full text] [doi: 10.1044/2020_JSLHR-19-00397] [Medline: 32791019]
19. Lingren T, Chen P, Bochenek J, Doshi-Velez F, Manning-Courtney P, Bickel J, et al. Electronic health record based algorithm to identify patients with autism spectrum disorder. PLoS One 2016 Jul 29;11(7):e0159621 [FREE Full text] [doi: 10.1371/journal.pone.0159621] [Medline: 27472449]
20. Khare R, Kappelman MD, Samson C, Pyrzanowski J, Darwar RA, Forrest CB, And the PEDSnet Computable Phenotype Working Group. Development and evaluation of an EHR-based computable phenotype for identification of pediatric Crohn's disease patients in a national pediatric learning health system. Learn Health Syst 2020 Aug 28;4(4):e10243 [FREE Full text] [doi: 10.1002/lrh2.10243] [Medline: 33083542]
21. Qin Y, Kernan KF, Fan Z, Park HJ, Kim S, Canna SW, et al. Machine learning derivation of four computable 24-h pediatric sepsis phenotypes to facilitate enrollment in early personalized anti-inflammatory clinical trials. Crit Care 2022 May 07;26(1):128 [FREE Full text] [doi: 10.1186/s13054-022-03977-3] [Medline: 35526000]
22. Phillips CA, Razzaghi H, Aglio T, McNeil MJ, Salvesen-Quinn M, Sopfe J, et al. Development and evaluation of a computable phenotype to identify pediatric patients with leukemia and lymphoma treated with chemotherapy using electronic health record data. Pediatr Blood Cancer 2019 Sep;66(9):e27876 [FREE Full text] [doi: 10.1002/pbc.27876] [Medline: 31207054]
23. Geva A, Gronsbell JL, Cai T, Cai T, Murphy SN, Lyons JC, Pediatric Pulmonary Hypertension Network and National Heart, Lung, and Blood Institute Pediatric Pulmonary Vascular Disease Outcomes Bioinformatics Clinical Coordinating Center Investigators. A computable phenotype improves cohort ascertainment in a pediatric pulmonary hypertension registry. J Pediatr 2017 Sep;188:224-31.e5 [FREE Full text] [doi: 10.1016/j.jpeds.2017.05.037] [Medline: 28625502]
24. Hedegaard H, Schoenbaum M, Claassen C, Crosby A, Holland K, Proescholdbell S. Issues in developing a surveillance case definition for nonfatal suicide attempt and intentional self-harm using international classification of diseases, tenth revision, clinical modification (ICD-10-CM) coded data. Natl Health Stat Report 2018 Feb(108):1-19 [FREE Full text] [Medline: 29616901]
25. Centers FDC. Emergency department surveillance of nonfatal suicide-related outcomes. Centers for Disease Control and Prevention. 2022 Jun 09. URL: https://www.cdc.gov/suicide/programs/ed-snsro/index.html [accessed 2022-12-19]
26. Zwald ML, Holland KM, Annor F, Kite-Powell AK, Sumner SA, Bowen D, et al. Monitoring suicide-related events using national syndromic surveillance program data. Online J Public Health Inform 2019 May 30;11(1):e440 [FREE Full text] [doi: 10.5210/ojphi.v11i1.9927]
27. Nitin R, Shaw DM, Rocha DB, Walters Jr CE, Chabris CF, Camarata SM, et al. Association of developmental language disorder with comorbid developmental conditions using algorithmic phenotyping. JAMA Netw Open 2022 Dec 01;5(12):e2248060 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.48060] [Medline: 36580336]
28. van Velzen LS, Toenders YJ, Avila-Parcet A, Dinga R, Rabinowitz JA, Campos AI, et al. Classification of suicidal thoughts and behaviour in children: results from penalised logistic regression analyses in the adolescent brain cognitive development study. Br J Psychiatry 2022 Apr;220(4):210-218. [doi: 10.1192/bjp.2022.7] [Medline: 35135639]
29. Coley RY, Johnson E, Simon GE, Cruz M, Shortreed SM. Racial/ethnic disparities in the performance of prediction models for death by suicide after mental health visits. JAMA Psychiatry 2021 Jul 01;78(7):726-734 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.0493] [Medline: 33909019]
30. Katki HA, Li Y, Edelstein DW, Castle PE. Estimating the agreement and diagnostic accuracy of two diagnostic tests when one test is conducted on only a subsample of specimens. Stat Med 2012 Feb 28;31(5):436-448 [FREE Full text] [doi: 10.1002/sim.4422] [Medline: 22139832]
31. Zima BT, Gay JC, Rodean J, Doupnik SK, Rockhill C, Davidson A, et al. Classification system for international classification of diseases, ninth revision, clinical modification and tenth revision pediatric mental health disorders. JAMA Pediatr 2020 Jun 01;174(6):620-622 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.0037] [Medline: 32202603]
32. Singh GK. Area deprivation and widening inequalities in US mortality, 1969-1998. Am J Public Health 2003 Jul;93(7):1137-1143. [doi: 10.2105/ajph.93.7.1137] [Medline: 12835199]
33. Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia classification algorithm of suicide assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. Am J Psychiatry 2007 Jul;164(7):1035-1043 [FREE Full text] [doi: 10.1176/ajp.2007.164.7.1035] [Medline: 17606655]
34. Franzen M, Keller F, Brown RC, Plener PL. Emergency presentations to child and adolescent psychiatry: nonsuicidal self-injury and suicidality. Front Psychiatry 2019 Jan 17;10:979 [FREE Full text] [doi: 10.3389/fpsyt.2019.00979] [Medline: 32010000]

35. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J R Stat Soc Series B Stat Methodol 2011 Jun;73(3):273-282 [FREE Full text] [doi: 10.1111/j.1467-9868.2011.00771.x]
36. Breiman L. Random forests. Mach Learn 2001;45(1):5-32 [FREE Full text] [doi: 10.1186/1478-7954-9-29] [Medline: 21816105]
37. Yang S, Varghese P, Stephenson E, Tu K, Gronsbell J. Machine learning approaches for electronic health records phenotyping: a methodical review. J Am Med Inform Assoc 2023 Jan 18;30(2):367-381. [doi: 10.1093/jamia/ocac216] [Medline: 36413056]
38. Zwald ML, Holland KM, Annor FB, Kite-Powell A, Sumner SA, Bowen DA, et al. Syndromic surveillance of suicidal ideation and self-directed violence - United States, January 2017-December 2018. MMWR Morb Mortal Wkly Rep 2020 Jan 31;69(4):103-108 [FREE Full text] [doi: 10.15585/mmwr.mm6904a3] [Medline: 31999688]
39. Rapid health information network (RHINO). Washington State Department of Health. URL: https://doh.wa.gov/public-health-healthcare-providers/healthcare-professions-and-facilities/public-health-meaningful-use/rhino [accessed 2022-11-30]
40. Tsui FR, Shi L, Ruiz V, Ryan ND, Biernesser C, Iyengar S, et al. Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts. JAMIA Open 2021 Mar 17;4(1):ooab011 [FREE Full text] [doi: 10.1093/jamiaopen/ooab011] [Medline: 33758800]
41. Walsh CG, Ribeiro JD, Franklin JC. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. J Child Psychol Psychiatry 2018 Dec;59(12):1261-1270. [doi: 10.1111/jcpp.12916] [Medline: 29709069]
42. Carson NJ, Mullin B, Sanchez MJ, Lu F, Yang K, Menezes M, et al. Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records. PLoS One 2019 Feb 19;14(2):e0211116 [FREE Full text] [doi: 10.1371/journal.pone.0211116] [Medline: 30779800]
43. Belsher BE, Smolenski DJ, Pruitt LD, Bush NE, Beech EH, Workman DE, et al. Prediction models for suicide attempts and deaths: a systematic review and simulation. JAMA Psychiatry 2019 Jun 01;76(6):642-651. [doi: 10.1001/jamapsychiatry.2019.0174] [Medline: 30865249]
44. Simon GE, Shortreed SM, Boggs JM, Clarke GN, Rossom RC, Richards JE, et al. Accuracy of ICD-10-CM encounter diagnoses from health records for identifying self-harm events. J Am Med Inform Assoc 2022 Nov 14;29(12):2023-2031. [doi: 10.1093/jamia/ocac144] [Medline: 36018725]
45. Velupillai S, Epstein S, Bittar A, Stephenson T, Dutta R, Downs J. Identifying suicidal adolescents from mental health records using natural language processing. Stud Health Technol Inform 2019 Aug 21;264:413-417. [doi: 10.3233/SHTI190254] [Medline: 31437956]
46. Buckland RS, Hogan JW, Chen ES. Selection of clinical text features for classifying suicide attempts. AMIA Annu Symp Proc 2021 Jan 25;2020:273-282 [FREE Full text] [Medline: 33936399]

Abbreviations

ADI: Area Deprivation Index
CAMHD-CS: Child and Adolescent Mental Health Disorders Classification System
C-CASA: Columbia Classification Algorithm of Suicide Assessment
CDC: Centers for Disease Control and Prevention
FIPS: Federal Information Processing System
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
LASSO: least absolute shrinkage and selection operator
MIMIC-III: Medical Information Mart for Intensive Care III
SITB: self-injurious thoughts and behaviors
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by J Torous; submitted 09.03.23; peer-reviewed by S Shortreed, E Chan, D Lekkas; comments to author 18.04.23; revised version received 11.05.23; accepted 29.05.23; published 21.07.23.

Please cite as:
Edgcomb JB, Tseng CH, Pan M, Klomhaus A, Zima BT
Assessing Detection of Children With Suicide-Related Emergencies: Evaluation and Development of Computable Phenotyping Approaches
JMIR Ment Health 2023;10:e47084
URL: https://mental.jmir.org/2023/1/e47084
doi:10.2196/47084
PMID:37477974

Original Paper

Using HIPAA (Health Insurance Portability and Accountability Act)–Compliant Transcription Services for Virtual Psychiatric Interviews: Pilot Comparison Study

Salman Seyedi[1*], PhD; Emily Griner[2*], MPH; Lisette Corbin[3], MD; Zifan Jiang[1,4], PhD; Kailey Roberts[5], MPH; Luca Iacobelli[2]; Aaron Milloy[6], MPH; Mina Boazak[7], MD; Ali Bahrami Rad[1], PhD; Ahmed Abbasi[8], PhD; Robert O Cotes[2], MD; Gari D Clifford[1,4], PhD

[1]Department of Biomedical Informatics, Emory University, Atlanta, GA, United States
[2]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, United States
[3]Department of Psychiatry, Duke University Health, Durham, NC, United States
[4]Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, United States
[5]Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA, United States
[6]Infection Prevention Department, Emory Healthcare, Atlanta, GA, United States
[7]Animo Sano Psychiatry, Durham, NC, United States
[8]Department of Information Technology, Analytics, and Operations, University of Notre Dame, Notre Dame, IN, United States
[*]these authors contributed equally

Corresponding Author:
Salman Seyedi, PhD
Department of Biomedical Informatics
Emory University
100 Woodruff Circle
Atlanta, GA, 30322
United States
Phone: 1 404 727 4562
Email: sseyedi@emory.edu

Abstract

Background: Automatic speech recognition (ASR) technology is increasingly being used for transcription in clinical contexts. Although there are numerous transcription services using ASR, few studies have compared the word error rate (WER) between different transcription services among different diagnostic groups in a mental health setting. There has also been little research into the types of words ASR transcriptions mistakenly generate or omit.

Objective: This study compared the WER of 3 ASR transcription services (Amazon Transcribe [Amazon.com, Inc], Zoom-Otter AI [Zoom Video Communications, Inc], and Whisper [OpenAI Inc]) in interviews across 2 different clinical categories (controls and participants experiencing a variety of mental health conditions). These ASR transcription services were also compared with a commercial human transcription service, Rev (Rev.Com, Inc). Words that were either included or excluded by the error in the transcripts were systematically analyzed by their Linguistic Inquiry and Word Count categories.

Methods: Participants completed a 1-time research psychiatric interview, which was recorded on a secure server. Transcriptions created by the research team were used as the gold standard from which WER was calculated. The interviewees were categorized into either the control group (n=18) or the mental health condition group (n=47) using the Mini-International Neuropsychiatric Interview. The total sample included 65 participants. Brunner-Munzel tests were used for comparing independent sets, such as the diagnostic groupings, and Wilcoxon signed rank tests were used for correlated samples when comparing the total sample between different transcription services.

Results: There were significant differences between each ASR transcription service's WER (P<.001). Amazon Transcribe's output exhibited significantly lower WERs compared with the Zoom-Otter AI's and Whisper's ASR. ASR performances did not significantly differ across the 2 different clinical categories within each service (P>.05). A comparison between the human transcription service output from Rev and the best-performing ASR (Amazon Transcribe) demonstrated a significant difference (P<.001), with Rev having a slightly lower median WER (7.6%, IQR 5.4%-11.35 vs 8.9%, IQR 6.9%-11.6%). Heat maps and

spider plots were used to visualize the most common errors in Linguistic Inquiry and Word Count categories, which were found to be within 3 overarching categories: Conversation, Cognition, and Function.

Conclusions: Overall, consistent with previous literature, our results suggest that the WER between manual and automated transcription services may be narrowing as ASR services advance. These advances, coupled with decreased cost and time in receiving transcriptions, may make ASR transcriptions a more viable option within health care settings. However, more research is required to determine if errors in specific types of words impact the analysis and usability of these transcriptions, particularly for specific applications and in a variety of populations in terms of clinical diagnosis, literacy level, accent, and cultural origin.

(JMIR Ment Health 2023;10:e48517) doi:10.2196/48517

KEYWORDS

ASR; automatic speech recognition; Health Insurance Portability and Accountability Act; HIPAA; Linguistic Inquiry and Word Count; LIWC; mental health; psychiatric interview; speech to text; WER; word error rate

Introduction

In 2020, 52.9 million (21%) of US adults experienced a mental illness, and of those, 41.4 million (17%) received mental health services either through inpatient treatment, outpatient treatment, or prescription medication [1]. Workforce shortages, specifically within mental health, have been well-documented and are projected to be a persistent concern in the future [2-4]. There is expected to be an insufficient supply of mental health practitioners to meet the need for psychiatric care by 2030, with the number of psychiatrists decreasing by 20% [2]. This shortage is even more pronounced in certain geographic areas due to an uneven distribution of psychiatrists and other mental health practitioners, further limiting access to care [2,4]. To assist with these shortages, many have proposed increased use of telehealth and other technology-assisted solutions to increase efficiency. One promising approach is to use automatic speech recognition (ASR) systems to convert speech into readable text or transcriptions.

The evolution of ASR systems over the years has been marked by a significant reduction in the word error rate (WER), a key metric in evaluating these systems. This is consistent with the decline observed in the WER across the literature. From approximately 30% in the early 2000s [5] to 10% to 15% in the 2010s [6], and subsequently dropping below 10% in recent years [7], the WER trend reflects this evolution. In recent years, the WER of ASR systems applied to the data set (Librispeech Other) demonstrates this trend: Panayotov et al [8] recorded 13.97%, Zeghidour et al [9] achieved 11.24%, Irie et al [10] attained 10.3%, and Whisper (OpenAI Inc) [11] impressively reached a WER of 5.2% [11]. Further insights from diverse data sets and models can be found elsewhere ([11] or Park et al [12]).

ASR has been explored in various clinical applications and continues to grow in popularity. One of the most notable uses of ASR in a clinical setting is to assist the practitioner with clinical documentation in the electronic health record (EHR) [13,14]. EHRs have been cited as a contributing factor to physician burnout due to the significant increase in time spent completing documentation, which has decreased time spent with patients [15,16]. Recommendations to alleviate these challenges have included improving EHRs through ASR technology [17]. Automated medical scribe services may decrease administrative burden and lessen physician burnout related to documentation [13]. Another application of ASR is to assist with clinical education. For example, automated transcriptions can be used in psychotherapy supervision contexts by reducing the time spent generating human transcriptions, providing more timely feedback, and quantifying other relevant information, such as the amount of time spent talking by both the therapist and the client [18]. Research has also begun to use ASR in predictive models to assess, diagnose, and track mental illness [19-21].

One of the most commonly used natural language tools in text analysis is the Linguistic Inquiry and Word Count (LIWC) [22]. The most up-to-date version, LIWC-22, has an internal dictionary of over 12,000 words categorized into various groups intended to assess different psychosocial constructs [22]. Numerous categories related to first-person pronoun use and negatively valenced emotion and tone words have been shown to be associated with depression symptom severity [23-25]. Similarly, people experiencing psychosis use more personal pronouns, negative emotion words, biological process words, and fewer words per sentence compared with controls [26,27].

Although there is a growing interest in ASR applications in clinical practice, few studies have compared the accuracy of commercially available regulatory-compliant, for example, the Health Insurance Portability and Accountability Act (HIPAA) of 1996 [28], transcription services in mental health settings [29]. HIPAA-compliant services provide a regulatory standard for protecting an individual's identity. Previous research that looked at ASR services' WERs used Google Cloud speech-to-text after going through an 8-month process to obtain a waiver from their university and having an existing business associate agreement with Google [29]. This process will not be reasonable for many settings; therefore, this study seeks to compare HIPAA-compliant services such as Amazon Transcribe (Amazon.com, Inc) and Zoom-Otter AI (Zoom Video Communications, Inc), along with the latest state-of-the-art open-source software, Whisper, implemented on a local HIPAA-compliant server on which the study data was located. Furthermore, this study also compares the WERs of these transcription services by analyzing groups with different clinical diagnoses. Although ASR services are constantly improving, there is a need to continue to compare these services for a variety of populations. This study compares the WER of multiple commercially available ASR services against human transcriptions using clinical interviews from controls and those experiencing various mental illnesses. The eventual aim of

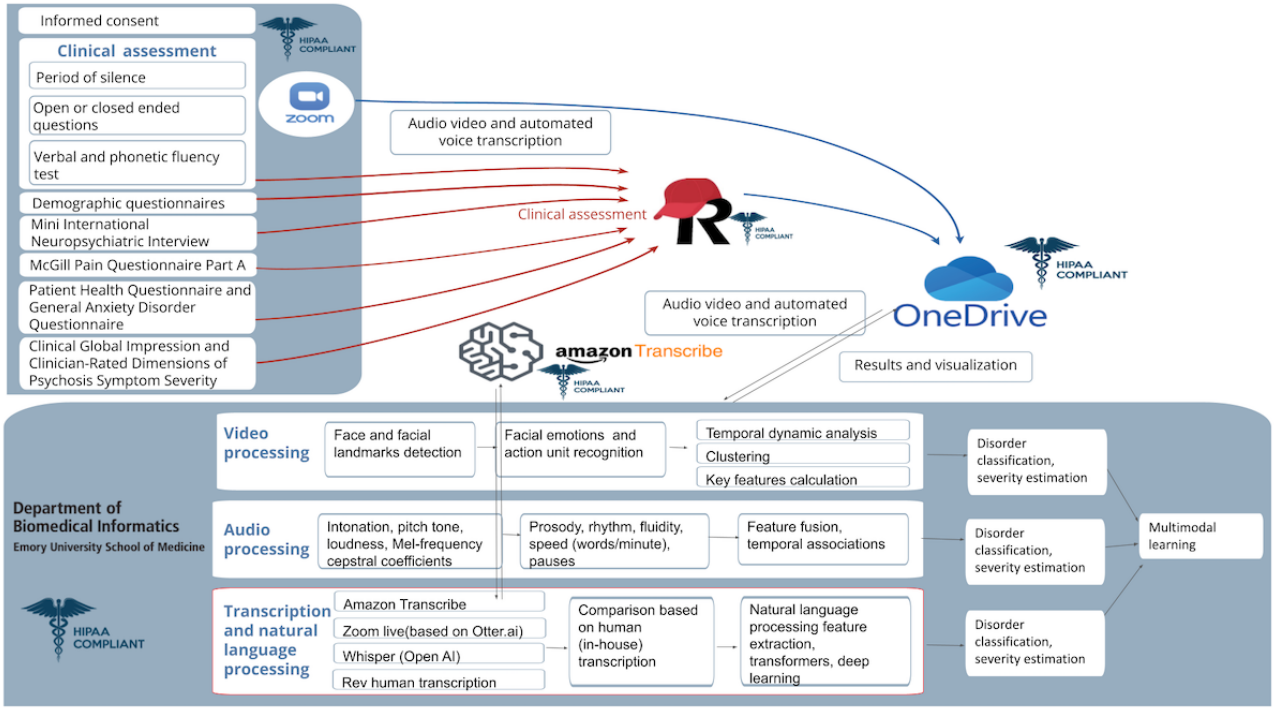
assessing ASR services is to develop a scalable, timely, and cost-effective system for automatically analyzing the semantic content of telemedicine visits to assist in diagnosis and treatment recommendations.

Methods

Overview

The overall project protocol can be found in Cotes et al [30]. In Figure 1, we have detailed only the parts of the project pertinent to evaluating the automatic and human transcriptions used in the study, highlighted in red.

Figure 1. Schematic flow of data collection, storage, and processing. The process of administering the interviews is listed on the left. The clinical categorizations then flow into REDCap (Research Electronic Data Capture; Vanderbilt University; red lines). Audio and video recordings are captured during the interview and placed in the OneDrive (Microsoft Corporation) folder. In this work, we focus on the part of the project shown in the lower red box. Otter AI transcriptions are produced afterward. Data are then synchronized to local servers, which process the same audio data to shorten the length of the files and then transcribe with OpenAI's Whisper software, Amazon Transcribe, and Rev transcription service (human). Further audio and video processing will be implemented when the transcription process has been fully validated. Adapted from Cotes et al (CC BY [Creative Commons Attribution license] open-source license).



Recruitment

Interviewees were recruited from Research Match [31], a National Institutes of Health–funded, web-based recruitment strategy designed to connect potential participants to research studies, and through Grady's Behavioral Health Outpatient Clinic using a database of interested research participants. Participants were aged between 18 and 65 years and were native English speakers. All interviewees were in the United States at the time of the interview. For the initial screening, interviewees were recruited for either a control group (no history of mental illness within the past 12 months) or a group currently experiencing depression. All diagnoses and group categorizations were verified and finalized by the overseeing psychiatrist and clinical team after the semistructured interview.

Interviews

All interviews were conducted remotely through Zoom's (Zoom Video Communications) secure, encrypted, and HIPAA-compliant platform. The interview guide and protocol were created by the study team with components that simulate a psychiatric intake interview [30]. The interview was divided into three parts: (1) a semistructured interview composed of a series of open-ended questions, a thematic apperception test (TAT) [32], a phonetic fluency test [33], and a semantic fluency test [34]; (2) a sociodemographic section; and (3) clinical assessments, which included the Mini-International Neuropsychiatric Interview (MINI) 6.0 [35], McGill Quality of Life Questionnaire [36], General Anxiety Disorder-7 [37], and Patient Health Questionnaire-9 [38]. The upper left side box in Figure 1 contains the visual representation of this flow.

Categorization

The final sample included 65 interviewees that were categorized into 1 of 2 groups: control (n=18) or mental health condition (MHC) (n=47). Over half (14/18, 78%) of the individuals in the control group did not code into any current or past major depressive disorder, and those with a past history of major depressive disorder (4/18, 22%) all confirmed that their last 2-week episode of depression did not occur within the prior previous 12 months. Individuals with a mental health condition (MHC) had either a primary diagnosis of major depressive disorder (35/47, 75%), some sort of psychosis or manic disorder (9/47, 19%), or a primary anxiety or obsessive-compulsive

disorder (3/47, 6%). These individuals also experienced comorbidities such as agoraphobia (18/47, 38%), generalized anxiety disorder (14/47, 30%), posttraumatic stress disorder (4/47, 9%), social anxiety disorder (6/47, 13%), or substance abuse or dependence disorders (5/47, 11%). All classifications were discussed and verified by the overseeing psychiatrist and research team. Interviewees who were unable to participate in the interview or who had a history of traumatic brain or neck injury were excluded. All interviewee demographics are shown in Table 1.

Table 1. Interviewee demographics for each interviewee group.

Demographics	Control group (n=18)	MHC[a] group (n=47)	All participants (n=65)	*P* values
Age (y)				.03[b]
Mean (SD)	46 (14.19)	38 (13.98)	40 (14.40)	
Median (IQR)	48 (33.25-59.00)	33 (26-49.5)	36 (26-52)	
Gender, n (%)				.10[c]
Female	10	35	45	
Male	8	9	17	
Nonbinary	0	3	3	
Race, n (%)				.19[c]
Asian	1	6	7	
Black or African American	7	7	14	
Hispanic or Latino	0	3	3	
White	9	30	39	
Mixed race	1	1	2	
Years of education				.78[b]
Mean (SD)	17 (5.20)	17 (2.52)	17 (3.45)	
Median (IQR)	16 (14.50-18.75)	17 (15-18)	17 (15-18)	
Education Level, n (%)				.27[c]
Some high school	1	1	2	
High school graduate	0	2	2	
Some college, trade, or vocational school	4	11	15	
College graduate	6	6	12	
Graduate or professional school	7	27	34	

[a]MHC: mental health condition.

[b]Mann-Whitney Test.

[c]Fischer exact test.

Automatic Transcription Process

The automatic transcription services used in this study were: Amazon Transcribe, Zoom live transcription (using OtterAI), and Whisper (an open-source ASR system by OpenAI). Amazon Transcribe and Zoom-Otter AI have HIPAA-compliant services that were used in this study. The Whisper ASR was downloaded and used on the local HIPAA-compliant servers. Zoom allows for recording separate audio tracks; therefore, the interviewees' audio files were used to produce all automatic transcripts. The interviewee-sided audio was played as input within a recorded Zoom meeting (muted with shared audio) with live transcription to produce a text file for only the interviewee-sided audio. To reduce costs, interviewee audio files were edited to remove silences longer than 1 second from the files based on an average amplitude threshold of 5 in 1 second. These shortened audio files showed lower errors and were used as input to create Amazon Transcribe and Whisper transcripts.

Human Transcription Process

Gold standard transcriptions were produced by the research team using a multiple-overread and consensus approach and were used throughout this study as the gold standard.

We followed the earlier work of Neamatullah et al [39] for the deidentification of medical data. Given that algorithms are sensitive but not specific and humans are the opposite, by combining the strengths of both and adding a human overread step, they demonstrated that this was a highly effective process that neither distorted medical data nor leaked protected health information. In this study, the automated transcriptions for Amazon Transcribe and Zoom-Otter AI were used to produce

a side-by-side comparison text document to serve as the basis for human overreads. A total of 2 clinically trained experts overread the entire transcription while listening to the audio. Where the 2 transcriptions disagreed, the computer code highlighted the section with an underscore to help draw the human's attention to the issue. The research team then edited a separate file to create a correct overread, or gold standard. All numerical quantities were transposed to their word equivalents, slang was written phonetically, and brackets were used to denote utterances such as laughter. Any discrepancies in the transcripts were resolved by the consensus of the 3 clinical transcribers.

The research team also created interviewee-sided transcriptions using a commercial human transcription service offered by Rev. These transcriptions were generated using the shortened audio files previously used to create the Amazon Transcribe and Whisper automated transcriptions. These transcriptions were not used in the process of creating the gold standard but, rather, were generated as a baseline for human performance.

WER Process

Preprocessing and Text Preparation

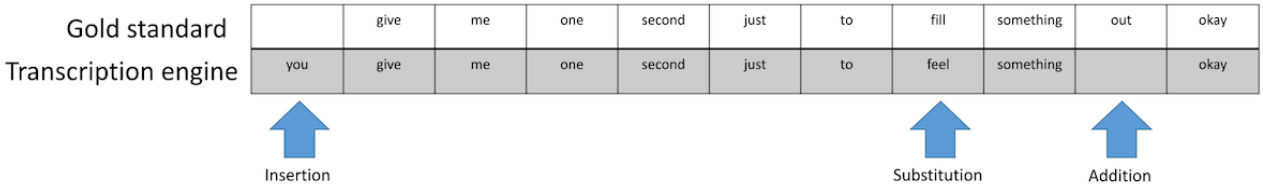
Each transcript was preprocessed by expanding contractions, removing annotations, and changing all characters to lowercase. Different notations for different transcripts were recognized, and labels and descriptions were removed (for instance, "[Laughter]"). Also, a specific dictionary was built to alleviate error counting between transcriptions based on stylistic preferences (eg, "twenty twenty two," "two thousand twenty-two," and "two thousand and twenty-two" were all replaced with "2022").

WER Calculation

The gold standard transcriptions produced by interviewers were used as the reference, and all other transcriptions' performances were compared with the gold standard using WER, a common and important metric for measuring the accuracy of transcriptions [40,41]. The WER, expressed as a percentage, is defined as:

where S is the number of substitutions, D is the number of deletions (words spoken but left out in the transcription, such as the word "out" in Figure 2), I is the number of insertions (words that are not spoken but have been inserted by the transcription, such as the word "you" in Figure 2), H is the number of hits (correct words), and N is the total number of words in the reference (gold standard). The *Jiwer* library [42] was used to calculate the WER. The library is based on the minimum-edit distance calculated using the Levenshtein distance [43].

Figure 2. Illustration of the word error rate (WER) calculation. The word "you" is inserted by the transcription engine but has not been spoken. The word "fill" is substituted by "feel" by the transcription engine. Both the words "fill" and "feel" are counted as substitutions, but "fill" would be counted as a substitution deletion (S-delete), and "feel" would be counted as a substitution insertion (S-insert) within this study. The word "out" is deleted by the transcription engine, although it has been spoken.



The substitution words that are counted in S can be divided into 2 parts. The S-deletes are the words that are being substituted ("fill" in Figure 2), and the S-inserts are substitution words ("feel" in Figure 2). We bundle the error words that are counted in insert (I) and those that are in S-insert in one category called "Error Insert." Then the other half, which are words in S-delete and words counted in D together, are called "Error Delete."

Statistical Tests

All P values were calculated for 2 sets at a time. A Wilcoxon signed rank test (a nonparametric test) was used for correlated samples, such as when comparing the WER for each interview between different transcription services. The Brunner-Munzel test was used for the independent sets, for example, when comparing the WER between genders or between groups of individuals with different clinical diagnoses. The statistics package *SciPy* (version 1.7.3; Python Library) was used for the calculations.

Ethical Considerations

The Emory University Institutional Review Board and the Grady Research Oversight Committee granted approval for this study (IRB #00105142). All participants provided informed consent to partake in the study, and all collected data were deidentified by providing a unique identifier to each participant. All participants were compensated for their time volunteering in this study and were provided a 1-time financial incentive (US $30).

Results

Overall, the interviews lasted 46 minutes on average, with the shortest interview lasting 25 minutes and the longest lasting 1 hour and 55 minutes. Before comparing the clinical groups, WER across genders was compared for each of the 4 services. We did not see any statistical difference for WER between male and female individuals for Amazon Transcribe (P=.71), Zoom-Otter AI (P=.39), Whisper (P=.79), or Rev (P=.42). The number of interviewees identifying as nonbinary was too low for any statistical analysis. The comparison of WER for race between individuals who identify as White versus those who did not identify as White did not show any statistically significant difference for Amazon Transcribe (P=.17), Zoom-Otter AI (P=.26), Whisper (P=.25), or Rev (P=.13). The

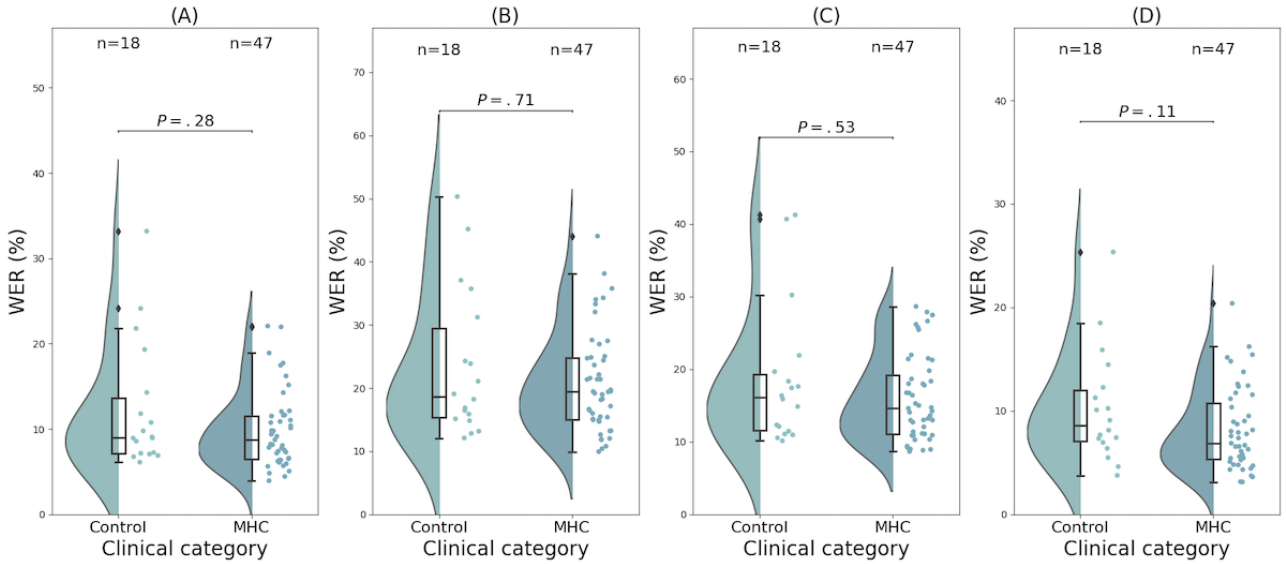
groups of individuals who did not identify as White had counts that were too low for any statistical analysis, and thus, we turned this into a binary analysis.

The median of total words spoken by each interviewee was 1280 (IQR 927.0-2041.0). The median word counts for the control and MHC groups were 1337 (IQR 802.5-1961.5) and 1232 (IQR 969.0-1964.0), respectively. Transcriptions were generated using Whisper for both medium and large models. The medium model was chosen due to its superior performance.

Comparing Clinical Categorizations by Transcription Service

The median WER values for the control group were lowest for Rev transcription at 8.6% (IQR 7.0%-12.0%), followed by Amazon Transcribe at 9% (IQR 7.2%-13.7%), Whisper at 16.1% (11.6%-19.3%), and Zoom-Otter AI at 18.6% (IQR 15.3%-29.5%). This same trend was seen for the MHC group with Rev transcription's WER value being 6.9% (IQR 5.0%-10.8%), followed by Amazon Transcribe's WER value of 8.7% (IQR 6.4%-11.6%), Whisper's WER value of 16.1% (IQR 11.1%-19.2%), and Zoom-Otter AI's WER value of 19.4% (IQR 15.0%-24.8%). Using the Brunner-Munzel statistical test comparing each group in a pair-wise manner, there was no statistically significant difference between the clinical groups' WER for any of the transcription services (all $P>.05$). The breakdown of the WER for transcripts based on the 2 clinical groupings and related P values can be seen in Figure 3. Since there were no significant differences found between the clinical groups for each transcription service, the total sample (N=65) was used to compare transcription services with one another for the remainder of all analyses.

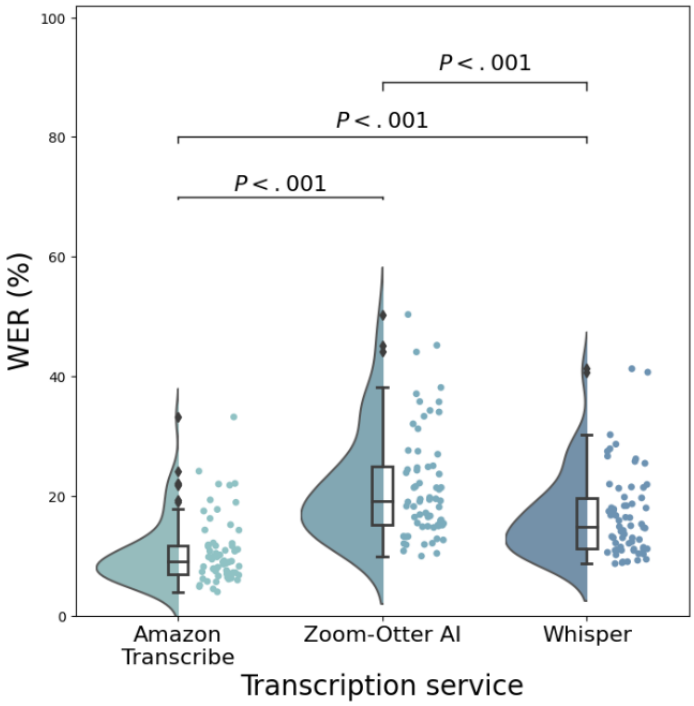
Figure 3. Distributions of the word error rates (WERs) for the control group and the mental health condition (MHC) group within (A) Amazon Transcribe, (B) Zoom-Otter AI, (C) Whisper, and (D) Rev human transcription. Distributions are estimated from actual values (dots) using a kernel density estimate. Box plots indicate the median with the 25th and 75th percentiles. Note that with ($P>.05$) for every Brunner-Munzel test applied between 2 categories, there were no statistically significant differences in WER between the control (n=18) and MHC (n=47) groups within each transcription method.



Comparing Automatic Transcription Services

The median WER of the 3 tested automatic transcription services was lowest in the Amazon Transcribe transcriptions at 8.9% (IQR 6.9%-11.6%); followed by Whisper at 14.8% (IQR 11.1%-19.7%); and lastly, Zoom-Otter AI at 19.2% (IQR 15.1-24.8). A Wilcoxon signed rank test applied to the WER showed a statistically significant difference between Amazon Transcribe and Zoom-Otter AI ($P<.001$), Amazon Transcribe and Whisper ($P<.001$), and Zoom-Otter AI and Whisper ($P<.001$). Figure 4 provides the distributions of the WER for each service. Amazon Transcribe had the lowest WER of all automatic transcription services and was then compared with the paid human transcription provided by Rev.

Figure 4. Distributions of the word error rates (WER) for Amazon Transcribe, Zoom-Otter AI, and Whisper transcriptions. Dots indicate the actual values of each WER for each of the (N=65) interviewees. Distributions are estimated from these values using a kernel density estimate. Box plots indicate the median with the 25th and 75th percentiles. *P* values are for a Wilcoxon signed rank test applied between distributions in a pair-wise manner.



Rev Human Transcription

For human transcription provided by Rev, the median WER was at 7.6% (IQR 5.4%-11.3%), and a Wilcoxon paired signed rank test against the Amazon Transcribe indicated a statistically significant difference between these 2 distributions of WERs (*P*<.001). The distributions of Amazon Transcribe's and Rev's WER are shown in Figure 5. Table 2 provides the medians, means, and IQRs for all transcription services. However, the errors for each method are not necessarily for the same underlying types of words. It may be incorrect to conclude that either approach is better solely based on aggregate error rate comparisons without deeper analysis of the most common categories or types of erroneous words and without considering their importance for diagnosis.

Figure 5. Distributions of the word error rates (WERs) for Amazon Transcribe and Rev human transcription. Dots indicate the actual values of each WER of each of the (N=65) interviewees with shortened audio. Distributions were estimated from the raw values using a kernel density estimate. Box plots indicate the median WER with the 25th and 75th percentiles. *P*<.001 is for a Wilcoxon signed rank test applied between the 2 distributions.

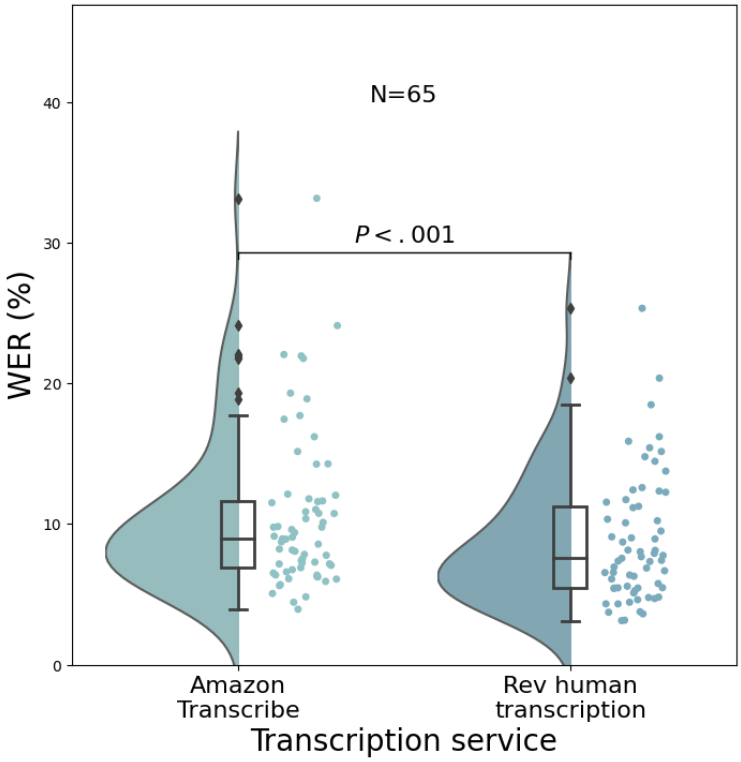


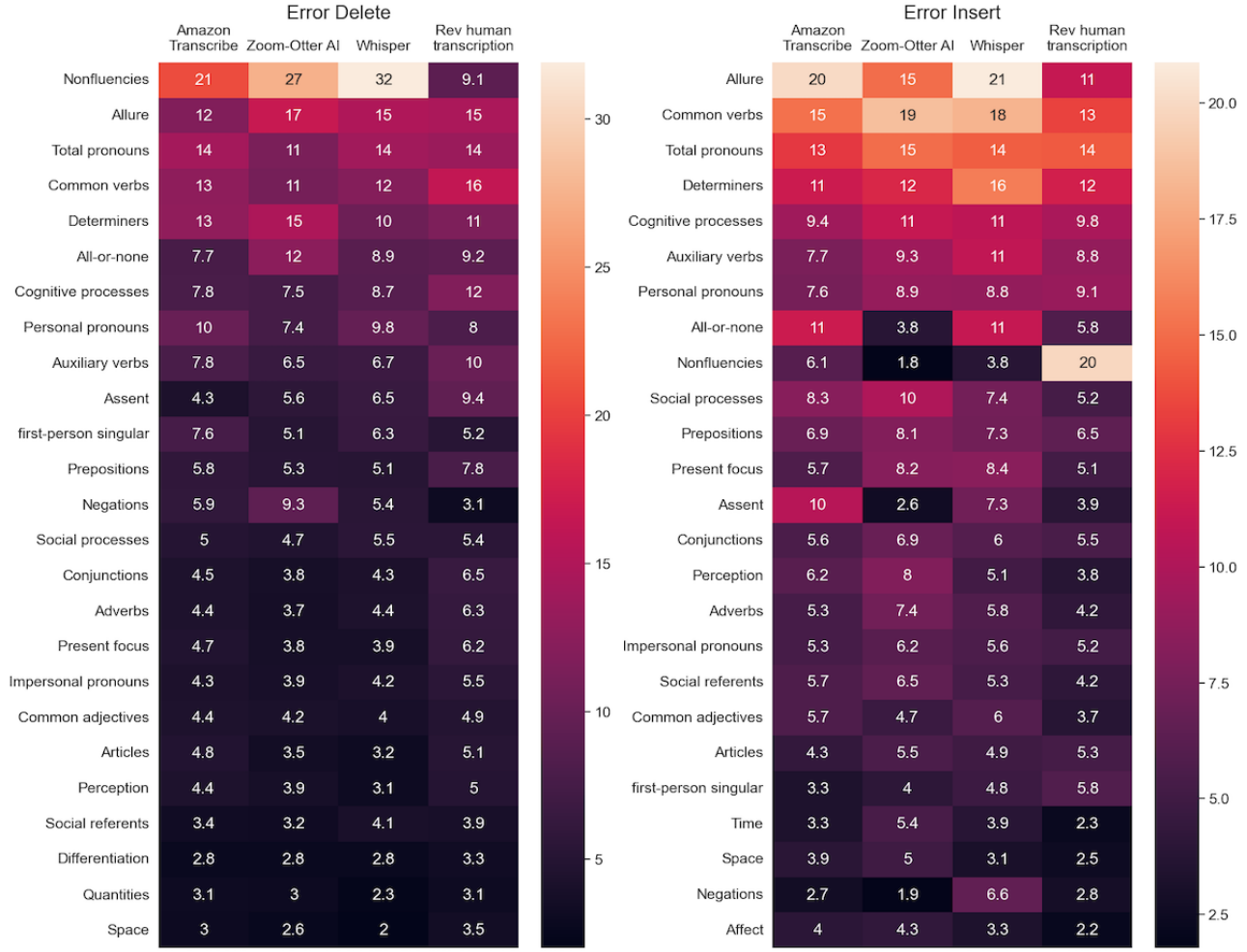
Table 2. Word error rate's (WER) median (IQR) and mean (SD) as percentages.

Transcription service	WER (%), median (IQR)	WER (%), mean (SD)
Amazon Transcribe	8.9 (6.9-11.6)	10.5 (5.4)
Zoom-Otter AI	19.2 (15.1-24.8)	21.6 (9.1)
Whisper	14.8 (11.1-19.7)	16.7 (7.0)
Rev	7.6 (5.4-11.3)	8.8 (4.5)

Error Insert and Error Delete by LIWC Category

While the WER tallies errors, it fails to distinguish between specific types of errors, such as omissions and insertions. In other words, it does not accurately differentiate between instances where words are mistakenly overrepresented or underrepresented. To further understand the types of errors within these transcription services, the Error Delete and Error Insert percentages were analyzed by LIWC category for all 4 transcription services. There are general overarching LIWC categories, such as "Conversation, Cognition, and Function," that are composed of subcategories. For example, Conversation includes categories such as assent and nonfluencies [22]. Nonfluencies refer to words such as "oh," "um," and "i i," which are often used in speech [22]. Cognition is a newly added overarching category that is meant to reflect differing ways people think or refer to their thinking, such as through the subcategory of all-or-none thinking [22]. Function words are made up of short, common words such as pronouns, verbs, and determiners [22]. Dropping the abovementioned general overarching categories, the 25 categories with the highest Error Delete and Error Insert based on the average values of all 4 transcriptions are shown through the heat maps in Figure 6. To further visualize these errors, spider plots were created for the top 9 LIWC categories that fell under the overarching categories of Conversation, Cognition, and Function, which were found to have a high percentage of Error Insert and Error Delete. Figure 7 visualizes the error delete by LIWC category, and Figure 8 visualizes the Error Insert by LIWC category.

Figure 6. Heat map of top Linguistic Inquiry and Word Count (LIWC) categories (not including the general categories like Conversation, Cognition, and Function) sorted based on average values of all 4 transcriptions for Error Insert (insertion + substitution insertion) as well as Error Delete (deletion and substitution deletion). The numbers represent the percentile contribution of each category to the respective error type, either Error Delete or Error Insert. Major categories (Conversation, Cognition, and Function) have been dropped to accentuate their subcategories (nonfluencies and assent subcategories of Conversation; all-or-none and cognitive processes sub-categories of Cognition; common verbs, auxiliary verbs, determiners, personal pronouns, and total pronouns subcategories of Function).



Discussion

Principal Findings

Using a mental health research assessment in an over-the-internet format, we found significant differences in performance, as measured by WER, between services from Amazon Transcribe, Zoom-Otter AI, Whisper, and Rev human transcription. We found no significant differences in transcription errors between the control and MHC groups for Amazon Transcribe, Zoom-Otter AI, and Whisper. Notably, Amazon Transcribe performed significantly better than other tested ASR systems and was very similar to human transcription services, with a marginally higher median transcription WER (8.9%, IQR 6.9%-11.6% vs 7.6%, IQR 5.4%-11.3%).

WER Performance Across Clinical Categories

To date, there has been a paucity of literature evaluating the performance of ASR across different psychiatric clinical categories. In theory, systematic performance differences could result in discrimination against a particular subset and limit clinical applicability. To alleviate this concern, analysis of the performance of each clinical category is crucial and allows us to detect differences in performance. We found that the services do not seem to discriminate between any particular clinical category, at least in a sample of controls and outpatients (Figure 3). The lack of a statistically significant differences ($P \geq .05$) between clinical categories for each transcription service could potentially be explained by a relatively small sample size compared to the large spread of differences. Further research is needed with larger samples; separate analyses of more specific diagnostic categories (individuals with bipolar disorder or schizophrenia-spectrum disorders); and the inclusion of diverse samples encompassing individuals with various cultural, racial, and ethnic backgrounds. In this study, we are focused on comparing the performance of different transcription services using the WER metric and investigating any potential systematic biases present in these errors. While our analysis presented here does not aim to use the transcriptions to differentiate between control and mental health groups, we anticipate in the future that natural language processing methods, and large language models in particular, will be used on the transcribed data to identify mental health status. The performance of these complex models as a function of WER rate in particular word types is unknown, and a statistically insignificant difference may still

be amplified by a complex classifier or predictor. Therefore, each of these must be stress-tested as a function of the specific categories of transcription errors.

WER Performance by Transcription Service

The WERs of 3 automatic services illustrated that Amazon Transcribe outperformed Zoom-Otter AI and Whisper transcriptions. However, when comparing Amazon Transcribe and a human transcription service, Rev performed statistically significantly better despite having a similar median WER. Nevertheless, the WERs were similar to Amazon Transcribe, with a notable difference in nonfluencies (Figures 7 and 8). It is unclear whether this difference is clinically significant, and this is left to future work when we have collected more data. Although these differences were found to be significant, the gap in performance between ASR and human transcription services appears to be narrowing. With overall improvements in ASRs and significantly different service costs compared with human transcription services, ASRs may be the preferred choice if selected wisely.

Figure 7. Top Linguistic Inquiry and Word Count (LIWC) categories for Error Delete (deletion and substitution deletion). For each transcription service, these show what percent of Error Delete words are in each category. Different colors represent different transcription services. The outermost labels are the major categories (Conversation, Cognition, and Function); each can be divided into subcategories (nonfluencies and assent for Conversation; all-or-none and cognitive processes for Cognition; common verbs, auxiliary verbs, determiners, personal pronouns, and total pronouns for Function).

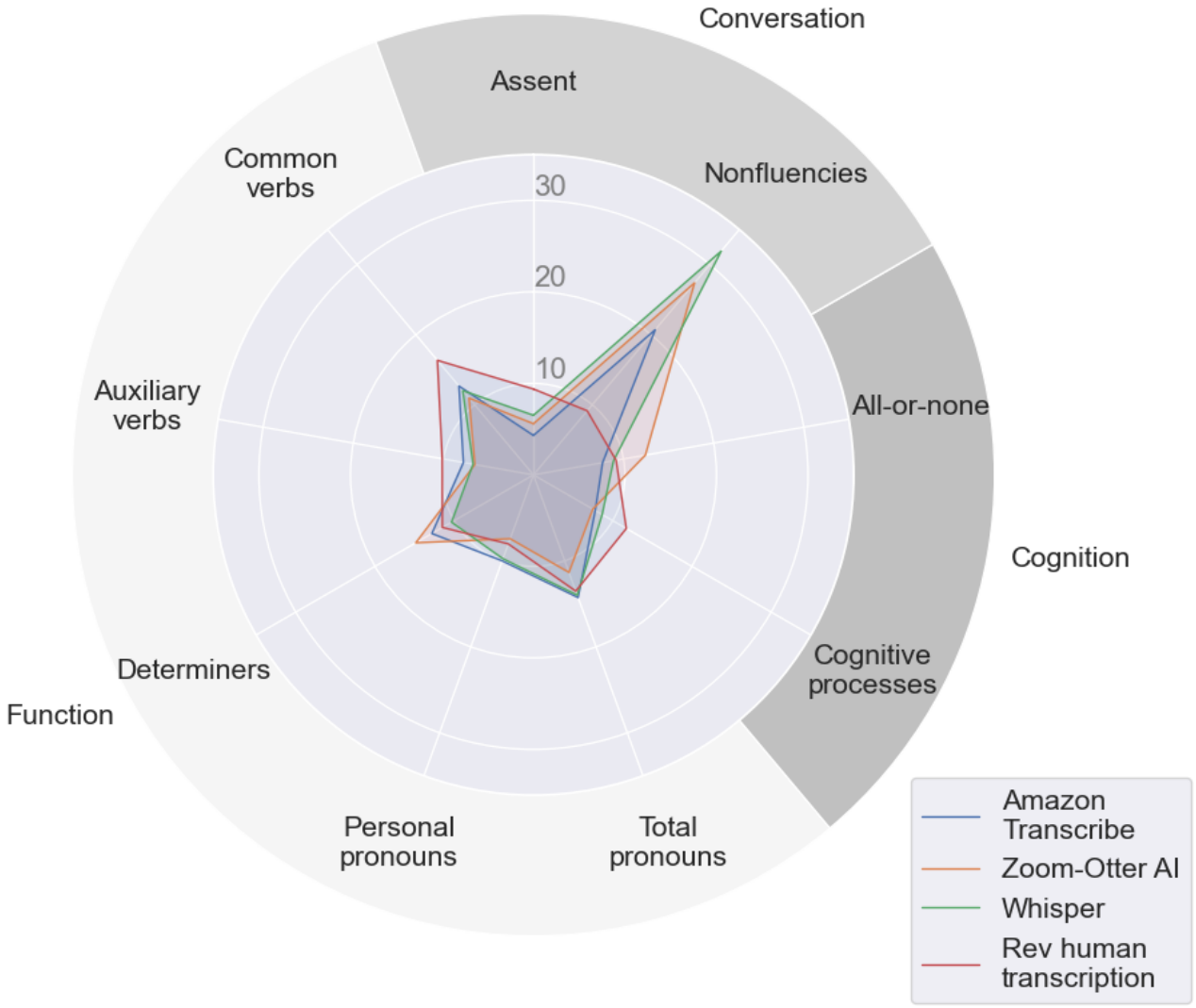
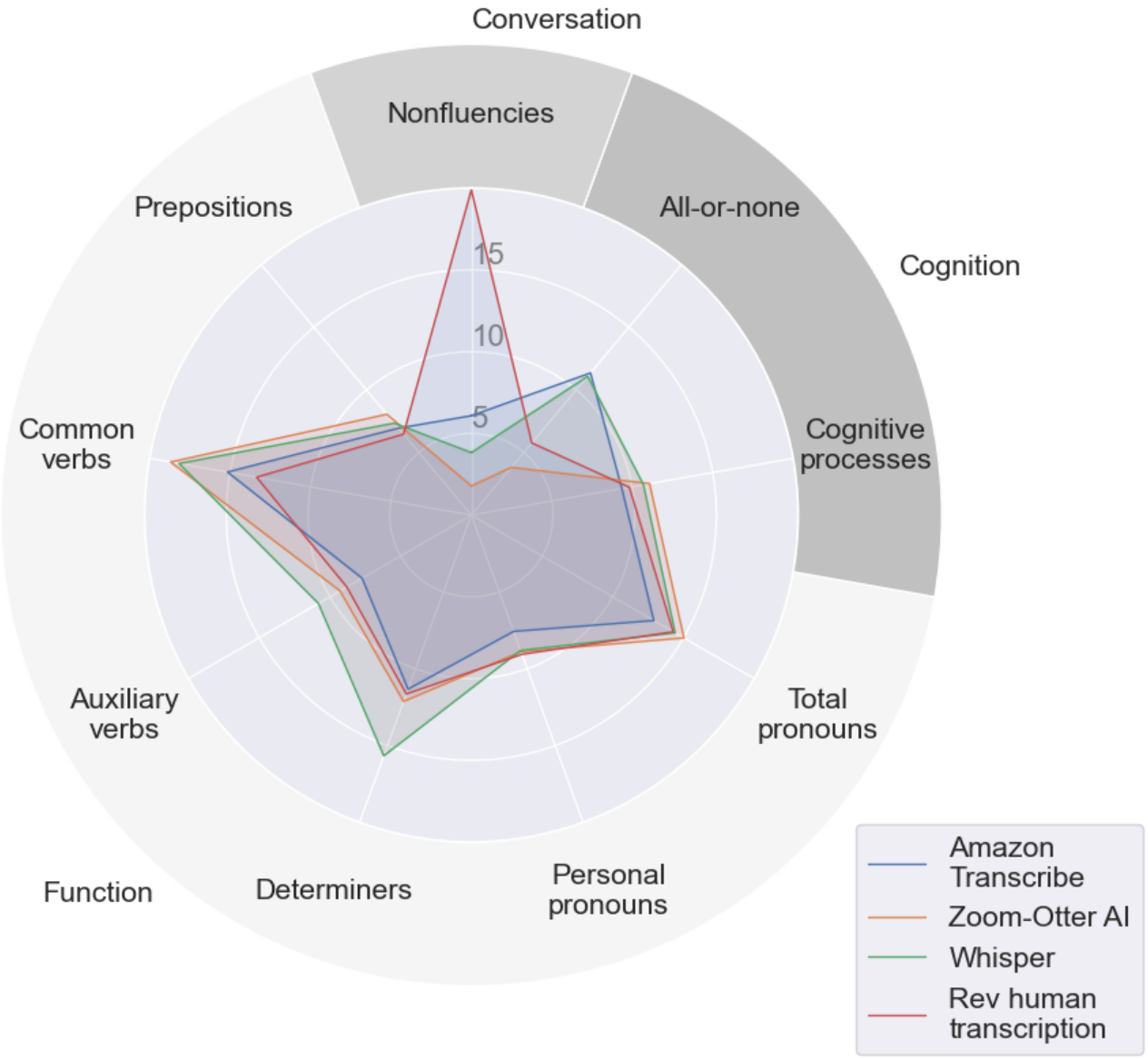


Figure 8. Top Linguistic Inquiry and Word Count (LIWC) categories for Error Insert (insertion + substitution insertion). For each transcription service, these show what percent of Error Insert words are in each category. Assorted colors represent different transcription services. The outermost labels are the major categories (Conversation, Cognition, and Function); each can be divided into subcategories (nonfluencies for Conversation; all-or-none and cognitive processes for Cognition; prepositions, common verbs, auxiliary verbs, determiners, personal pronouns, and total pronouns for Function).



Cost and Scalability

Although the Rev human transcription service provided a statistically significantly lower WER ($P<.01$) with a promised delivery time of fewer than 12 hours, it is costly (US $1.50 per min, 50 times more expensive compared with Amazon Transcribe, which is US $0.024 per min) and not as fast as automatic transcriptions (which are almost instantaneous). Also, when scalability is a concern, any human-dependent process can be a rate-limiting step. Moreover, the performance of Amazon Transcribe and human transcription was comparable when looking at both the distributions and the median WER for Amazon at 8.9% (IQR 6.9%-11.6%) versus Rev at 7.6% (IQR 5.4%-11.3%). Our results provide some tentative justification that Amazon Transcribe may serve as a valuable substitute for human transcription, albeit with a few caveats (see the *Limitations and Future Directions* section).

Errors by LIWC Category

It is important to not only understand the overall WER of these various transcription services but to contextualize the types of words being deleted or inserted in error. As shown through the heat maps and spider plots, the majority of errors fall within 3 overarching LIWC categories: Function, Cognition, and Conversation. Pronouns, specifically first-person and personal pronouns, are often cited as potential predictors of various mental illnesses, most notably depression [23,25,44]. Thus, depending on the population being examined, the use of ASR within LIWC research may provide a misrepresentation of pronoun usage. The same can be seen for words related to cognition, specifically in the all-or-none and cognitive processes categories. There were also a large number of nonfluencies that were deleted in error by the ASR platforms. Nonfluencies have been found to correlate strongly with depression and anxiety in both written and spoken text [45,46]. The nonfluencies category in LIWC is comprised of only 21 words [22], yet it has one of

the highest Error Delete percentages for all 3 ASR transcription services. Other LIWC categories commonly noted as potential predictors of mental illness include numerous affect-related categories, such as sadness or negative emotion words [23,24]. However, these affect-related categories were not found to be largely represented in Error Delete or Error Insert within this sample. This could result from this category being less used by participants in these transcriptions or from these types of words being correctly transcribed.

Limitations and Future Directions

While this study revealed significant differences among ASR transcription services, it is important to acknowledge that 1 of the 3 ASR services (Zoom-Otter AI) used live transcription to produce the transcriptions. This approach could potentially increase the difficulty of the task and lead to reduced accuracy due to limitations in using upcoming parts of speech. However, Amazon Transcribe and Whisper transcribe the audio data in a rolling buffer rather than using the entire recording, and Zoom-Otter AI also has a delay and retrospective correction buffer, which potentially makes the algorithms' implementations essentially equivalent. We note, however, that pauses and silences longer than 1 second were therefore not removed before transcription by Zoom-Otter AI.

Another potential limitation lies in the fact that the 2 human transcribers had the opportunity to see the agreements and disagreements of the Amazon Transcribe and Zoom-Otter AI outputs before making their corrections. (As noted in the methods, we followed earlier work of Neamatullah et al [39] for deidentification of medical data by combining the strengths of sensitive algorithms and specific humans, which was shown to be highly effective.) This leads to the potential that both human overreaders were "primed" in some manner (and in the same manner) by the transcriptions of the 2 comparative algorithms. However, humans are also primed in some manner by their formative experiences, and there is no evidence to suggest that this is necessarily any more "unbiased" than the algorithms with a human overread. In fact, our Rev human transcription results indicate that humans are very similar to algorithms (on average) but make different mistakes. To identify and remove any residual bias, we would have to have a large (or unknown) number of humans from varying backgrounds and cultures. In other medical data experiments combining human decisions, we have found that the number of individuals required to provide a confident decision or label can be as high as 9 [47]. While identifying the exact number of humans needed to create a near-perfect (or unbiased) transcription is an exciting potential research avenue, it is beyond the scope (and means) of this study.

Finally, we note that this study's findings are limited by the relatively small sample size and the demographics skewing toward White, highly educated females. It is also important to note that there was relatively little background noise in our recordings, and some individuals, particularly those with fewer resources, may not be able to find quiet locations to talk over video. However, innovations in background noise cancellation, particularly other voices, have improved enormously over the last few years and somewhat mitigate this issue.

Given the increasing use of ASR transcription services in clinical settings, independent evaluations of WERs are crucial to ensuring these services are accurate for specific contexts. It is currently unclear whether a given WER would alter the clinical decision-making process or outcomes for a particular patient or for a given algorithm that uses the transcribed text as input. Future research should further examine the types of errors these transcription services are getting and whether the contents of those words are clinically impactful. Even seemingly minor errors have the potential to cause clinically significant errors in diagnoses or treatment recommendations, for example, if ASRs are not thoroughly tested in the context in which they are to be used. Moreover, biases in ASR algorithms (eg, those that have been trained on "standard English") may further exacerbate diagnostic disparities or lead clinicians to select improper treatments, particularly in the case of underrepresented minority groups, women, nonnative English speakers, and individuals from low socioeconomic or low literacy groups. It will be important to incorporate measures for such subgroups and implement methods to mitigate these disparities at both the algorithmic and user-implementation levels. Such analysis will require a much larger corpus of data. These issues will be the subject of subsequent publications on our corpus as we continue to increase the size and diversity of our population.

Conclusions

The gap in performance between ASR and human transcription services continues to narrow, and our results appear to indicate that they are close to being equivalent. This is consistent with the trend in the literature where, depending on the context, WER has dropped from around 30% in the early 2000s [5] to 10% to 15% in the 2010s [6] to under 10% in recent years [7]. With overall improvements in ASRs and significantly lower service costs (around 50 times less expensive) compared with human transcription services, ASRs are increasingly likely to be the preferred choice for medical transcription. However, further research needs to evaluate various clinical populations with larger, more diverse sample sizes to determine whether these errors impact the analysis and usability of these ASR transcriptions in applied settings. Of course, human transcription services should be subject to the same analysis, as they also have the potential to generate the same biases and errors.

Acknowledgments

The research presented in this publication was supported in part by Imagine, Innovate, and Impact Funds from the Emory School of Medicine and through a Georgia Clinical and Translational Science Alliance National Institutes of Health award (UL1-TR002378).

Conflicts of Interest

ROC received institutional research funding from Alkermes, Roche, Karuna, and Otsuka; is a consultant to Saladax Biomedical and the American Psychiatric Association; and is a speaker for Clinical Care Options. In the last 3 years, GDC has received research funding from the National Science Foundation, National Institutes of Health, Nextsense Inc, LifeBell AI, and Otsuka UA and unrestricted donations from AliveCor Inc, Amazon Research, the Center for Discovery, the Gates Foundation, Google, the Gordon and Betty Moore Foundation, MathWorks, Microsoft Research, One Mind Foundation, the Rett Research Foundation, and Samsung Research. He has a financial interest in AliveCor Inc and Nextsense Inc and has received licensing royalties from Linus Health, Medibio Inc, and MindChild Medical Inc. He is also the chief technology officer of MindChild Medical and the chief strategy officer of LifeBell AI, and he has ownership in both companies. These relationships are unconnected to the work presented here.

References

1. Key substance use and mental health indicators in the United States: results from the 2020 national survey on drug use and health. Substance Abuse and Mental Health Services Administration. Rockville, MD: Center for Behavioral Health Statistics, Quality, Substance Abuse, and Mental Health Services Administration; 2021. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35325/NSDUHFFRPDFWHTMLFiles2020/2020NSDUHFFR1PDFW102121.pdf [accessed 2023-10-11]
2. Aggarwal R, Balon R, Beresin EV, Coverdale J, Morreale MK, Guerrero APS, et al. Addressing psychiatry workforce needs: where are we now? Acad Psychiatry 2022;46(4):407-409 [FREE Full text] [doi: 10.1007/s40596-022-01690-5] [Medline: 35882768]
3. Behavioral Health Workforce Projections, 2020-2035. National Center for Health Workforce Analysis. 2022. URL: https://bhw.hrsa.gov/data-research/projecting-health-workforce-supply-demand [accessed 2023-10-11]
4. Pheister M, Cowley D, Sanders W, Keeble T, Lu F, Pershern L, et al. Growing the psychiatry workforce through expansion or creation of residencies and fellowships: the results of a survey by the AADPRT Workforce Task Force. Acad Psychiatry 2022;46(4):421-427 [FREE Full text] [doi: 10.1007/s40596-021-01509-9] [Medline: 34292538]
5. Zechner K, Waibel A. Minimizing word error rate in textual summaries of spoken language. 2000 Presented at: 1st Meeting of the North American Chapter of the Association for Computational Linguistics; April 29-May 4, 2000; Seattle Washington p. 186-193.
6. Su H, Li G, Yu D, Seide F. Error back propagation for sequence training of context-dependent deep networks for conversational speech transcription. 2013 Presented at: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; May 26-31, 2013; Vancouver, BC, Canada p. 6664-6668. [doi: 10.1109/icassp.2013.6638951]
7. Zhou L, Blackley SV, Kowalski L, Doan R, Acker WW, Landman AB, et al. Analysis of errors in dictated clinical documents assisted by speech recognition software and professional transcriptionists. JAMA Netw Open 2018;1(3):e180530 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.0530] [Medline: 30370424]
8. Panayotov V, Chen G, Povey D, Khudanpur S. Librispeech: an ASR corpus based on public domain audio books. 2015 Presented at: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); April 19-24, 2015; South Brisbane, QLD, Australia p. 5206-5210. [doi: 10.1109/icassp.2015.7178964]
9. Zeghidour N, Xu Q, Liptchinsky V, Usunier N, Synnaeve G, Collobert R. Fully convolutional speech recognition. arXiv. Preprint posted online on April 09 2019 2018 [FREE Full text]
10. Irie K, Zeyer A, Schlüter R, Ney H. Language modeling with deep transformers. arXiv. Preprint posted online on July 11 2019 2019 [FREE Full text] [doi: 10.21437/interspeech.2019-2225]
11. Radford A, Kim JW, Xu T, Brockman G, McLeavey C, Sutskever I. Robust speech recognition via large-scale weak supervision. 2023 Presented at: Proceedings of the 40th International Conference on Machine Learning; 2023; Honolulu, Hawaii, USA p. 28492-28518 URL: https://proceedings.mlr.press/v202/radford23a/radford23a.pdf
12. Park DS, Chan W, Zhang Y, Chiu CC, Zoph B, Cubuk ED, et al. Specaugment: a simple data augmentation method for automatic speech recognition. ArXiv. Preprint posted online on December 03 2019 2019 [FREE Full text] [doi: 10.21437/interspeech.2019-2680]
13. van Buchem MM, Boosman H, Bauer MP, Kant IMJ, Cammel SA, Steyerberg EW. The digital scribe in clinical practice: a scoping review and research agenda. NPJ Digit Med 2021;4(1):57 [FREE Full text] [doi: 10.1038/s41746-021-00432-5] [Medline: 33772070]
14. Johnson M, Lapkin S, Long V, Sanchez P, Suominen H, Basilakis J, et al. A systematic review of speech recognition technology in health care. BMC Med Inform Decis Mak 2014;14:94 [FREE Full text] [doi: 10.1186/1472-6947-14-94] [Medline: 25351845]
15. Alexander AG, Ballou KA. Work-Life balance, burnout, and the electronic health record. Am J Med 2018;131(8):857-858 [FREE Full text] [doi: 10.1016/j.amjmed.2018.02.033] [Medline: 29608876]
16. Collier R. Electronic health records contributing to physician burnout. CMAJ 2017;189(45):E1405-E1406 [FREE Full text] [doi: 10.1503/cmaj.109-5522] [Medline: 29133547]

17. Mertz H. Electronic health record reform: an alternative response to physician burnout. Am J Med 2021;134(9):e498 [FREE Full text] [doi: 10.1016/j.amjmed.2021.04.022] [Medline: 34462089]
18. Flemotomos N, Martinez VR, Chen Z, Singla K, Ardulov V, Peri R, et al. Automated evaluation of psychotherapy skills using speech and language technologies. Behav Res Methods 2022;54(2):690-711 [FREE Full text] [doi: 10.3758/s13428-021-01623-4] [Medline: 34346043]
19. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Commun 2015;71:10-49 [FREE Full text] [doi: 10.1016/j.specom.2015.03.004]
20. Holmlund TB, Chandler C, Foltz PW, Cohen AS, Cheng J, Bernstein JC, et al. Applying speech technologies to assess verbal memory in patients with serious mental illness. NPJ Digit Med 2020;3:33 [FREE Full text] [doi: 10.1038/s41746-020-0241-7] [Medline: 32195368]
21. Xu S, Yang Z, Chakraborty D, Chua YHV, Dauwels J, Thalmann D, et al. Automated verbal and non-verbal speech analysis of interviews of individuals with schizophrenia and depression. Annu Int Conf IEEE Eng Med Biol Soc 2019;2019:225-228. [doi: 10.1109/EMBC.2019.8857071] [Medline: 31945883]
22. Boyd RL, Ashokkumar A, Seraj S, Pennebaker JW. The development and psychometric properties of LIWC-22. Austin: University of Texas; 2022. URL: https://www.liwc.app/static/documents/LIWC-22%20Manual%20-%20Development%20and%20Psychometrics.pdf [accessed 2023-10-11]
23. Gortner EM, Rude SS, Pennebaker JW. Benefits of expressive writing in lowering rumination and depressive symptoms. Behav Ther 2006;37(3):292-303 [FREE Full text] [doi: 10.1016/j.beth.2006.01.004] [Medline: 16942980]
24. Morales MR, Levitan R. Speech vs. text: a comparative analysis of features for depression detection systems. 2016 Presented at: 2016 IEEE Spoken Language Technology Workshop (SLT); December 13-16, 2016; San Diego, CA, USA p. 136-143. [doi: 10.1109/slt.2016.7846256]
25. Burkhardt HA, Alexopoulos GS, Pullmann MD, Hull TD, Areán PA, Cohen T. Behavioral activation and depression symptomatology: longitudinal assessment of linguistic indicators in text-based therapy sessions. J Med Internet Res 2021;23(7):e28244 [FREE Full text] [doi: 10.2196/28244] [Medline: 34259637]
26. Buck B, Penn DL. Lexical characteristics of emotional narratives in schizophrenia: relationships with symptoms, functioning, and social cognition. J Nerv Ment Dis 2015;203(9):702-708 [FREE Full text] [doi: 10.1097/NMD.0000000000000354] [Medline: 26252823]
27. Fineberg SK, Leavitt J, Deutsch-Link S, Dealy S, Landry CD, Pirruccio K, et al. Self-reference in psychosis and depression: a language marker of illness. Psychol Med 2016;46(12):2605-2615 [FREE Full text] [doi: 10.1017/S0033291716001215] [Medline: 27353541]
28. Health Insurance Portability and Accountability Act of 1996 (HIPAA). 1996. URL: https://www.cdc.gov/phlp/publications/topic/hipaa.html [accessed 2023-10-11]
29. Miner AS, Haque A, Fries JA, Fleming SL, Wilfley DE, Wilson GT, et al. Assessing the accuracy of automatic speech recognition for psychotherapy. NPJ Digit Med 2020;3:82 [FREE Full text] [doi: 10.1038/s41746-020-0285-8] [Medline: 32550644]
30. Cotes RO, Boazak M, Griner E, Jiang Z, Kim B, Bremer W, et al. Multimodal assessment of schizophrenia and depression utilizing video, acoustic, locomotor, electroencephalographic, and heart rate technology: protocol for an observational study. JMIR Res Protoc 2022;11(7):e36417 [FREE Full text] [doi: 10.2196/36417] [Medline: 35830230]
31. ResearchMatch. URL: https://www.researchmatch.org/ [accessed 2023-10-11]
32. Moretti RJ, Rossini ED, Hilsenroth MJ, Segal DL. The Thematic Apperception Test (TAT). In: Hersen M, editor. Comprehensive Handbook of Psychological Assessment, Volume 2: Personality Assessment. Hoboken, NJ: John Wiley & Sons, Inc; 2004:356-371.
33. Benton AL, Hamsher KD, Sivan AB. Multilingual Aphasia Examination: Manual of Instructions. Iowa City, IA: AJA Associates; 1994.
34. Goodglass H, Kaplan E. The Assessment of Aphasia and Related Disorders. Philadelphia: Lea & Febiger; 1972.
35. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33;quiz 34-57 [FREE Full text] [Medline: 9881538]
36. Cohen SR, Sawatzky R, Russell LB, Shahidi J, Heyland DK, Gadermann AM. Measuring the quality of life of people at the end of life: the McGill quality of life questionnaire-revised. Palliat Med 2017;31(2):120-129 [FREE Full text] [doi: 10.1177/0269216316659603] [Medline: 27412257]
37. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097 [FREE Full text] [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
38. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
39. Neamatullah I, Douglass MM, Lehman LH, Reisner A, Villarroel M, Long WJ, et al. Automated de-identification of free-text medical records. BMC Med Inform Decis Mak 2008;8:32 [FREE Full text] [doi: 10.1186/1472-6947-8-32] [Medline: 18652655]

40. Klakow D, Peters J. Testing the correlation of word error rate and perplexity. Speech Commun 2002;38(1-2):19-28 [FREE Full text] [doi: 10.1016/s0167-6393(01)00041-3]
41. Park Y, Patwardhan S, Visweswariah K, Gates SC. An empirical analysis of word error rate and keyword error rate. Interspeech 2008:2070-2073 [FREE Full text] [doi: 10.21437/interspeech.2008-537]
42. Evaluate your speech-to-text system with similarity measures such as Word Error Rate (WER). jitsi. Jiwer. 2022. URL: https://github.com/jitsi/jiwer [accessed 2022-04-25]
43. The levenshtein python C extension module contains functions for fast computation of levenshtein distance and string similarity. Max Bachmann. Levenshtein. URL: https://github.com/maxbachmann/Levenshtein [accessed 2022-04-25]
44. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]
45. O'Dea B, Boonstra TW, Larsen ME, Nguyen T, Venkatesh S, Christensen H. The relationship between linguistic expression in blog content and symptoms of depression, anxiety, and suicidal thoughts: a longitudinal study. PLoS One 2021;16(5):e0251787 [FREE Full text] [doi: 10.1371/journal.pone.0251787] [Medline: 34010314]
46. Weintraub MJ, Posta F, Ichinose MC, Arevian AC, Miklowitz DJ. Word usage in spontaneous speech as a predictor of depressive symptoms among youth at high risk for mood disorders. J Affect Disord 2023;323:675-678 [FREE Full text] [doi: 10.1016/j.jad.2022.12.047] [Medline: 36528134]
47. Zhu T, Johnson AEW, Behar J, Clifford GD. Crowd-sourced annotation of ecg signals using contextual information. Ann Biomed Eng 2014;42(4):871-884 [FREE Full text] [doi: 10.1007/s10439-013-0964-6] [Medline: 24368593]

Abbreviations

ASR: automatic speech recognition
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
LIWC: Linguistic Inquiry and Word Count
MHC: mental health condition
MINI: Mini-International Neuropsychiatric Interview
TAT: thematic apperception test
WER: word error rate

Edited by J Torous; submitted 26.04.23; peer-reviewed by H Burkhardt, L Dupuy; comments to author 06.07.23; revised version received 25.08.23; accepted 12.09.23; published 31.10.23.

Please cite as:
Seyedi S, Griner E, Corbin L, Jiang Z, Roberts K, Iacobelli L, Milloy A, Boazak M, Bahrami Rad A, Abbasi A, Cotes RO, Clifford GD Using HIPAA (Health Insurance Portability and Accountability Act)–Compliant Transcription Services for Virtual Psychiatric Interviews: Pilot Comparison Study
JMIR Ment Health 2023;10:e48517
URL: https://mental.jmir.org/2023/1/e48517
doi:10.2196/48517
PMID:37906217

Original Paper

Investigating How People Who Self-harm Evaluate Web-Based Lived Experience Stories: Focus Group Study

Lizzy Winstone[1], PhD; Becky Mars[1,2], PhD; Jennifer Ferrar[3], PhD; Paul Moran[1,2], PhD; Ian Penton-Voak[2,3], PhD; Lydia Grace[4], PhD; Lucy Biddle[1,2], PhD

[1]Population Health Sciences, University of Bristol, Bristol, United Kingdom

[2]NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust, Bristol, United Kingdom

[3]School of Psychological Science, University of Bristol, Bristol, United Kingdom

[4]Samaritans, Surrey, United Kingdom

Corresponding Author:
Lizzy Winstone, PhD
Population Health Sciences, University of Bristol
Oakfield House
Oakfield Grove
Bristol, BS8 2BN
United Kingdom
Phone: 44 07843301453
Email: lizzy.winstone@bristol.ac.uk

Abstract

Background: The positive and negative effects of interacting with web-based content on mental health, and especially self-harm, are well documented. Lived experience stories are one such type of static web-based content, frequently published on health care or third-sector organization websites, as well as social media and blogs, as a form of support for those seeking help via the web.

Objective: This study aimed to increase understanding about how people who self-harm engage with and evaluate web-based lived experience stories.

Methods: Overall, 4 web-based focus groups were conducted with 13 people with recent self-harm experience (aged 16-40 years). In total, 3 example lived experience stories were read aloud to participants, who were then asked to share their reactions to the stories. Participants were also encouraged to reflect on stories previously encountered on the web. Data were analyzed thematically.

Results: Overall, 5 themes were generated: stories of recovery from self-harm and their emotional impact, impact on self-help and help-seeking behaviors, identifying with the narrator, authenticity, and language and stereotyping.

Conclusions: Lived experience stories published on the web can provide a valuable form of support for those experiencing self-harm. They can be motivating and empowering for the reader, and they have the potential to distract readers from urges to self-harm. However, these effects may be moderated by age, and narratives of recovery may demoralize older readers. Our findings have implications for organizations publishing lived experience content and for community guidelines and moderators of web-based forums in which users share their stories. These include the need to consider the narrator's age and the relatability and authenticity of their journey and the need to avoid using stigmatizing language.

(JMIR Ment Health 2023;10:e43840) doi:10.2196/43840

KEYWORDS

self-harm; lived experience stories; web-based support; self-help; recovery; focus groups

Introduction

Background

Self-harm is an important public health priority, with increased prevalence observed over the past 2 decades, particularly among young women [1]. It comprises both suicidal and nonsuicidal behaviors and is predictive of poor social and emotional outcomes and future suicide attempts [2].

The positive and negative effects of interacting with web-based content on mental health, and especially on self-harm, are well

documented [3,4]. The internet offers an important avenue for supporting people who self-harm, particularly those who are unable or reluctant to access formal, in-person support [5,6]. Increasingly, young people are likely to initially self-disclose self-harm behavior via the web owing to affordances of anonymity and perceptions of being judged less [7,8]. Young people experiencing suicidal feelings are also particularly likely to search for advice or peer support on the web [4,9,10]. However, web-based help seeking is also associated with risks, such as unintentional exposure to graphic or prosuicide and self-harm content and engagement with discussion forums in which self-harm is sometimes normalized and negative feelings are reinforced [3,8,9,11]. Therefore, it is vital that the web-based support content provided is safe, effective, and evidence-based.

Lived experience stories are a frequently used source of information and support on the websites of third-sector or health care organizations and can include user-generated content on social media [5,9]. Stories published by more formal help sites, such as Mind or Samaritans, typically involve a personal narrative of help seeking and recovery. Although limited research has focused on self-harm lived experience stories so far, there is emerging evidence suggesting that these narratives can inspire hope and help seeking in the reader [9]. A recent qualitative study compared *recovery narratives* presented live and recorded (in audio, video, or text-based formats) with adults using statutory mental health services in England [12]. Findings suggested that recorded narratives are valued for their availability *on demand*, for access to narrators without geographical constraint, and for the ability to access support without a need for social interaction [12]. In a mixed methods study including interviews with young people and general practitioners (GPs), Cohen et al [9] highlighted suicide-specific lived experience stories as a potentially useful and engaging source of web-based support. Simultaneously, concerns were raised by a GP that reading about someone else's recovery could lead to negative comparisons and feelings of failure. A young participant suggested that detailed descriptions of the context in which an individual experienced suicidal ideation could be triggering to some readers.

Objectives

Despite some evidence suggesting that creative storytelling [13] or digital storytelling [14] may be beneficial to oneself and others in terms of educating an audience and destigmatizing mental illness, little is known about how people who self-harm use web-based lived experience stories and how this may shape their behavior. With this in mind, we aimed to address the following research question—how do people who self-harm engage with, interpret, and evaluate lived experience stories accessed via the web?

Methods

Design

This was a qualitative focus group study.

Recruitment

Participants for focus groups were recruited from another experimental study [15]. Interested participants were invited to complete an eligibility check and to provide their contact details. Participants were eligible to participate if they were aged ≥16 years, reported self-harm in the past year, and were fluent in English, with no other exclusion criteria. Purposive sampling—based on age and gender—was then used to invite a subset of those participants to participate in the focus groups. Written consent was obtained before each group session and confirmed verbally at the beginning of discussions.

Overall, 4 focus groups were conducted from June 2022 to July 2022, with a total of 13 participants in the United Kingdom. Owing to the sensitive nature of discussions, we limited the number of participants to 4 per group and attempted to assign participants to groups with others of similar age. This was not possible for gender, as only 2 men returned the consent forms. Groups consisted of 2 to 4 participants each.

Procedure

Each focus group lasted between 40 and 60 minutes and was conducted via the web using Zoom (version 5.12.2; Zoom Video Communications, Inc) videoconferencing software. In total, 2 researchers were present to moderate each group. A semistructured approach was used. First, the lead moderator (LW) shared on the screen and read aloud 3 examples of lived experience stories about self-harm (Multimedia Appendix 1), followed by group discussion. The 3 example stories were created by combining extracts of *real-life* stories selected from different sources—including the websites of third-sector organizations, web-based forums, and personal blogs—to reflect a range of help-seeking circumstances. Each of the stories included a recovery narrative and did not include any graphic details of self-harm or encouragement of self-harm. Story 1 focused on self-help strategies and was written from the perspective of an individual aged 21 years. Story 2 did not explicitly report help seeking but conveyed a narrative of nonlinear recovery and a message of hope. Story 3 presented a journey of finding a therapist, medication, and social support. In the second and third stories, the age of the narrator was not specified. The example stories were used to prompt discussion, but participants were also encouraged to think about and discuss their own experience and reflect on other lived experience stories they had encountered on health care or third-sector organization websites or on social media.

Discussion was facilitated by flexible use of a topic guide [16] and included participants' experiences of engaging with lived experience material on the web, their initial reactions to the example stories, specific aspects of stories (both in the examples and more broadly) that were supportive or not supportive, and the impact of reading lived experience stories on one's own thoughts about help seeking and self-harm behavior. Participants received a shopping voucher worth £20 (US $24.64) by way of thanks, in addition to a debriefing sheet and signposting to support services. Discussions were audio-recorded and transcribed verbatim.

Ethics Approval

The authors assert that all procedures contributing to this study comply with the ethical standards of the relevant national and institutional committees on human experimentation and with

the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human participants were approved by the School of Psychological Science research ethics committee at the University of Bristol (reference 10504).

Analysis

Data were thematically analyzed, by following the 6 phases outlined by Braun and Clarke [17]. Anonymized transcripts were checked for accuracy and familiarization and then imported into NVivo (version 1.6.1; QSR International) for coding. LW and LB each openly coded the extracts from each of the 4 focus groups, collaboratively developing a coding frame to identify broad topics around which the codes clustered. Then, LW coded the full data set using the coding framework, added new codes, and iteratively refined the thematic descriptions and boundaries. Discussion with the study team led to further amendments to themes, with the transcripts then revisited to check accuracy. Our results are not presented separately for each story because the focus group discussions combined all 3 example stories and lived experience stories encountered elsewhere.

Results

Overview

Participants (11/13, 85% women and 2/13, 15% men) were aged between 16 and 40 years, and mean age was 24.5 (SD 7.3) years (Table 1).

Analysis generated 5 main themes related to how people with recent experience of self-harm interpret, evaluate, and engage with lived experience stories. These themes were labeled as follows: (1) stories of recovery from self-harm and their emotional impact, (2) impact on self-help and help-seeking behaviors, (3) identifying with the narrator, (4) authenticity, and (5) language and stereotyping.

In the descriptions of these themes in the following sections, we refer mainly to the discussions about specific features of the example stories presented to participants. Where participants referred to previous experiences of encountering stories on the web, these were most often accessed on mental health–related websites, such as Mind, but some participants also considered stories as user-generated content shared in web-based self-harm community forums or social media groups.

Table 1. Summary of characteristics of participants in each focus group (N=13).

Characteristics	Focus group 1 (n=2, 15%)	Focus group 2 (n=4, 31%)	Focus group 3 (n=4, 31%)	Focus group 4 (n=3, 23%)
Gender, n (%)				
Men	0 (0)	0 (0)	1 (25)	1 (33)
Women	2 (100)	4 (100)	3 (75)	2 (67)
Age (years), mean (SD; range)	24 (4.9; 20-27)	32 (8.1; 23-40)	18 (1.8; 16-20)	24 (1; 23-25)
16-18, n (%)	0 (0)	0 (0)	2 (50)	0 (0)
19-21, n (%)	1 (50)	0 (0)	2 (50)	0 (0)
22-25, n (%)	0 (0)	1 (25)	0 (0)	3 (100)
≥26, n (%)	1 (50)	3 (75)	0 (0)	0 (0)

Stories of Recovery From Self-harm and Their Emotional Impact

Most participants described stories of recovery from self-harm—including positive framing of scars—as hopeful, motivating, or inspiring reminders of past struggles, helping to reassure them that they also could eventually recover in time. This included lived experience stories widely encountered on the web, with references made to specific features of the example stories:

> *Story one is just more look towards the future and kind of inspiring, maybe. Like that person did it so, so can I and everyone else...it's definitely good to like know that once again, I am not the only one and it is reassuring and promising that other people can do it so, there is a chance for me too.* [Woman; aged 19 years]

Some spoke about general positive impact on their mood when reading the example stories with a positive outcome or motivational message that encourages patience and strength throughout a recovery journey:

> *...When they like kind of do their reflection at the end saying it is possible, I find that really does just pick me up a bit.* [Woman; aged 20 years]
>
> *...It just feels like the sort of thing, quotes that I'd like relate to or would like on Instagram. It's quite nice little reminders, little pick me ups.* [Woman; aged 24 years]

However, older participants found it challenging if they encountered a story in any context that framed someone's recovery within a certain period or by a certain age. The first example story was a recovery narrative from an individual aged 21 years. For participants aged >21 years, hearing about a young narrator's recovery could be disheartening and could reinforce a sense of having failed to recover themselves:

> *There's definitely an element of, you know, just the age thing again...if this is a teenage problem and I'm in my 20's that means there's no hope for me, that means I'm doomed, I'll be like this forever.* [Woman; aged 23 years]

Some participants described the inclusion of numeric metrics in a lived experience story as being potentially problematic because it could encourage demoralizing comparisons. In discussions about the example stories presented, this commonly referred to the age of the narrator. Participants also reflected on the second example story and other stories they had encountered on social media, which included the number of days the narrator had not self-harmed:

> *Sometimes when people put the numbers or they're like, "oh, I am so and so days clean" it's quite hard because you feel like you are comparing all the time.* [Woman; aged 20 years]

Impact on Self-help and Help-Seeking Behaviors

Several participants reported having actively sought out lived experience stories on organizational websites or on social media as a form of help seeking. This could be to lift their mood (refer to the *Stories of Recovery From Self-harm and Their Emotional Impact* section), as a means of distracting themselves from the immediate urge to self-harm, or to reduce feelings of social isolation:

> *It's usually when I am feeling really low and that I don't really know what to do with myself, you know, I don't want to self-harm, I want to do something else. So, I try to see who else out there has gotten through it and just to see if there is hope out there because in those moments, there is literally nothing I would rather do than basically punish myself for whatever I am feeling...I think they just kind of calm me down enough to be able to actually do something that is healthier for me than self-harm.* [Woman; aged 16 years]

Although some participants reflected on the usefulness of reading lived experience stories on the web during urges to self-harm, others—reflecting on the example stories—discussed low mood as a barrier to the positive impact of a story. Several participants suggested that a reader's frame of mind would influence their perception of a narrative or advice:

> *Once you have read this over and over and you think this is, in the immediate moment, the be all and end all of what is available to help you, it becomes a bit like, "fantastic, you can tell me the strategies – I've tried them, they're not working. You can tell me that you're supposed to have a support network – guess what, I don't right now.*" [Man; aged 25 years]

A participant highlighted that, although she considered herself to be recovered, reading lived experience stories on social media was a cue that she should check on her own well-being and speak to her support network:

> *...If I read one of these and I am like, "Oh, actually I have been thinking about this a bit more" because sometimes the sign that you are reading one of these is a sign that actually, you're thinking about it again.* [Woman; aged 20 years]

Reading lived experience stories—regardless of whether they were actively sought—was commonly reported to reduce feelings of isolation. This was often achieved simply by the recognition of others going through similar experiences or feelings. Some participants reflected on engagement with web-based communities to access user-generated stories, which also promoted a sense of social inclusion:

> *I'm just trying to find a place where I'm feeling less alone with these feelings.* [Woman; aged 28 years]

Practical ideas for self-help strategies were a key feature of the first example story and were described by most participants as particularly useful. Several participants reported encountering similar suggestions in stories read on organizational websites, and they had been inspired to try some suggested strategies in the past. Some participants suggested that further practical advice about where to go for help would also benefit readers:

> *I think I found the first story most helpful because of the suggestions of things that have helped. Things that you can try because it's about thinking, actually what's helped for someone else, that might help me. And having access to as many things as possible that might help because everything is different for everyone else...so that's where I found that one most helpful.* [Woman; aged 40 years]

Highlighting the fundamental concept of self-help in the first example story was also described as motivating and important for those who do not have a support network or access to formal therapy. However, for some participants who reflected on content previously found on the web and the example stories, if the suggested strategies had been tried and tested or if the reader was feeling particularly low, advice could feel clichéd or irritating:

> *...Sometimes [friends who are trying to be supportive] will repeat the exact same phrases over and over and over. It just kind of makes you want to scream and I think stories like this can kind of have a similar effect if it's something you've already seen a lot.* [Man; aged 25 years]

More broadly, lived experience stories on organizational websites could be sought to empower readers to better understand their own feelings and experiences:

> *I was more so actively seeking them perhaps when I wasn't in treatment because I guess, it was a bit unknown or it was difficult to put pieces together. I guess that was more so perhaps with the trauma and sort of seeing how [self-harm] sort of is a coping mechanism for control or that other people [self-harm] for different reasons. So, I guess, last year, it was more when I was trying to figure things out maybe as a reassurance.* [Woman; aged 27 years]

Some participants described using lived experience stories previously accessed on organizational websites as a source of ideas for talking more confidently about their own experiences to others. This included speaking to their support network or therapist and sharing their own experiences to educate and help others:

> *I think I wanted to be able to talk about my own experience and found the best way to do that was to*

> *hear how other people talked about their own experience. Kind of, what do you include, what you don't, what do people not include, those kind of things, how much detail do people go into, things like that.* [Woman; aged 40 years]

By learning about other people's experiences and routes to help seeking, participants appeared to recognize a need for help in others, and in turn, recognized their own need to seek support. This could be through learning about others' routes to therapy. More frequently, participants described the knowledge that they were not alone being a catalyst for help seeking when they had encountered lived experience stories in the past. The following quote refers specifically to the example stories presented to participants:

> *It's hard to connect to just like words on a screen but I think these stories do a pretty good job at that and I think it will help a lot of people to then seek help because they realise, "I am not alone. I am not crazy, and I deserve help for this, and I need help."* [Man; aged 17 years]

Identifying With the Narrator

All participants reported identifying with at least one aspect of ≥1 example lived experience stories. This type of web-based content was suggested to be intrinsically more relatable than advice from a health care professional:

> *You can have a therapist or a professional telling you do X and Y but I found for me, I didn't really listen to them as much as I would another person who may or may not have been a similar age but has lived through a similar thing as me...Hearing it from someone who has gone through it is so much more impactful.* [Woman; aged 20 years]

Most participants suggested that it was more important to be able to relate to a narrator's feelings or experiences within a lived experience story than to their demographic background. By keeping the narrator's age and gender ambiguous in the example stories (the age was specified in the first example story), older readers, in particular, seemed to be able to project their own characteristics onto the story, thus making it feel more relevant and engaging to them. However, some young participants found it easier to relate to stories from similarly aged narrators, which brought them "comfort" (man; aged 17 years).

Some participants felt that organizational websites should provide several stories that reflect a range of experiences to encourage readers to identify and connect with them. However, others noted that being very specific about the circumstances preceding a narrator's self-harm—or about co-occurring mental health disorders or methods of self-harm—could alienate some readers by invalidating their feelings:

> *I don't find it helpful when people give like loads of reasons, like this bad thing happened, and then I self-harmed. Because I found – I don't come across that so much now but when I was younger, I found that made me think I don't deserve to self-harm, I've not had those experiences. And it just felt like another way of invalidating me, but there's not a huge amount of that in these [example] stories...It's not "oh look at me and all the horrible things I've been through."* [Woman; aged 38 years]

Several participants noted the challenge of organizational websites needing to provide stories from multiple perspectives to appeal to the broadest audience, acknowledging that an exhaustive collection would be impossible:

> *...If it's too generic and vague, it's almost like it's not real, but if you make it too specific then obviously you narrow down the people who can perhaps relate to it. But then who's going to want to sit and read through ten different accounts [stories] to find one that matches their own experience?* [Woman; aged 38 years]

Identifying with the narrator also appeared to be an important factor in encouraging help seeking through the recognition of need in others and thus oneself (refer to the *Impact on Self-help and Help-Seeking Behaviors* section) but was simultaneously reported by most participants as unlikely to affect the likelihood of future self-harm behavior, as each individual's self-harm experience is unique:

> *For me, these stories don't really affect what I would be doing because obviously, it is someone else's experience and not mine. So, I feel like it wouldn't impact me as much and what I do in the future.* [Woman; aged 19 years]

Authenticity

The example lived experience stories, which were moderated to avoid being triggering, were reported by some participants as feeling sanitized and somewhat unrealistic. Although some participants suggested that more "uncomfortable" stories with more extreme accounts of crisis would be more likely to prompt them to seek help, others recognized the risk that this could be triggering to some readers:

> *I think these stories are kind of sanitised versions of the actual experience. So I don't think for them there's really anything that could be particularly triggering. I'm kind of in two minds in terms of a more extreme story being helpful. But I think some people it definitely can be because it kind of makes you face up to the potential thing that's going to happen. I also think that could cause different people to spiral a little bit thinking, "oh God, this is going to happen to me," not "this could happen to me." But I think that is very much on an individual basis, I'm not trying to say that story would be bad or that story would be good, it's just it would potentially be a difficult one to kind of balance the benefits with the potential risks for some.* [Man; aged 25 years]

Some aspects of the example stories were felt to increase a sense of authenticity and "humanize" the story without the need for the storyline to be extreme or graphic. First, having realistic representations about recovery being a nonlinear process was considered to be helpful in terms of managing expectations and pressure on participants for their own recovery. The second

example story reflected the likelihood of setbacks or relapses during the process of recovery. This was interpreted by several participants as more authentic and particularly relatable (refer to the *Identifying With the Narrator* section):

> *I liked this one perhaps not so much because of the self-care ideas or things but because it shows sort of the ups and downs of the trajectory and I think that sort of humanizes it a bit and I think that's important. I think accepting that is probably one of the most important things, and accepting that it's not an upwards linear trajectory, it's up, down, down, up, so, I found that one quite helpful because of that.* [Woman; aged 27 years]

Second, acknowledgment of the challenges attached to accessing therapy (in the third example story) was considered by some participants to be helpful in setting realistic expectations. However, others noted that excessive focus on the difficulties around therapy could be discouraging:

> *It's tricky, but from my point of view, it's about trying to find that middle ground between saying "I went to therapy and everything's good now because therapy's amazing" and saying "I've been to 97 therapists and 96 of them were dreadful and it's taken 30 years" or whatever and then you think, "oh well there's no point then." Because obviously everyone's experience is going to be different, so I think it's trying to find a middle ground without being too scripted.* [Woman; aged 23 years]

Third, references to scars (in the third example story) were not reported by any participant to be triggering but served to make a story feel more real and "one of the more lived parts of that lived experience" (man; aged 25 years). The positive framing of scars in this example story was felt by several participants as contributing to a general uplifting and motivating effect. A participant further noted that including a reference to scars could be an important deterrent of self-harm, by reminding the reader about its lasting effects:

> *If anything that's probably like one of the lines. I mean it wouldn't stop me from doing it, but if any of the lines were that would probably be the most likely to stop me from self-harming; to know that there are lasting effects, and actually it does leave scars. Because I think it's hard to remember when you're doing it at the time that you're going to be reminded of it for the rest of your life.* [Woman; aged 24 years]

Finally, in the second example story, the narrator reflected on the importance of their self-harm as a means of coping and as part of their identity. This seemed to particularly appeal to and be respected by participants. Acknowledgment of self-harm as a coping mechanism appeared to reduce feelings of stigma and contributed to increased understanding of participants' own experiences:

> *At the end of the day you've sometimes got to accept that in the relative scheme of things, that might be the healthiest coping mechanism you've got at hand.* [Man; aged 25 years]

Although this was presented positively by some participants as feeling authentic, it is also important to note the potential negative impact this normalization of self-harm may have on help-seeking behavior.

Several participants reflected on the authentic representation of self-harm as an important part of the narrator's personal or social identity in the second example story. Some participants linked this to engaging with lived experience stories on social media and interacting with other members of web-based self-harm discussion boards or groups. Reflections on identity and engagement with web-based groups were felt to be helpful in providing a sense of social support by knowing that there were others going through similar experiences. However, the same participants simultaneously felt this sense of identity could hinder someone's recovery if stopping self-harm was accompanied by detachment from the self-harm community:

> *Yeah, I think it's like we're all part of this kind of secret community that nobody knows about unless you're part of it. And then if something happens and you leave the community then you're not part of it anymore. It feels like [someone else] said about having [self-harm] as part of your identity and this is how I am, then it's not a part of who you are anymore it's almost like you're losing that, but then you're losing that community as well.* [Woman; aged 28 years]

This was felt most acutely when referring to engagement with lived experience stories as user-generated content on social media and web-based forums, which could sometimes feel like an unhealthy, competitive space in which self-harm was normalized:

> *I think although it's reassuring to know there's other people like you, it can also be a bit of a toxic environment because everybody is talking about the same thing and it's almost like normalizing it. And I suppose although it shouldn't be something you're ashamed of, you know, it almost is encouraging this behavior sometimes, I think. And can be – it can kind of lull you into a false sense of security I think in these online spaces particularly. And it can kind of feed that negative mindset of like, okay I'm doing this but if I don't want to stop doing it then this is the place I'll go.* [Woman; aged 28 years]

Language and Stereotyping

Several participants reflected on the importance of lived experience stories using language that does not stigmatize self-harm or alienate people who engage in self-harm. When lived experience stories referred to self-harm as an unhealthy coping mechanism—discussed hypothetically and in reference to the second example story—some participants felt that it reinforced a sense of moralistic judgment, which could feel devaluing. This also applied to the use of the word, "clean," in the second example story, when the narrator described the number of days they had not self-harmed. Although some participants reported feeling ambivalent about this terminology, others highlighted the stigmatizing effect this may have on

people who self-harm. A participant discussed the potential harm this stigma presents in deterring people from seeking support:

> *I didn't think of the comparison to drugs first, but comparison to STD's and in particular HIV and the stigma around that. And I think all three of them kind of interplay in this idea that it's a moral judgement on a medical issue. It isn't helpful to put that moralistic judgement on something like that, because it isn't helpful to people either for feeling able to seek help or finding help in a safe way. And then it doesn't make it as accessible because it isn't as easy to talk about. So I kind of particularly think that [being "clean"] is questionable wording in that respect. I think also by using it to give it that moralistic judgement, people who are going through that are more likely to feel worse in themselves.* [Man; aged 25 years]

A participant highlighted the term, "cutting," in the third example story—but also encountered as "cutter" in other stories found on the web—as "derogatory" and "crass" (woman; aged 28 years). This sort of terminology was also felt to be problematic for being overly specific and thus potentially alienating some readers (refer to the *Identifying With the Narrator* section):

> *I suppose also there's other ways of self-harm and it's not just cutting so that – by saying that it almost invalidates saying "oh it's not self-harm because it's head banging, or it's not self-harm because it's overdoses" or something.* [Woman; aged 40 years]

The example stories and those previously encountered on organizational websites were sometimes felt to perpetuate the stereotype of self-harm as a problem for teenage girls (refer to the *Identifying With the Narrator* section). Several participants suggested that providing a wide range of stories on organizational websites, including some from older narrators, would be beneficial:

> *They're quite young in all the stories. I think it would be nice every now and then to see a story of like somebody a bit older. Because I feel like a lot of self-harm is sort of people assume it's always like teenagers who are depressed or something. But it would be nice if there was more like 20, 30-year-old stories. So, you know that actually it's not just depressed teenagers that do it.* [Woman; aged 24 years]

Discussion

Principal Findings

People with recent self-harm experience reported positive engagement with lived experience stories encountered on organizational websites and sometimes on social media. Example stories used to evaluate different features of this type of web-based content were generally found to be most engaging and helpful if the feelings recounted by the narrator were relatable, practical help seeking and self-help strategies were included, the stories felt human and authentic, and the language used did not perpetuate stigma or stereotypes about self-harm. Our findings suggest that lived experience stories can support readers to feel validated and help them to better understand their own feelings and experiences. Participants also told us that exposure to these stories can empower people to share their experiences with professionals and their support network and that they may also provide a helpful way of educating others about self-harm, which can reduce stigma and self-stigma [18].

The narratives of recovery were generally felt to be inspiring and motivating to those thinking about their own recovery from self-harm, and lived experience stories appeared to be highly valued as a form of web-based support, which could also serve as a temporary distraction for those experiencing an immediate urge to self-harm. However, our findings also highlight that voicing such narratives from a young or emerging adult's perspective could have a demoralizing effect on older readers. Strategies to mitigate this problem may include (1) maintaining ambiguity regarding the age of the narrator to minimize exclusion of some readers; (2) including more stories from older narrators to broaden the representation of different age groups; or (3) tailoring age references depending on the target audience, for example, using young narrators on youth mental health websites.

Our findings also highlight the need for a balance between providing detail and specificity to make stories feel relatable to readers and avoiding the exclusion or invalidation of readers who do not identify with such details. Previous studies suggest that high specificity in stories about self-harm or suicide—particularly regarding methods—can risk overidentification with the individual or narrator, thus increasing the risk of contagion [7,19,20]. This was not captured in our findings as we deliberately excluded detailed references to methods of harm for ethical reasons. A balance should also be sought between the avoidance of narratives that may be triggering for some readers and the use of stories that could be overly formulaic and sanitized. In our sample, including a positive reference to scars in an example story was not felt to be triggering but contributed a sense of the lived experience story being more raw and real. Authenticity was felt to be further enhanced through portraying recovery journeys as nonlinear and acknowledging relapses and thus managing expectations regarding one's own recovery. Although care must be taken to avoid normalization of self-harm, acknowledgment of self-harm as a key coping mechanism or part of someone's identity may also feel less sanitized to some readers, thus increasing their engagement with a story.

Our findings highlight lived experience stories as a setting for the reader to compare themselves with the narrator. This can be helpful in managing feelings of isolation, knowing that others are experiencing similar feelings, and engaging in self-harm as a coping mechanism. It can also be positive when viewing stories of recovery as motivating. Feeling as if one is in a position comparable with that of the narrator may inspire help seeking by recognizing need in others and oneself by extension and by instilling hope that recovery can be attained. However, not all comparisons are positive. We found that the helpfulness of comparisons between reader and narrator may depend on

their relative age and the reader's stage of recovery. Having few negative experiences or those perceived to be less traumatic compared with those of the narrator can feel invalidating and may deter some from seeking help as a result. Comparing oneself against a narrator who is more advanced in their recovery may instill feelings of dejection and pessimism in the reader. Providing numeric metrics, such as days without self-harm, can make comparisons particularly salient. Although comparing with others going through similar experiences can reduce feelings of isolation and encourage a sense of community, web-based forums can also foster a sense of toxic competition and normalization of self-harm.

Feelings of connectedness, enhanced understanding of one's own emotions and behavior, use of alternative coping strategies for emotion regulation, and self-disclosure of self-harm to others have been described elsewhere as important factors in recovery from self-harm [21]. Thus, our findings suggest that engagement with web-based lived experience stories may provide a worthy contribution to the self-harm recovery process.

Limitations

It is important to recognize that the example stories presented to participants in this study were selected by the research team to be recovery oriented and consistent with safety guidelines regarding web content and thus excluded graphic content. This is reflected in some of the comments about stories feeling *sanitized* (refer to the *Authenticity* section). Although these stories were felt to be broadly representative of content published by third-sector organizations, they may not reflect unmoderated, user-generated content shared on social media. Participants' accounts of encountering uncensored or less-helpful stories were limited. Therefore, our findings highlight several helpful aspects of lived experience stories, but we are limited in the extent to which we can draw conclusions about the potentially harmful aspects. This is particularly relevant to stories shared on social media, which may include more graphic content.

Similarly, although our focus was on specific features of the example stories and those encountered on organizational websites, there was limited discussion about user-generated content encountered on the web. Therefore, we have not explored the potential impact of purposeful reading of a lived experience story—one that is expected and consented to—compared with unexpectedly encountering a story on social media by its appearance on a feed or by algorithmic promotion.

We purposely limited the numbers in each focus group owing to the sensitive nature of the topic; however, group 1 was smaller than intended. Although the 2 participants in group 1 appeared to be comfortable with reflecting on personal and upsetting experiences, the discussion flowed more freely in the other groups, each with 3 or 4 participants. Participants in the slightly larger groups were better able to respond to one another, and concurrence and disagreement among group members resulted in particularly rich data.

Our sample was not diverse in terms of gender, with boys and men being particularly underrepresented. This is a common challenge in self-harm studies [5], but our additional attempts to purposively sample this group were unsuccessful. Information about participants' ethnicity, sexual orientation, socioeconomic status, geographical location, or whether they were transgender individuals was not collected, making it impossible to evaluate the diversity of the sample. However, a broad age range was represented in our sample, thus providing valuable insight into the age-related factors discussed previously. In addition, many participants described being at different stages of a recovery journey, thus contributing useful insight into differences in acceptance, reactions to, and engagement with lived experience stories, depending on the current circumstances or mindset of the reader.

Comparison With Previous Studies

There has been little previous research on engagement with lived experience stories related to self-harm. Interviews with young people and GPs in the United Kingdom about web-based support services for suicide prevention indicated the overall value placed on lived experience content [9]. Consistent with our findings, young people reported the feeling of reassurance that they were not alone in their experiences, recognition of the need for support in others and themselves in turn, and concerns around content being potentially triggering to vulnerable readers. The potential for negative comparison was also raised by GPs, which we highlight as a particular risk for adults who self-harm beyond adolescence. We further emphasize that the inclusion of numeric metrics (such as age and days free from self-harm) in a story may increase the likelihood of comparison.

People who engage in web-based help seeking for self-harm have described stumbling across graphic social media content or comments that reinforce self-harm and suicide [6,19]. Our findings suggest several positive aspects of lived experience stories similar to those provided on organizational websites (such as Mind or Samaritans). Although we did not focus in depth on graphic content within lived experience stories on social media, our findings indicate that there may be risks of competition attached to engagement with user-generated content, which is often accompanied by unmoderated interactions [5]. Such feelings of competitiveness have previously been identified as a negative outcome of (and motivation for) engagement in web-based self-harm forums [6,7].

Our findings align with the framework for impact and moderators of impact developed by Rennick-Egglestone et al [22] in a systematic review of recovery narratives related to eating disorder and psychosis. By synthesizing 4 qualitative studies on the impact of recovery narratives (the key thread of most lived experience stories), the authors identified 6 themes, all of which were reinforced by our own findings specific to self-harm. The themes identified in the review by Rennick-Egglestone et al [22] were *connectedness to the narrator or to others*; *understanding of mental illness and how to recover*; *reduction of stigma, including self-stigma*; *validation of self and of personal experiences*; *affective responses*; and *behavioral responses*. Several moderators of impact were also identified regarding reader and narrator characteristics and are reflected in our findings; however, contextual characteristics (such as internet access) were not relevant in our study. Consistent with our findings, positive reception of a story was

facilitated by the reader having similar life experiences to the narrator, whereas barriers included inability to relate to the narrator's recovery journey or feeling worse than the narrator. As evidenced in our findings, Rennick-Egglestone et al [22] suggest the perceived authenticity of the narrator to be a key moderator of the impact of a narrative. This review was part of an in-depth body of work by the Narrative Experiences Online team to evaluate the impact of recovery narratives on adults who experience psychosis or other mental health problems [23]. In addition, within the Narrative Experiences Online program, Ng et al [20] have developed a causal chain model, using data from interviews with adult users of UK mental health services, to demonstrate the key mechanisms in whether a reader connects with a recovery narrative. These were (1) comparison with the narrator, in terms of shared experiences, stage of recovery, and narrator characteristics; (2) learning, in terms of gaining insight into the perspectives of others and developing new coping techniques; and (3) empathy, in terms of insight into the struggles, successes, and tone of the narrator. Our findings provide further support for these mechanisms of connection, extended to the self-harm context. We additionally highlight empowerment through enhanced understanding of one's own experience as a key part of the *learning* mechanism and both avoidance of stigmatizing language and authentic representation of a self-harm recovery journey as key parts of *empathy*.

However, in contrast to the findings by Ng et al [20], and perhaps specific to the self-harm context, we found that lived experience stories were felt to be a useful distraction and source of support during moments of crisis for people who self-harm. Participants in our study also interpreted nonlinear narratives of recovery as more authentic and easy to connect with, whereas disjointed or circular narratives were interpreted pessimistically by Ng et al [20]. Nonlinear narratives of recovery from self-harm may be useful if they project an overall upward trajectory.

As mentioned in the *Limitations* section, the example stories presented to participants were carefully selected and were more representative of stories published on organizational websites than on social media, in terms of more *sanitized* content and a clear narrative structure. Although our findings point to several suggestions for helpful aspects to be included in a lived experience story, people posting a story on the web or those moderating web-based forums should also be mindful of existing guidelines around safe communication about suicide and self-harm on social media [6,24,25]. Broadly consistent with our findings (refer to the theme, *Language and Stereotyping*), #chatsafe guidelines recommend removing unhelpful language or graphic references to self-harm and using trigger warnings if appropriate [24].

Conclusions

This study highlights the value placed on lived experience stories as an accessible form of web-based support and information for those who self-harm. Our findings have several implications for health care or third-sector organizations publishing this type of content and for community guidelines for web-based forums in which users share their stories. First, it is important to consider the narrator's age when describing a successful recovery. Second, ensuring that stories feel human and relatable by including references to challenges and relapses in a recovery journey will be valuable. Third, care should be taken to not perpetuate stereotypes of self-harm as a problem for teenage girls, and stigmatizing language or implied moral judgments should be avoided. We urge caution to those publishing or moderating lived experience stories, in terms of removing unhelpful language or graphic references to self-harm, using trigger warnings if appropriate, and being mindful of the risk of negative comparisons or competition.

Lived experience stories—shared carefully—have the potential to inspire those who self-harm to seek help, empower people to understand and share their own experiences, distract those experiencing an urge to self-harm in the moment, and reduce feelings of self-stigma and isolation.

Acknowledgments

This study is funded by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at the University Hospitals Bristol and Weston National Health Service Foundation Trust and the University of Bristol. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. BM is funded by a Medical Research Foundation fellowship (MRF-058-0017-F-MARS-C0869). PM is supported by the NIHR Applied Research Collaboration West and the NIHR Biomedical Research Centre at the University Hospitals Bristol and Weston National Health Service Foundation Trust and the University of Bristol. The funders had no involvement in the study design; collection, analysis, and interpretation of the data; writing of the report; or decision to submit the paper for publication. The authors are extremely grateful to the people who participated in this study. The authors would also like to thank Nina di Cara and Zoe Haime for their support in moderating the focus groups.

Authors' Contributions

BM, JF, PM, IPV, LG, and LB contributed to conception and design of the study. LW conducted the study, with support from JF and BM. LW and LB collaboratively coded a subsection of the data, and LW conducted the remaining analysis and drafted the output. LW, BM, JF, and LB contributed to interpretation of data. BM, JF, PM, IPV, LG, and LB critiqued the output for important intellectual content. All authors have read and approved the final version of the manuscript. LW serves as guarantor for the contents of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The 3 example lived experience stories presented to focus group participants for discussion.
[DOCX File , 18 KB - mental_v10i1e43840_app1.docx]

References

1. McManus S, Gunnell D, Cooper C, Bebbington PE, Howard LM, Brugha T, et al. Prevalence of non-suicidal self-harm and service contact in England, 2000-14: repeated cross-sectional surveys of the general population. Lancet Psychiatry 2019 Jul;6(7):573-581 [FREE Full text] [doi: 10.1016/S2215-0366(19)30188-9] [Medline: 31175059]
2. Mars B, Heron J, Crane C, Hawton K, Lewis G, Macleod J, et al. Clinical and social outcomes of adolescent self harm: population based birth cohort study. BMJ 2014 Oct 21;349:g5954 [FREE Full text] [doi: 10.1136/bmj.g5954] [Medline: 25335825]
3. Brennan C, Saraiva S, Mitchell E, Melia R, Campbell L, King N, et al. Self-harm and suicidal content online, harmful or helpful? A systematic review of the recent evidence. J Public Mental Health 2022 Jan 05;21(1):57-69. [doi: 10.1108/jpmh-09-2021-0118]
4. Lavis A, Winter R. #Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media. J Child Psychol Psychiatry 2020 Aug;61(8):842-854. [doi: 10.1111/jcpp.13245] [Medline: 32459004]
5. Biddle L, Derges J, Goldsmith C, Donovan JL, Gunnell D. Online help for people with suicidal thoughts provided by charities and healthcare organisations: a qualitative study of users' perceptions. Soc Psychiatry Psychiatr Epidemiol 2020 Sep;55(9):1157-1166 [FREE Full text] [doi: 10.1007/s00127-020-01852-6] [Medline: 32157324]
6. Abou Seif N, John-Baptiste Bastien R, Wang B, Davies J, Isaken M, Ball E, et al. Effectiveness, acceptability and potential harms of peer support for self-harm in non-clinical settings: systematic review. BJPsych Open 2022 Jan 17;8(1):e28 [FREE Full text] [doi: 10.1192/bjo.2021.1081] [Medline: 35034666]
7. Murray CD, Fox J. Do internet self-harm discussion groups alleviate or exacerbate self-harming behaviour? Australian e J Advancement Mental Health 2014 Dec 17;5(3):225-233. [doi: 10.5172/jamh.5.3.225]
8. Rowe SL, French RS, Henderson C, Ougrin D, Slade M, Moran P. Help-seeking behaviour and adolescent self-harm: a systematic review. Aust N Z J Psychiatry 2014 Dec;48(12):1083-1095. [doi: 10.1177/0004867414555718] [Medline: 25335872]
9. Cohen R, Rifkin-Zybutz R, Moran P, Biddle L. Web-based support services to help prevent suicide in young people and students: a mixed-methods, user-informed review of characteristics and effective elements. Health Soc Care Community 2022 Nov;30(6):2404-2413. [doi: 10.1111/hsc.13819] [Medline: 35514202]
10. Mars B, Heron J, Biddle L, Donovan JL, Holley R, Piper M, et al. Exposure to, and searching for, information about suicide and self-harm on the Internet: prevalence and predictors in a population based cohort of young adults. J Affect Disord 2015 Oct 01;185:239-245 [FREE Full text] [doi: 10.1016/j.jad.2015.06.001] [Medline: 26150198]
11. Rodway C, Tham SG, Richards N, Ibrahim S, Turnbull P, Kapur N, et al. Online harms? Suicide-related online experience: a UK-wide case series study of young people who die by suicide. Psychol Med 2022 May 19:1-12. [doi: 10.1017/S0033291722001258] [Medline: 35587034]
12. Rennick-Egglestone S, Ramsay A, McGranahan R, Llewellyn-Beardsley J, Hui A, Pollock K, et al. The impact of mental health recovery narratives on recipients experiencing mental health problems: qualitative analysis and change model. PLoS One 2019;14(12):e0226201 [FREE Full text] [doi: 10.1371/journal.pone.0226201] [Medline: 31834902]
13. Kosyluk K, Marshall J, Conner K, Macias DR, Macias S, Michelle Beekman B, et al. Challenging the stigma of mental illness through creative storytelling: a randomized controlled trial of this is my brave. Community Ment Health J 2021 Jan;57(1):144-152. [doi: 10.1007/s10597-020-00625-4] [Medline: 32388690]
14. De Vecchi N, Kenny A, Dickson-Swift V, Kidd S. How digital storytelling is used in mental health: a scoping review. Int J Ment Health Nurs 2016 Jun 21;25(3):183-193. [doi: 10.1111/inm.12206] [Medline: 26900000]
15. Investigating the impact of lived experience stories on self harm, mood, and help-seeking intentions. OSF. 2022 Mar 23. URL: https://osf.io/cesda [accessed 2022-12-01]
16. Winstone L, Mars B, Ferrar J, Moran P, Penton-Voak I, Biddle L. How do people who self-harm engage with, interpret, and evaluate lived experience stories? OSF. 2022 Nov 1. URL: https://osf.io/sdm6h/ [accessed 2022-11-01]
17. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology, Vol. 2. Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. Washington, D.C., United States: American Psychological Association; 2012:57-71.
18. Lindstrom G, Sofija E, Riley T. "Getting better at getting better": how sharing mental health stories can shape young people's wellbeing. Community Ment Health J 2021 Nov;57(8):1604-1613. [doi: 10.1007/s10597-021-00786-w] [Medline: 33580469]

19. Biddle L, Derges J, Goldsmith C, Donovan JL, Gunnell D. Using the internet for suicide-related purposes: contrasting findings from young people in the community and self-harm patients admitted to hospital. PLoS One 2018;13(5):e0197712 [FREE Full text] [doi: 10.1371/journal.pone.0197712] [Medline: 29795637]
20. Ng F, Charles A, Pollock K, Rennick-Egglestone S, Cuijpers P, Gillard S, et al. The mechanisms and processes of connection: developing a causal chain model capturing impacts of receiving recorded mental health recovery narratives. BMC Psychiatry 2019 Dec 21;19(1):413 [FREE Full text] [doi: 10.1186/s12888-019-2405-z] [Medline: 31864315]
21. Ryan-Vig S, Gavin J, Rodham K. The presentation of self-harm recovery: a thematic analysis of YouTube videos. Deviant Behav 2019 Apr 01;40(12):1596-1608. [doi: 10.1080/01639625.2019.1599141]
22. Rennick-Egglestone S, Morgan K, Llewellyn-Beardsley J, Ramsay A, McGranahan R, Gillard S, et al. Mental health recovery narratives and their impact on recipients: systematic review and narrative synthesis. Can J Psychiatry 2019 Oct;64(10):669-679 [FREE Full text] [doi: 10.1177/0706743719846108] [Medline: 31046432]
23. The NEON Collection. Research into Recovery. URL: https://www.researchintorecovery.com/neoncollection [accessed 2022-11-01]
24. Robinson J, Hill NT, Thorn P, Battersby R, Teh Z, Reavley NJ, et al. The #chatsafe project. Developing guidelines to help young people communicate safely about suicide on social media: a Delphi study. PLoS One 2018;13(11):e0206584 [FREE Full text] [doi: 10.1371/journal.pone.0206584] [Medline: 30439958]
25. Thorn P, Hill NT, Lamblin M, Teh Z, Battersby-Coulter R, Rice S, et al. Developing a suicide prevention social media campaign with young people (The #Chatsafe Project): co-design approach. JMIR Ment Health 2020 May 11;7(5):e17520 [FREE Full text] [doi: 10.2196/17520] [Medline: 32391800]

Abbreviations

GP: general practitioner

Edited by J Torous; submitted 10.11.22; peer-reviewed by S Rowe, C Brennan; comments to author 16.12.22; revised version received 05.01.23; accepted 05.01.23; published 31.01.23.

Please cite as:
Winstone L, Mars B, Ferrar J, Moran P, Penton-Voak I, Grace L, Biddle L
Investigating How People Who Self-harm Evaluate Web-Based Lived Experience Stories: Focus Group Study
JMIR Ment Health 2023;10:e43840
URL: https://mental.jmir.org/2023/1/e43840
doi:10.2196/43840
PMID:36719729

Original Paper

A Digital Single-Session Intervention Platform for Youth Mental Health: Cultural Adaptation, Evaluation, and Dissemination

Akash Shroff[1], BS; Chantelle Roulston[1], BA; Julia Fassler[1], BA; Nicole A Dierschke[2], MPH, DrPH; Jennifer San Pedro Todd[2], RN, JD; Ámbar Ríos-Herrera[2], MS, LPC; Kristen A Plastino[2], MD; Jessica Lee Schleider[1], PhD

[1]Department of Psychology, Stony Brook University, Stony Brook, NY, United States

[2]University of Texas Teen Health, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

Corresponding Author:
Jessica Lee Schleider, PhD
Department of Psychology
Stony Brook University
Psychology B-340
Stony Brook, NY, 17794
United States
Phone: 1 9174391872
Email: jessica.schleider@stonybrook.edu

Abstract

Background: Despite the proliferation of evidence-based digital mental health programs for young people, their low uptake and inconsistent implementation preclude them from benefiting youths at scale. Identifying effective implementation strategies for evidence-based supports is especially critical in regions where treatment access is lowest owing to mental health provider shortages.

Objective: The goal of this academic-community partnership, funded by the City of San Antonio Metropolitan Health District, was to culturally adapt, disseminate, and gauge the acceptability and utility of an evidence-based digital mental health platform—Project Youth Empowerment and Support (YES)—among English- and Spanish-speaking youths living in south Texas.

Methods: Project YES is an open-access, anonymous platform containing 3 evidence-based, self-guided interventions for youth mental health. Project YES was culturally adapted via focus groups and co-design sessions with San Antonio youths with lived experience of depression and anxiety; translated into Spanish; and disseminated throughout San Antonio, Texas, via community and school partnerships.

Results: During the project period (April 2021 to December 2021), 1801 San Antonio youths began and 894 (49.64%) of them completed a 30-minute, single-session intervention within Project YES (aged 11-17 years; n=718, 39.87% male; n=961, 53.36% female; and n=3, 0.17% intersex; n=1477, 82.01% Hispanic; n=77, 4.28% non-Hispanic White; n=113, 6.27% Black; n=28, 1.55% Asian; and n=93, 5.16% other). This completion rate (49.64%) surpassed those previously observed for Project YES (eg, 34% when disseminated via social media). San Antonio youths rated Project YES as highly acceptable across all metrics, both in English and Spanish. In addition, the youths who completed Project YES—ENGLISH reported significant improvements in hopelessness (Cohen d=0.33; P<001), self-hate (Cohen d=0.27; P<001), and perceived agency (Cohen d=0.25; P<001) from before to after the intervention, and the youths who completed Project YES—SPANISH reported significant improvements in self-hate (Cohen d=0.37; P=.049) from before to after the intervention.

Conclusions: The results indicate that Project YES—an open-access, free, and anonymous web-based single-session intervention platform—is an acceptable, accessible, and applicable mental health support for English- and Spanish-speaking San Antonio youths.

(JMIR Ment Health 2023;10:e43062) doi:10.2196/43062

KEYWORDS

single-session intervention; cultural adaptation; web-based intervention; adolescents; mobile phone

Introduction

Background

Most adolescents with mental health needs in the United States are unable to access evidence-based interventions when and where they are needed [1]. This need-to-access gap has remained stagnant for decades because of a myriad of factors, including provider shortages nationwide; a dearth of therapists trained in evidence-based approaches; premature dropouts from multisession treatments; cost and transportation challenges; and language barriers, as many evidence-based treatments are accessible only to English-speakers [2-4]. All these access barriers grew starker with the onset of the COVID-19 pandemic when public health mandates and social distancing precautions further limited access to face-to-face treatment options [5]. Given these compounding factors, interest has dramatically grown in digital, self-guided mental health supports (ie, those accessible via computers and smartphones, which are accessible to >95% of the US adolescents) [6], which hold the potential to be more accessible and scalable than in-person interventions. However, the real-world uptake and completion rates of self-guided digital interventions are exceptionally low, on average [7]; many clinically tested digital self-help tools are never made openly accessible [8], few evidence-based self-help tools are available in languages other than English, and popular mental health apps often include no elements of evidence-based interventions [9]. Thus, there is a pressing need for digital youth mental health supports that are simultaneously evidence-based, easily accessible to diverse youths in moments of need, and optimized for uptake and completion in community settings.

Digital, evidence-based single-session interventions (SSIs) may help address the need for accessible, easily completable, and effective mental health interventions for adolescents. SSIs are defined as "structured programs that intentionally involve only one visit or encounter with a clinic, provider, or program; they may serve as stand-alone or adjunctive clinical services" [10]. SSIs can take multiple forms (eg, provider delivered and self-guided); however, the flexible and scalable nature of *digital, self-administered* SSIs allows them to be delivered and completed in low-resource settings relatively rapidly and broadly. Existing literature supports the effectiveness of SSIs for adolescents' depression and anxiety, even when delivered as stand-alone mental health supports [11-14]. Furthermore, in a recent nationwide randomized trial including 2452 adolescents with elevated depressive symptoms, 2 different digital, self-guided SSIs (one teaching behavioral activation and another teaching that personal characteristics are malleable) significantly reduced 3-month depressive symptoms, hopelessness, and restrictive eating compared with a supportive control [15]. In another randomized, placebo-controlled trial of 555 adolescents endorsing self-injurious behaviors, a digital, self-guided SSI significantly reduced self-hatred and increased the desire to stop future nonsuicidal self-injury [16]. In both trials, adolescents completed the digital SSIs in the midst of the COVID-19 pandemic (between May and December 2020), and more than 80% of the adolescents who began a digital SSI completed it in full. Even when evaluated in naturalistic settings, where the completion rates for digital self-help tools drop as low as 1% (vs 44%-99% in paid research studies) [7], 34% of US adolescents completed digital mental health–focused SSIs that were publicly and freely accessible on the web [17]. When comparing the engagement rates of digital SSIs in randomized controlled trials versus naturalistic evaluations, the presence of participant payments emerged as the sole predictor of higher versus lower rates of SSI completion (neither the level of distress nor sexual orientation, race, ethnicity, and gender identity predicted differential completion rates) [17]. Nonetheless, even in the absence of monetary incentives, SSIs show feasibility and clinical utility. The completion rates for digital SSIs substantially surpass those seen for previously studied digital self-help tools in formal trials and naturalistic contexts alike.

Given their clinical utility, high completion rates, and low cost (the abovementioned SSIs are free and available on the web) and because youths increasingly turn to digital tools such as websites and apps for mental health information and support [18], SSIs represent a promising approach to supporting youths' in-the-moment mental health needs. However, many questions remain regarding the dissemination and implementation of the existing SSIs within underserved communities. To date, no evidence-based SSIs have been systematically adapted to the culture, language, or context of specific underresourced and underserved populations—although such adaptations may both facilitate successful implementation and minimize the unintended consequences of failing to center population-specific needs (ie, inadvertently worsening treatment access disparities by disseminating English-only digital self-help tools in communities with many monolingual Spanish speakers). Cultural adaptations of interventions have demonstrated increased acceptability and retention in marginalized populations by including their stories and translations by talking to clinicians well versed in the cultural differences of specific populations [19,20]. However, startlingly, only a few youth-focused intervention trials—even those that include youths of color—actually use strategies to address culture within the interventions themselves [21]. Furthermore, very few culturally adapted interventions for adolescent anxiety and depression have been developed; those that exist rely mainly on clinician feedback, instead of youth involvement, to drive cultural adaptations [21]. This reflects an important and critical gap in the available digital supports.

Culturally adapted digital tools may be especially useful for low-resourced or minoritized communities because individuals belonging to these communities often rely on digital resources for health-related information [22]. A digital intervention app created for Spanish-speaking adults showed greater retention rates (72%) than most digital mental health apps. A pilot study testing this app demonstrated a preliminary reduction in depression severity and greater acceptability in the Spanish-speaking study population [23]. However, the use of digital interventions in these minoritized populations *without* careful adaptation can potentially have unintended negative consequences, such as alienating users with marginalized identities [22], making it crucial to rely on user and stakeholder feedback throughout the adaptation process. Overall, few evidence-based mental health interventions have been systematically adapted for specific minoritized or marginalized

youth communities, and even when those adaptations *have* been deployed, youth feedback and perspectives on the interventions have rarely been formally incorporated (eg, via user-centered design approaches) [24]. Thus, efforts to culturally adapt youth-focused SSIs will need to proactively center on end users' perspectives and experiences.

Latino, Latina, or Latine young people are a population with substantial and unmet mental health needs for whom culturally tailored digital mental health supports (including SSIs) may carry value. Latino, Latina, or Latine individuals are the largest ethnic minority in the United States, accounting for 18.5% of the US population [23]. Up to 13 million Latine individuals meet the criteria for a mental illness; however only 9.6% of these individuals access any evidence-based mental health support in a given year [24]. Barriers to treatment access vary widely, from limited insurance coverage to legal status, stigma, and language barriers (up to 33% Latine individuals report speaking English less than "very well" [25]); together, these create a clear need for scalable mental health tools tailored to the needs of Latine and Spanish-speaking individuals. Not only do most Latine and Spanish-speaking individuals in the United States have internet access via mobile devices (up to 80%), but a large majority also endorse openness to using digital health supports [26]. In addition, there is some evidence that Spanish-translated digital mental health tools are effective (3 studies investigating Spanish-language apps reported decreases in the target mental health scores such as depression or substance abuse [27]). However, few such tools have been built and evaluated to date [27]. Accordingly, the creation of translated and culturally relevant digital mental health supports is a critical next step for digital and SSI research.

Importantly, both Spanish- and English-speaking communities across different regions of the United States are highly diverse in themselves, and as such, taking a *community*- or *region-level* approach to adapting digital mental health tools may yield supports that are particularly acceptable to (and likely to be used by) specific communities. A community that may benefit from such efforts is that of Texas, one of the largest states in the United States, which is ranked by Mental Health America as among America's worst-performing states when it comes to treatment access. For instance, among the youths in Texas with a major depressive episode last year, >73% did not access any form of mental health treatment—this was the worst percentage among all the states last year [28]. Specifically, San Antonio, Texas, and the surrounding area are designated as a high-need, low-income mental health professional shortage area by the Health Resources and Services Administration [29]. In addition, of all the people in San Antonio, approximately 57% speak only English and 39% speak only Spanish [30]. Therefore, accessible, culturally tailored, digital, and brief mental health supports are critically needed to help youths in San Antonio, Texas, manage their mental health needs flexibly, affordably, and in real-time moments of need. Adapting and evaluating digital, evidence-based SSIs for youths in this city promises to increase the adolescents' odds of accessing *any* form of evidence-based mental health support, given that many San Antonio youths with mental health needs are currently accessing no support at all.

Accordingly, we led an academic-community partnership project to systematically adapt, translate, and disseminate digital SSIs for use by youths in the high-need city of San Antonio, Texas. In collaboration with youth stakeholders and local health care providers, we adapted an evidence-based single-session web-based intervention platform [31] and tested its acceptability, feasibility, and utility when disseminated as a city-wide resource to English- and Spanish-speaking young people in San Antonio. Our approach to cultural adaptation fits with the core principles of Heim and Kohrt's [32] Cultural Adaptation Framework for Scalable Interventions, which centers on the integration of the *cultural concepts of distress* (here, reflected within San Antonio youths' lived experience narratives and descriptions of their own difficulties and coping strategies), along with adaptations to nonspecific intervention factors to increase cultural relevance (eg, the presence of a Spanish-translated version, youth feedback–informed language adaptations, and the inclusion of San Antonio youth voices throughout all programs). The specific platform adapted for this nonexperimental, observational project is called Project Youth Empowerment and Support (YES), which is a free-of-charge, open, and anonymous website wherein young people can anonymously and flexibly complete any of 3 digital SSIs, which have demonstrated short- and long-term effectiveness in reducing hopelessness, self-hate, and depression symptoms [12,15,31]. The 3 SSIs that adolescents may complete within Project YES are called "Project Personality" (which teaches that personal traits are malleable), "The ABC Project" (which teaches values-based activity engagement to elicit pleasure and accomplishment), and "Project CARE" (which teaches the benefits of self-kindness in social and academic success). These 30-minute SSIs were designed to instill adaptive self-relevant beliefs with known links to lower levels of depression and anxiety symptoms. The SSIs within Project YES have shown acceptability and both short- and long-term utility in reducing hopelessness, increasing agency, and mitigating depression and anxiety symptoms in young people across randomized and open trials alike [12,15,17,33]. Each SSI was created per a routinely used SSI design framework, which is detailed elsewhere [10]. Youths (1) learn the brain science that normalizes a core concept, (2) are invited to help researchers learn about their perspectives as "youth experts," (3) are asked to convey the program's messages in their own words and offer advice to their peers, and (4) hear stories from peers who used the program in their lives. Regardless of program selection, youths can choose to offer their "best, anonymous coping advice" to other youths coping with depression or anxiety. Youths are also given the opportunity to share this advice in a public "YES advice center" [10].

Goal of This Study

In this project, the 3 digital SSIs in Project YES were adapted and revised in collaboration with youth stakeholders, translated into Spanish, and disseminated to youths across San Antonio, Texas, via collaborations with community organizations, clinics, and schools in the city. For both the English and Spanish versions of the interventions, we investigated youths' acceptability of the programs and immediate effects on proximal, clinically relevant outcomes that the SSIs were designed to target (hopelessness, agency, perceived control, and self-hate).

We examined use-pattern variables (eg, SSI selection and SSI completion rates), user characteristics (age range, sex, gender identity, race and ethnicity, and depressive symptoms), and acceptability metrics to gauge which youths use Project YES and whether they view Project YES as valuable, helpful, and user-friendly.

Methods

Recruitment

The participants in this project were youths aged 11 through 17 years from San Antonio, Texas, and the surrounding areas (within a 100-mile [161 km] radius) who interacted with the Project YES platform between April 31, 2021, and December 31, 2021. The participants learned about Project YES through several sources, including paid advertisements on Instagram (Meta Platforms, Inc), friends, teachers, and the University of Texas Teen Health (UTTH) team. UTTH provided 12 opportunities for representatives from potential community partners (eg, schools and community centers) and community members to receive an orientation to Project YES, which covered the following topics: (1) overview of Project YES, (2) introduction to Project YES in San Antonio, (3) goals for Project YES San Antonio, (4) what youths will do in Project YES, (5) community partner roles and responsibilities, and (6) additional resources. The purpose of the orientation was to provide the community with an understanding of Project YES and resources to refer youths to the program. A total of 79 people attended the orientation. From April 1, 2021, to June 30, 2021, of those who attended the orientation, 2 (3%) school partners and 4 (5%) community organizations reported referring youths to Project YES. The partners made 2763 referrals during this reporting period.

To facilitate youth-level recruitment efforts, UTTH created advertising postcards with the QR code to access the website, along with phone numbers, on the back of the postcard for youths who may be in crisis and need immediate assistance. These postcards were distributed to youths in schools, community organizations, counseling offices, and clinics and anyone interested in the program. During the project year, a total of 12,839 postcards were distributed. UTTH also created flyers, with the assistance of the Stony Brook University team, which were posted on the Lab for Scalable Health's Instagram page. These were also posted on UTTH's social media sites and the UTTH program website and added to UTTH staff's email signature blocks.

To facilitate community partners' efforts to increase Project YES's uptake, UTTH created a social media toolkit to aid partners in sharing Project YES with the youths in their organization. This comprehensive guide has information on the need for the project, the evidence base, information on how to engage with Project YES, sample social media posts, a sample script for introducing Project YES to young people, and frequently asked questions. When possible, UTTH also attended health fairs or school functions to inform attendees about Project YES, with an express focus on the San Antonio area.

Facilitating the Uptake of Project YES in Schools

Owing to COVID-19–related public health restrictions at the beginning of the 2021 school year, visitors were not allowed on campus for Project YES–related presentations. However, in October 2021, UTTH received the opportunity to present Project YES during a physical education class to middle school– and high school–aged students at a local school district. UTTH coordinated with district-based physical education coaches to offer Project YES to the students on the campus. Each coach reserved the computer laboratory or gymnasium as per availability. The coaches asked the students to bring their school laptops to the session. During these sessions, the students were informed about the program and logged in using the QR code for their phone or a computer desktop or laptop. There were some challenges with poor internet connectivity, specifically in the gymnasium, which prevented some students from completing an SSI within Project YES. Subsequently, this process was replicated with a second school district partnering with UTTH, again targeting the district's middle school– and high school–aged students.

Ethical Considerations

Before launching Project YES, all procedures were reviewed and deemed "exempt" (as a program evaluation) by the UT Health San Antonio's institutional review board (IRB).

Project Procedures

Overview

There are no inclusion or exclusion criteria for Project YES, as it is publicly accessible, and the data collected are nonidentifiable. Although the Project YES website states that the activities are designed for youths aged >18 years (primarily preadolescent and adolescent youths), individuals of any age may choose to participate. The participants are required to report their age range (11 to 13, 14 to 16, or 17 years) and whether they are aged ≥18 years (yes or no) before initiating Project YES. The participants are required to report whether they were from San Antonio (or within 100 miles; yes or no). Using these data, we limited the analyses to our target population (youths aged 11 to 17 years from San Antonio).

The structure of Project YES has been described in a previous manuscript [10]. To summarize, Project YES is a nonrandomized, ongoing, and exploratory program evaluation whose original procedures were preregistered on Open Science Framework [34]. Parent permission is not required to participate in Project YES (waived by the University's IRB) to minimize access barriers, including discomfort disclosing psychological distress (as many parents are not aware of their children's depressive symptoms), and to ensure that youths remain unidentifiable.

Once the participants begin Project YES, advance past the "Project Information" page, and agree to participate, they are prompted to provide nonidentifying demographic information. Next, the participants select 1 of 3 SSIs to complete before completing pre-SSI questionnaires, the SSI itself, and post-SSI questionnaires, which were designed to measure the SSI's short-term effects and acceptability. Finally, the participants

have the option to anonymously share their "best advice for others dealing with depression, anxiety, or stress," which is posted on the "YES Advice Center" for others to read.

The current version of Project YES was adapted in several ways to better serve the San Antonio community. Specifically, Project YES was translated entirely into Spanish, including the pre- and postintervention questionnaires and the 3 digital SSIs. Before participating in the SSIs, the participants are given the option to select the language (either English or Spanish) in which they would like to hear or view the SSI. In addition, the question "Do you live in San Antonio, TX, or the surrounding areas (ie, within 100 miles, like Atascosa, Medina, Bandera)?" was added. Using these data, we limited the analyses to our target population (youths from San Antonio or its surrounding areas).

Inviting Teens to Share Their Stories for Inclusion in Project YES

San Antonio youth volunteers were invited to contribute their lived experience–based stories of coping with depression, anxiety, and related difficulties for inclusion in the culturally adapted Project YES SSIs. The procedures for youths' involvement in SSI adaptation were reviewed and deemed "Not Regulated Research" by the University IRB. Participating youths were members of the UTTH's Youth Leadership Council who expressed interest in sharing their stories and being a part of a program geared toward helping San Antonio–based teens. UTTH also shared the opportunity with teenagers via their Facebook (Meta Platforms, Inc) and Instagram pages. Youths interested in sharing their stories were then contacted and provided with further information about Project YES and the process of gathering stories. In total, 14 San Antonio youths (aged 12 to 17 years) elected to contribute their personal stories of mental health and coping to Project YES; parents' or guardians' written permission was obtained for all 14 volunteers. Ultimately, UTTH gathered 33 stories from these 14 youths, all of which were integrated into Project YES.

Collecting Youth Stories for Inclusion in Project YES

UTTH and Stony Brook University representatives met with the youth volunteers via Zoom (Zoom Video Communications, Inc), in groups of 3 to 5, on multiple dates to ensure that each volunteer had an opportunity to write and submit their stories to UTTH. Each volunteer was asked to attend 2-hour-long Zoom meetings to construct their stories for inclusion in the interventions. Our focus on adapting and embedding San Antonio youths' personal stories into Project YES reflects Heim and Kohrt's Cultural Adaptation Framework for Scalable Interventions [32], which highlights the core importance of integrating cultural concepts of distress into adapted interventions. In other words, we aimed to situate each SSI's core concepts and skills (eg, behavioral activation and self-kindness) within San Antonio youths' lived experiences, perspectives, and narratives of *how they have experienced and understood their own distress* as well as *how a particular skill has helped them overcome or manage that experience of distress.* By framing the purpose and relevance of the SSI content through the lens of San Antonio youths' voices and perspectives and by making key "non-specific" adaptation to signal cultural relevance (eg, a full Spanish translation), we aimed to increase the program's relevance to users' worldviews, thus improving the acceptability and clinical impact of the interventions.

Stony Brook University and UTTH staff facilitated small breakout sessions. Confidentiality was reviewed at the beginning of the session, and a mental health professional was on hand to provide assistance if any youth volunteer needed help when telling their stories. The youths were asked questions about their experiences at school, home, and community where they may have been stressed, anxious, or challenged and how they overcame those situations. The youth volunteers verbally shared their answers, and one of the facilitators scribed each volunteer's stories. The youths were also offered the option of writing their stories, rather than narrating them, based on their comfort level and personal preference. At the end of the meeting, a written summary of their stories was provided to each youth volunteer; they were then invited to provide feedback and edit their transcribed stories to ensure that each story optimally reflected their lived experience and perspective. All youths' stories were included (anonymously or using pseudonyms) in the final, adapted version of Project YES for San Antonio youths—either within one of the digital SSIs or immediately after the SSIs as part of a collection of "coping narratives" from peers, which all Project YES users had the chance to review.

Translating Project YES Into Spanish

UTTH staff, who were either native Spanish speakers or fluent in Spanish, translated each story into Spanish. Several youth volunteers were bilingual and translated their own stories into Spanish. The Project YES home page [35] and all 3 SSIs within Project YES were also translated into Spanish by an externally hired professional translator, along with input from UTTH.

Recording Youths' Stories for Inclusion in Project YES

Within the Project YES SSIs, narratives from young people are available as both *written stories* and *audio recordings* to help users feel more connected to their peers' mental health and coping narratives. Therefore, UTTH contracted the UT Health San Antonio Media Services Department to provide the studio and equipment to record all stories in English and Spanish. All youth volunteers were offered the opportunity to self-record their own stories, although none was required to do so.

Multiple recording dates were arranged to accommodate the youth volunteers' school schedules. Each volunteer received a copy of the story or stories they had written, and they were guided by the media services director on how to properly use the recording equipment. The youths were able to make changes to their stories' wording if deemed necessary, and the director lent periodic guidance to improve the flow and intonation. Two of the bilingual youth volunteers recorded each of their stories in Spanish. Media services staff edited the files to accommodate the multiple takes in each volunteer's recording and then provided a final file to UTTH (1 audio file per story). After the recordings, the youth volunteers were provided with a "swag bag" and gift cards in appreciation for their time, dedication, and contribution to the project (US $85 for each volunteer, for several hours of effort). Parents of the youth volunteers also

received gift cards for supporting their children in this work (US $50).

Interventions

Overview

The SSIs in Project YES were designed to incorporate 4 design features common across effective, self-guided SSIs for youth mental health, as proposed by Schleider et al [10]. Each SSI within Project YES (1) incorporates true stories from trustworthy individuals, including older peers (in this case, stories and narratives from San Antonio youths; refer to Multimedia Appendix 1 for examples) and scientific experts; (2) use brain science–based explanations to increase the credibility of the content; (3) empowers users to act as an "expert" or "helper" throughout the SSI by leveraging their own experience to offer advice to others; and (4) offers guided writing activities, often referred to as "saying-is-believing" or "self-persuasion" exercises. A detailed explanation of these 4 SSI design features has been provided elsewhere [10]. The culturally adapted version of Project YES, which was included in the present evaluation, incorporated new narratives constructed with San Antonio youths via the processes described earlier.

Materials for all SSIs within Project YES are publicly available via the Open Science Framework (Project Personality [36], Project CARE [37], and The ABC Project [38]). Each SSI is a 30-minute, self-delivered program. All the materials that youths who participate in Project YES view during the program are publicly viewable on the Project YES website [35].

There are currently 2 published reports on the preliminary acceptability and utility of the Project YES interventions across all youths who completed an intervention between September 2019 and August 2020 [10,12]. The results demonstrated acceptability and positive effects on both hopelessness and perceived agency for all 3 interventions [10].

Project Personality

The original (unadapted) version of Project Personality has been described elsewhere [10], and its structure is summarized here for convenience. Project Personality opens with an introduction to the human brain and a lesson on neuroplasticity. This SSI includes true stories from older youths, which explain their views that personality traits are malleable and describe moments in their lives when they used "growth mindsets" to persevere during setbacks. Project Personality also explains how and why one's personality can change and includes an exercise in which youths incorporate scientific information in the notes they write to younger youths about people's capacity for change. This SSI was adapted to include stories provided by San Antonio youths for San Antonio youths. Project Personality has shown effectiveness in multiple trials [32,39,40].

Project CARE

The original (unadapted) version of Project CARE has been described elsewhere [10]; however, its structure is summarized here for convenience. Project CARE begins with an introduction to the science that explains why many youths believe that they must dislike themselves to be successful and thus fear self-compassion. This SSI contains scientific evidence and true stories from other youths explaining that being self-compassionate predicts greater social and academic success. Evidence-based tips for overcoming the fear of self-compassion–based obstacles are provided. Youths have the opportunity to author notes to younger adolescents using scientific information to explain the different ways in which self-kindness can be beneficial. This SSI was adapted to include stories provided by San Antonio youths for San Antonio youths.

The ABC Project

The original (unadapted) version of The ABC Project has been described elsewhere [10]; however, its structure is summarized here for convenience. The ABC Project incorporates components of behavioral activation and introduces the idea that participating in value-based activities can combat low self-esteem and a sad mood. Through The ABC Project, youths receive psychoeducation about depression, specifically about how one's behavior can shape one's thoughts and feelings. Youths are walked through a life-values assessment in which they identify key areas from which they gain enjoyment and meaning; next, they identify and personalize 3 activities to focus on for change. Finally, The ABC Project offers an activity in which youths write about engaging in each of their targeted activities and the benefits that may come from engaging in them; they also pinpoint obstacles that may stop them from doing the activities and outline strategies to overcome the obstacles.

Measures

The following measures on Project YES are the same as those that prior reports have described; however, in the context of this data collection, we focused on the version of Project YES that contains both the English and Spanish versions.

Demographics

The participants identified their age range (age was provided in ranges to maintain anonymity: 11 to 13, 14 to 16, or 17 years); biological sex; gender identity; sexual orientation; race and ethnicity; how they learned about Project YES; and whether they were from San Antonio, Texas, or the surrounding areas. If they selected "Yes," they were asked in what part of San Antonio they lived (north, east, south, or west).

Mood and Feelings Questionnaire–Short

The Mood and Feelings Questionnaire—Short (SMFQ) is a valid, reliable, and commonly used assessment for depression symptoms in youths, which was previously translated into English and Spanish [41]. Before using an SSI, the participants rated their agreement with 13 statements reflecting thoughts and feelings over the previous 2 weeks (eg, "I felt lonely," "I felt miserable or unhappy," and "I felt I was no good anymore") on a 3-point Likert scale (0="not true," 1="sometimes," and 2="true"). Internal consistencies among the San Antonio youths who completed Project YES in English and Spanish, respectively, were Cronbach α=.93 and Cronbach α=.91. Note that the SMFQ does not measure suicidal ideation or suicidality, which are not assessed in the context of Project YES.

State Hope Scale

The State Hope Scale is a 6-item self-report scale created to evaluate hope in youths. The State Hope Scale includes 2

reliable subscales: agency and pathways [42]. The "agency" subscale measures one's perceived ability to create plans and work to achieve goals (eg, "I can think of many ways to reach my current goals"); the "pathways" subscale reports one's perceived success in meeting one's goals (eg, "At this time, I am meeting the goals I have set for myself"). For Project YES, we anticipated shifts in *agency* but not in *pathways* scores. This is because participants may not have opportunities to pursue their goals in new ways immediately after SSI. Thus, hope was indexed using the 3-item agency subscale of the State Hope Scale. The participants rated 3 statements immediately before and after intervention to reflect how they felt about themselves at that moment on an 8-point Likert scale (1="definitely false" to 8="definitely true"). Internal consistency among the San Antonio youths who completed Project YES in English was Cronbach α=.78 and Cronbach α=.91 before and after SSI, respectively. Internal consistency among the San Antonio youths who completed Project YES in Spanish was Cronbach α=.86 and Cronbach α=.86 before and after SSI, respectively.

Beck Hopelessness Scale-4

The Beck Hopelessness Scale-4 is a reliable, commonly used, shortened version of the 20-item scale that measures hopelessness in youths [42,43]. The participants rated 4 statements immediately before and after intervention to indicate their sense of hopelessness "right now, in this moment" on a 4-point Likert scale (0="absolutely disagree," 1="somewhat disagree," 2="somewhat agree," and 3="absolutely agree"). Internal consistency among the San Antonio youths who completed Project YES in English was Cronbach α=.87 and Cronbach α=.91 before and after SSI, respectively. Internal consistency among the San Antonio youths who completed Project YES in Spanish was Cronbach α=.76 and Cronbach α=.91 at before and after SSI, respectively.

Self-hate Scale

The Self-hate Scale is a reliable, 3-item measure designed to assess feelings of self-hate ([15,44]; adapted from the study by Turnell et al [44]). Immediately before and after the intervention, the participants rated how true each of 3 statements was for them at that moment ("I hate myself," "I feel disgusted when I think about myself," and "I feel ashamed of myself") on a 6-point Likert scale (1="not at all true for me" to 6="very true for me"). Internal consistency among the San Antonio youths who completed Project YES in English was Cronbach α=.93 and Cronbach α=.96 before and after SSI, respectively. Internal consistency among the San Antonio youths who completed Project YES in Spanish was Cronbach α=.86 and Cronbach α=.95 at before and after SSI, respectively.

Perceived Change in Hopelessness and Problem-solving

The participants were asked 2 questions immediately after intervention, which evaluated their perceived change in hopelessness and ability to solve problems. Questions were developed for this study based on established guidelines for assessing subjectively perceived change following an intervention [45]. These questions asked, "to what extent are you feeling hopeless right now?" and "to what extent are you able to solve the problems facing you right now?" when "compared to before doing this activity." Both perceived change in hopelessness and problem-solving ability were rated on a 5-point Likert scale ("much more hopeless" to "a lot less hopeless"; "much less able to solve problems" to "a lot more able to solve problems"). These measures were developed based on previously established methods used to calculate the "smallest effect size of interest" (the smallest possible effect size associated with a detectable, subjective change within individuals) [46].

Program Feedback Scale

The Program Feedback Scale (PFS) is commonly used to evaluate the acceptability and user perceptions of SSIs [16,20,46,47]. It asks participants to rate their level of agreement with 7 statements that indicate the perceived acceptability and feasibility of the SSI they chose (eg, "I enjoyed the program") on a 5-point Likert scale (1="really disagree," 2="disagree," 3="neutral," 4="agree," and 5="totally agree"). The PFS was adapted from the existing, validated acceptability assessments of digital interventions. To exclude items that did not apply to web-based SSIs (ie, items that reference one's interest in revisiting the program), adaptations from the existing scales were necessary. The PFS also assesses open-response feedback from participants. The PFS item scores can be evaluated either individually or across items via a mean score. Internal consistency across PFS items for the youths who completed Project YES in English and Spanish was Cronbach α=.93 and Cronbach α=.87, respectively; mean responses to *each* PFS item were evaluated separately to gain insight into the acceptability in specific domains (eg, ease of use and understanding and enjoyability).

Analytic Plan

Sample Characterization and Use Patterns

To evaluate the use patterns of Project YES, we quantified the number of youths who started Project YES and those who chose, began, and completed an SSI. In addition, we identified which SSI the youths chose as well as the demographics, symptom levels, hopelessness, and self-hate levels for youths who chose, began, and completed an SSI. Responses from Spanish and English respondents were evaluated separately. We calculated the overall and item-specific means for youths who completed an SSI and the PFS to evaluate each SSI's feasibility and acceptability. Mean scores of >3 for any PFS item demonstrated item endorsement (eg, satisfactory acceptability). Mean scores of >3 across all items demonstrated overall SSI acceptability. Descriptive statistics for pre-to-post SSI "perceived change" items were calculated using responses from the subsample of individuals who completed an SSI. Mean ratings >0 for each item demonstrated a subjectively detectable overall pre-to-post SSI change for that dimension (problem-solving ability or hopelessness).

SSI Effects on Proximal Outcomes

We calculated within-group effect sizes (Cohen *d*, including 95% CIs), which reflected the change in pre-to-post SSI levels of each post-SSI outcome variable (agency, hopelessness, and self-hate) across SSIs. Because the previous reports of outcomes for Project YES did not indicate significant differences across

SSI selection [10,12], all outcome data were collapsed for analytical purposes. There are several ways to compute Cohen *d* for within-subject designs [48]. Here, we report Cohen d_z, which accounts for within-subject correlations between pre- and post-SSI measures:

$$\text{Cohen } d_z = M_{diff} / \sqrt{(\sum((x_{diff} - M_{diff})^2 / (N - 1))} \quad \textbf{(1)}$$

Data Quality and Exclusions

Within self-report surveys, particularly within community-based program evaluations (such as this project), it is common for a subset of participants not to pay sufficient attention or put sufficient effort into providing valid questionnaire responses. Because prior work demonstrates that even a small percentage of invalid responses can undermine the interpretability of results [49], it is necessary to identify and filter careless or invalid effort (CIE) responses to optimize the interpretation and improve the accuracy of results. Because of the naturalistic nature of data collection in this project, we used a filtering strategy that prioritized specificity (detecting and removing *only the most clearly CIE* responses) over sensitivity (detecting and removing *all potentially CIE responses*). Specifically, rather than filtering data based on less-diagnostic metrics (such as time spent on the survey), we identified and excluded responses that demonstrated a "straightlining" pattern, whereby an individual provides identical responses to all survey items regardless of their content and direction [50]. Of the various approaches to identifying CIE responses, straightlining has been shown to have the most pronounced impact on data properties [50]. Here, we defined "straightlining" as providing numerically identical responses across *all 3* Likert-scale based outcomes, either before SSI or after SSI (the Beck Hopelessness Scale-4 item version, for which higher numeric responses reflect poorer functioning; the State Hope Scale—Agency subscale, on which higher numeric responses reflect more positive functioning; and the Self-hate Scale, on which higher numeric responses reflect poorer functioning). To optimize the validity of the results, the individuals who followed "straightlining" response patterns (ie, responding to all pre- or post-SSI Likert-scale items with "1," regardless of item direction or content) were removed from the effect size computation analyses.

All available data meeting the above-described inclusion criteria were used for each test described above, and data from Spanish and English respondents were evaluated separately. As use patterns (including attrition) were of direct empirical interest, missing data rates are reported but not imputed. Anonymized data and codes for all analyses are available via the Open Science Framework.

Results

Sample and Use Patterns in Project YES

Between April and December 2021, a total of 1801 San Antonio youths began Project YES, of whom 894 (49.64%) completed a 30-minute, SSI. Specifically, 1705 youths began Project YES in English and 855 of them completed it (completion rate: 50.15%). A total of 96 youths began Project YES in Spanish and 39 of them completed it (completion rate: 41%). The overall completion rate of 49.64% is considerably higher than those previously observed for Project YES (34.2%). The youths who participated in Project YES were demographically diverse. Across all youths who began Project YES in English, 94.78% (1616/1705) identified as people of color, 53.96% (920/1705) as biologically female, and 39.3% (670/1705) as nonheterosexual. There were 1464 youths who began an SSI activity in English. Among those who started an SSI in English, 29.71% (435/1464) chose Project CARE (257/435, 59.1% completion rate among those who began the SSI), 23.43% (343/1705) chose The ABC Project (206/343, 60.1% completion rate among those who began the SSI), and 46.86% (686/1464) chose Project Personality (392/686, 57.1% completion rate among those who began the SSI). Completion rates did not substantially differ between the 3 SSIs.

Across all youths who began Project YES in Spanish, 99% (95/96) identified as people of color, 43% (41/96) as biologically female, and 33% (32/96) as nonheterosexual. There were 77 youths who began an SSI in Spanish. (Notably, beginning an SSI requires the completion of all pre-SSI questionnaires; as such, 96 youths began Project YES in Spanish, but only 77 youths completed the pre-SSI questionnaires and initiated an intervention in Spanish.) Among those who started an SSI in Spanish, 29% (22/77) chose Project CARE (8/22, 36% completion rate among those who began the SSI), 25% (19/77) chose The ABC Project (11/19, 58% completion rate among those who began the SSI), and 47% (36/77) chose Project Personality (20/36, 56% completion rate among those who began the SSI). Completion rates did not substantially differ between the 3 SSIs. Across the total sample (both Spanish and English completers), sexual orientation (nonheterosexual vs heterosexual) was associated with the odds of completion; specifically, nonheterosexual participants were less likely to complete their selected SSI. Across the total sample, racial and ethnic identity, biological sex (female vs intersex vs male), and depressive symptom severity (total SMFQ score) were not associated with the odds of completion.

Youths who completed Project YES in English in its entirety spent 34.6 (SD 21.2; median 32.7; range 6.27-409) minutes on the YES website on average; this included the time spent on all questionnaires, the SSI they chose, and authoring anonymous advice for the YES Advice Center. Youths who completed Project YES in Spanish in its entirety spent 40.3 (SD 20.0; median 37.2; range 19.1-105) minutes on the YES website on average. Across all youths who accessed the YES website (including those who neither started nor completed an SSI), the average amount of time spent on Project YES was 29.8 minutes for the English version and 28.6 minutes for the Spanish version.

Did the Youths Perceive Project YES as Acceptable?

The youths who completed both an SSI and the PFS (794/1801, 44.09%) in English, collapsing both across and within each SSI, found Project YES to be acceptable. Overall, these youths rated the SSI they chose as enjoyable (3.57/5.00), easy to understand (3.94/5.00), easy to use (3.99/5.00), likely to help their peers (3.89/5.00), and worth recommending to others (3.77/5.00). In addition, the youths endorsed that they tried their hardest on the SSI they chose (3.65/5.00) and that they agreed with the SSIs message (4.02/5.00).

The youths who completed both an SSI and the PFS (37/1801, 2.05%) in Spanish, collapsing both across and within each SSI, found Project YES to be acceptable. Overall, these youths rated the SSI they chose as enjoyable (3.95/5.00), easy to understand (4.05/5.00), easy to use (4.22/5.00), likely to help their peers (4.14/5.00), and worth recommending to others (4.24/5.00). In addition, the youths endorsed that they tried their hardest on the SSI they chose (4.16/5.00) and that they agreed with the SSIs message (4.27/5.00; Table 1).

Table 1. Means and SDs of the Program Feedback Scale items among single-session intervention (SSI) completers, across all SSIs in English and Spanish.

Item	All SSIs—English, mean (SD)	All SSIs—Spanish, mean (SD)
Enjoy	3.57 (1.04)	3.95 (1.13)
Understood	3.94 (0.97)	4.05 (1.15)
Easy to use	3.99 (1.00)	4.22 (1.03)
Tried hardest	3.65 (1.10)	4.16 (1.14)
Helpful	3.89 (1.11)	4.14 (1.23)
Recommend to friend	3.77 (1.17)	4.24 (0.93)
Agree with message	4.02 (1.03)	4.27 (1.04)
Full scale	3.84 (0.88)	4.15 (0.83)

The respondents who completed the PFS in English left the following comments in it: “User friendly and the examples of kids my age made me feel understood and not alone”; “I like that it is interactive and has reflection built in. The balance of science and real life experiences was great”; “I learned how to be nicer to myself”; and “It was very helpful for me to improve on things.” The respondents who completed the PFS in Spanish left the following comments: “Me gusto que le cambió la vida a otros y ami misma” (I liked that it changed the lives of others and myself); “Que puede aplicar mis futuros problemas en el ejemplo” (That I can apply my future problems in the example); and “Me gusto como me hizo pensar en vez de nada más dando ejemplos sobre otros” (I liked how it made me think instead of just giving examples about others).

Did Hopelessness, Self-hate, and Perceived Agency Improve From Before to After Project YES?

To optimize the validity and accuracy of the effect size estimates, we identified and excluded the youths who demonstrated a “straightlining” pattern of responses across all pre- or post-SSI Likert-scale items that assessed the psychosocial outcomes of interest (hopelessness, self-hate, and agency). In total, 3.9% (33/855) and 2.9% (25/855) of the English-language respondents showed a straightlining pattern in the pre-SSI and post-SSI surveys, respectively; 3% (1/39) and 8% (3/39) of the Spanish-language respondents showed a straightlining pattern in the pre-SSI and post-SSI surveys, respectively. The results reported subsequently reflect the effect sizes observed after the removal of these responses.

SSI short-term utility was tested both across SSIs, as they have shared principles and common structures, and separately, as each SSI is distinct in content. The effect sizes Cohen d_z and 95% CIs both across SSIs and for each SSI are reported (Table 2). Across all 3 SSIs in the English version of Project YES, the youths reported significant improvements in all proximal outcomes pre-to-post SSI. Regarding overall reductions in hopelessness, small-medium effects emerged (2-tailed t_{712}=8.87; P<.001; 95% CI 0.26-0.41, Cohen d_z=0.33), with post-SSI hopelessness showing a 57.6% chance of being lower than pre-SSI hopelessness (per the “common language effect size” estimate; refer to the study by Lakens [48]). For overall reductions in self-hate, small-medium effects emerged (t_{698}=7.11; P<.001; 95% CI 0.19-0.34, Cohen d_z=0.27), with post-SSI self-hate showing a 55% chance of being lower than pre-SSI self-hate. For overall improvements in perceived agency, small-medium effects emerged (t_{707}=6.54; P<.001; 95% CI 0.17-0.32, Cohen d_z=0.25), with post-SSI perceived agency showing a 56.2% chance of being higher than pre-SSI perceived agency.

Across all 3 SSIs in the Spanish version of Project YES, the youths reported a significant improvement in self-hate from before to after SSI. For overall reductions in self-hate, small-medium effects emerged (t_{29}=2.05; P=.049; 95% CI 0.001-0.74, Cohen d_z=0.37), with post-SSI self-hate showing a 55% chance of being lower than pre-SSI self-hate (per the “common language effect size” estimate; refer to the study by Lakens [48]). The youths did not report a substantial improvement in hopelessness or perceived agency.

SSI-specific effect sizes were not calculated, as the programs do not appear to have differential effects; therefore, there was no reason to expect differential effects in this study [10,15].

Table 2. Means, SDs, and effect sizes by single-session intervention (SSI) and across all SSIs in English and Spanish[a].

Outcome variable	English				Spanish			
	ABC Project	Project CARE	Project Personality	All SSIs	ABC Project	Project CARE	Project Personality	All SSIs
Agency								
Before SSI, mean (SD)	5.3 (1.7)	5.2 (1.8)	5.5 (1.7)	5.3 (1.7)	6.3 (1.5)	5.9 (1.6)	5.6 (2.0)	5.8 (1.8)
After SSI, mean (SD)	5.8 (1.9)	5.5 (2.1)	5.9 (1.9)	5.7 (2.0)	6.6 (1.8)	5.1 (2.7)	5.8 (2.0)	5.9 (2.1)
Cohen d_z (95% CI)	*0.25 (0.10 to 0.40)* [b]	*0.24 (0.10 to 0.38)*	*0.25 (0.14 to 0.36)*	*0.25 (0.17 to 0.32)*	0.07 (−0.56 to 0.69)	0.39 (−0.46 to 1.20)	*0.57 (0.01 to 1.11)*	0.12 (−0.23 to 0.47)
Hopelessness								
Before SSI, mean (SD)	1.2 (0.9)	1.1 (0.9)	1.0 (0.8)	1.0 (0.9)	1.0 (0.8)	1.2 (1.0)	1.9 (0.8)	1.0 (0.8)
After SSI, mean (SD)	0.8 (0.9)	0.8 (0.8)	0.7 (0.8)	0.8 (0.9)	1.1 (1.0)	1.1 (1.2)	1.7 (0.9)	0.9 (1.0)
Cohen d_z (95% CI)	*0.37 (0.22 to 0.52)*	*0.38 (0.24 to 0.52)*	*0.28 (0.17 to 0.39)*	*0.33 (0.26 to 0.41)*	0.34 (−0.31 to 0.97)	0.17 (−0.58 to 0.91)	0.02 (−0.46 to 0.49)	0.00 (−0.34 to 0.34)
Self-hate								
Before SSI, mean (SD)	2.7 (1.8)	2.6 (1.7)	2.3 (1.6)	2.5 (1.7)	1.9 (1.3)	2.1 (1.6)	2.2 (1.6)	2.1 (1.5)
After SSI, mean (SD)	2.2 (1.6)	2.3 (1.6)	2.0 (1.5)	2.1 (1.6)	2.5 (2.0)	1.9 (2.0)	1.4 (0.8)	1.8 (1.6)
Cohen d_z (95% CI)	*0.37 (0.22 to 0.53)*	*0.21 (0.07 to 0.35)*	*0.25 (0.14 to 0.36)*	*0.27 (0.19 to 0.34)*	0.10 (−0.56 to 0.75)	0.55 (−0.34 to 1.39)	0.47 (−0.07 to 0.99)	*0.37 (0.00 to 0.74)*

[a]For hopelessness and self-hate measures, lower scores indicate better functioning, and for perceived agency, higher scores indicate better functioning. Where applicable, Cohen *d* values are corrected (multiplied by −1.0) such that positive values indicate greater improvements from before to after SSI.

[b]Significant improvements are italicized.

Did the Youths Subjectively Detect Changes in Hopelessness and Problem-Solving Ability From Before to After Project YES?

Among the youths who completed Project YES in English, after the intervention, 32.9% (259/785) reported feeling "much less hopeless," 30.8% (242/785) felt "a little less hopeless," 23.5% (185/785) felt "the same amount hopeless," 5.3% (42/785) felt "a little more hopeless," and 7.5% (59/785) felt "a lot more hopeless" compared with before beginning the SSI. Separately, 27.2% (214/785) of the youths reported feeling "much more able to solve problems," 31.3% (246/785) felt "a little more able to solve problems," 28.6% (225/785) felt "the same amount able to solve problems," 7% (55/785) felt "a little less able to solve problems," and 5.9% (47/785) felt "a lot less able to solve problems" compared with before beginning the SSI.

Among the youths who completed Project YES in Spanish, after the intervention, 38% (14/37) reported feeling "Mucho menos desesperanzado" ("much less hopeless"), 27% (10/37) felt "Un poco menos desesperanzado" ("a little less hopeless"), 8% (3/37) felt "Lo mismo de desesperanzado" ("the same amount hopeless"), 8% (3/37) felt "Un poco más desesperanzado" ("a little more hopeless"), and 19% (7/37) felt "Mucho más desesperanzado" ("a lot more hopeless") compared with before beginning the SSI. Separately, 43% (16/37) of the youths reported feeling "Mucho más capaz de resolver los problemas" ("much more able to solve problems"), 24% (9/37) felt "Un poco más capaz de resolver los problemas" ("a little more able to solve problems"), 22% (8/37) felt "La misma capacidad de resolver los problemas" ("the same amount able to solve problems"), 8% (3/37) felt "Un poco menos capaz de resolver los problemas" ("a little less able to solve problems"), and 3% (1/37) felt "Mucho menos capaz de resolver los problemas" ("a lot less able to solve problems") compared with before beginning the SSI.

Discussion

Principal Findings

This academic-community partnership project sought to evaluate the acceptability and short-term utility of an open-access, culturally adapted digital platform for English- and Spanish-speaking adolescents, providing three 30-minute self-administered SSIs to young people living in the city of San Antonio, Texas. From before to after the intervention, the youths who completed Project YES in English reported short-term improvements in hopelessness (Cohen *d*=0.33), self-hate (Cohen *d*=0.27), and perceived agency (Cohen *d*=0.25). In addition, the

youths who completed Project YES in Spanish reported improvements in self-hate (Cohen *d*=0.37) from before to after the intervention. The youths who completed an SSI in either language rated it as enjoyable, easy to understand, likely to help peers, and worth recommending to others. On the basis of the largely overlapping CIs for Cohen d_z, no evidence emerged for the differential effects of individual SSIs based on proximal outcomes. That is, the youths who completed Project YES reported similar short-term improvements in self-hate, agency, and hopelessness regardless of the specific SSI they opted to complete—consistent with prior trials comparing the impacts of the SSIs in Project YES [10,15].

Findings build on a growing body of evidence for single-session digital interventions as scalable and accessible mental health tools for young people [10,15]. Past studies have reported the acceptability and utility of culturally adapted internet-based interventions for mental disorders [51]. This project provides evidence of the acceptability of cultural adaptations of SSIs. The youth stories, focus groups, and translations involved teens directly from the local community during the adaptation process. Furthermore, the mean SMFQ score for the sample was 9.97, with users' scores ranging from 0 to 26, suggesting usefulness for both young people with lower symptoms and those with higher symptoms (ie, as a preventive tool), and for youths already experiencing substantial distress. This project suggests its utility and acceptability as a mental health resource when adapted for the needs of a specific underserved community or population of San Antonio, Texas.

The composition of the sample was highly representative of the community and very diverse relative to most youth mental health studies, with >94% the teens identifying as people of color (predominantly Latino, Latina, Latine, or Hispanic; African American; Asian; and other). Therefore, the findings of this study are likely to be generalizable to other youths in this area. Furthermore, the high percentage of lesbian, gay, bisexual, transgender, queer and more (LGBTQ+) youths in the sample (approximately 33%-39%) provides evidence for the project's generalizability and usability. However, LGBTQ+ youths were less likely to complete their selected SSI, and further work should investigate this notable finding. The sample in this study consisted of a more even gender or biological sex balance than those in other SSI trials [10,11,15], potentially because of community-facilitated intervention dissemination (eg, in classrooms). Future work on digital mental health supports may benefit from using community-facilitated dissemination as a tool for engaging more males.

Similar to other web-based SSI evaluations [10,15,52], the youths in this project showed an SSI completion rate of 49.6%. Across interventions, among the youths who began an SSI, 36.6% to 60.1% completed the full activity. This relatively high completion rate provides further evidence of youth acceptability for web-based SSIs, as prior studies of self-guided digital mental health programs have shown substantially lower completion rates (0%-28%, on average) [7]. Furthermore, pre-SSI youth dropout (ie, dropout occurring during the pre-SSI questionnaire stage) was only 14.4% in this project, which was substantially lower than the pre-SSI dropout rate observed in a previous program evaluation of Project YES (37.2%) [17]. Proactive, community-partnered dissemination efforts (rather than relying exclusively on social media–based dissemination) may have contributed to the notably higher completion rates in this project. Cultural adaptations may have likewise increased the youths' engagement with the SSIs and willingness to complete the programs in full.

In future efforts to disseminate youth-focused web-based SSIs, to further minimize pre-SSI dropouts (which was 14.4% in this project), we strongly suggest minimizing or eliminating pre-SSI questionnaires. As evidence for the acceptability, utility, and safety of these SSIs in diverse populations continues to grow, offering SSIs beyond the context of program evaluations and randomized trials will become increasingly feasible. Relatively few youths chose to complete the SSIs in Spanish. Among the youths who did complete Project YES in Spanish, the completion rates did not differ as a function of the selected SSI, and they reported all SSIs as acceptable and likely to help. Many youths in the San Antonio area are bilingual English-Spanish speakers. Although many of their parents might have been monolingual Spanish speakers, the youths may have chosen to complete Project YES in English out of personal preference or comfort level. Nonetheless, it is possible that the mere availability of a Spanish-translated version of Project YES might have been viewed positively by San Antonio youths. For example, it might have signaled that the program was built with their needs and community in mind, thereby strengthening the feelings of belonging and relatedness in the context of the SSIs. Future studies are needed to empirically examine this possibility. However, it is notable that the exact percentages of primarily Spanish- and English-speaking students at the participating schools were not publicly available; therefore, the possibility of differential Project YES engagement based on primary language remains unclear. Additional research is needed to gauge whether treatment access disparities are maintained or reduced by making Spanish-language youth mental health supports freely available.

Although our results demonstrated overall acceptance of the culturally adapted English and Spanish SSIs in Project YES, our project is not without limitations. As this was a nonexperimental, observational, and anonymous project because of the goals of the study, follow-up data were not collected; therefore, the findings should be interpreted cautiously. However, the same SSIs (albeit not the present culturally adapted versions) have demonstrated efficacy in large randomized control trials across 3- to 9-month follow-up periods [33,38]. Additional trials oversampling Spanish-speaking youths and using the newly adapted versions of the SSIs are needed to gauge the broader and longer-term utility of Project YES, both within and beyond the San Antonio youth community. Furthermore, implementation challenges will be ongoing for the community. Although the resource is free, raising awareness about the platform poses various challenges. Moving ahead, there is a need to examine pathways for integrating Project YES and similar tools into community settings, including schools, to ensure that youths who might benefit from the tool are aware of it and can access it easily, privately, and independently.

Conclusions

Overall, the cultural adaptation of Project YES demonstrated an acceptable, accessible, and useful mental health support for English- and Spanish-speaking San Antonio youths. Community-facilitated dissemination of Project YES may have helped lower the dropout rates, relative to recent evaluations of web-based youth-focused SSIs, and reach a more racially and ethnically diverse population, including engaging more males than in prior SSI trials, and a sizable LGBTQ+ youth population. There remains a need for additional work on effective strategies for disseminating and implementing web-based SSI platforms as a community-wide resource. Moving ahead, web-based SSIs may remain a valuable resource and a low-cost, accessible, and youth-centered means for augmenting, complementing, and expanding resource-strapped mental health care systems such as that of San Antonio, Texas.

Acknowledgments

This work was funded by the City of San Antonio Metropolitan Health District (HSC20200877N) for KAP (principal investigator).

JLS received funding from National Institutes of Health Office of the Director (DP5OD028123), National Institute of Mental Health (R43MH128075), National Science Foundation (grant 2141710), Health Resources and Service Administration (U3NHP45406-01-00), Upswing Fund for Adolescent Mental Health, Society for Clinical Child and Adolescent Psychology, and Klingenstein Third Generation Foundation, outside of this work.

Authors' Contributions

KAP, JLS, AR-H, NAD, and JSPT conceptualized and designed the project. JLS led the development of the original version of Project Youth Empowerment and Support and its web-based single-session interventions. AR-H, NAD, and JSPT led the cultural adaptation and Spanish translation process for Project Youth Empowerment and Support. KAP, AR-H, NAD, and JSPT facilitated the outreach and recruitment efforts. JLS, AS, CR, and JF supported the web-based recruitment efforts. AS, CR, and JF led data management and cleaning. AS performed all data analyses and wrote the initial draft of the *Introduction*, *Results*, and *Discussion* sections; CR and NAD wrote the initial draft of the *Methods* section. All the authors contributed to the review and editing of the final manuscript.

Conflicts of Interest

JLS serves on the Scientific Advisory Board for Walden Wise and the Clinical Advisory Board for Koko, is a Cofounder and Codirector of Single Session Support Solutions Inc, and receives book royalties from New Harbinger, Oxford University Press, and Little Brown Book Group.

Multimedia Appendix 1
Stories from San Antonio youths.
[DOCX File , 17 KB - mental_v10i1e43062_app1.docx]

References

1. Kazdin AE. Annual research review: expanding mental health services through novel models of intervention delivery. J Child Psychol Psychiatry 2019 Apr;60(4):455-472. [doi: 10.1111/jcpp.12937] [Medline: 29900543]
2. Kepley HO, Streeter RA. Closing behavioral health workforce gaps: a HRSA program expanding direct mental health service access in underserved areas. Am J Prev Med 2018 Jun;54(6 Suppl 3):S190-S191 [FREE Full text] [doi: 10.1016/j.amepre.2018.03.006] [Medline: 29779541]
3. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
4. Garcia ME, Ochoa-Frongia L, Moise N, Aguilera A, Fernandez A. Collaborative care for depression among patients with limited English proficiency: a systematic review. J Gen Intern Med 2018 Mar;33(3):347-357 [FREE Full text] [doi: 10.1007/s11606-017-4242-4] [Medline: 29256085]
5. Gruber J, Prinstein MJ, Clark LA, Rottenberg J, Abramowitz JS, Albano AM, et al. Mental health and clinical psychological science in the time of COVID-19: challenges, opportunities, and a call to action. Am Psychol 2021 Apr;76(3):409-426 [FREE Full text] [doi: 10.1037/amp0000707] [Medline: 32772538]
6. Anderson M, Jiang J. Teens, social media and technology 2018. Pew Research Center. 2018 May 31. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2022-03-15]
7. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]

8. Schleider JL. Repairing the research–service rupture in clinical psychological science. Nat Rev Psychol 2022 Jan 12;1(1):2-4. [doi: 10.1038/s44159-021-00005-z]
9. Wasil AR, Gillespie S, Patel R, Petre A, Venturo-Conerly KE, Shingleton RM, et al. Reassessing evidence-based content in popular smartphone apps for depression and anxiety: developing and applying user-adjusted analyses. J Consult Clin Psychol 2020 Nov;88(11):983-993. [doi: 10.1037/ccp0000604] [Medline: 32881542]
10. Schleider JL, Dobias M, Sung J, Mumper E, Mullarkey MC. Acceptability and utility of an open-access, online single-session intervention platform for adolescent mental health. JMIR Ment Health 2020 Jun 30;7(6):e20513 [FREE Full text] [doi: 10.2196/20513] [Medline: 32602846]
11. Dobias ML, Schleider JL, Jans L, Fox KR. An online, single-session intervention for adolescent self-injurious thoughts and behaviors: results from a randomized trial. Behav Res Ther 2021 Dec;147:103983. [doi: 10.1016/j.brat.2021.103983] [Medline: 34688102]
12. McDanal R, Rubin A, Fox KR, Schleider JL. Associations of LGBTQ+ identities with acceptability and efficacy of online single-session youth mental health interventions. Behav Ther 2022 Mar;53(2):376-391. [doi: 10.1016/j.beth.2021.10.004] [Medline: 35227411]
13. Schleider JL, Weisz JR. Reducing risk for anxiety and depression in adolescents: effects of a single-session intervention teaching that personality can change. Behav Res Ther 2016 Dec;87:170-181 [FREE Full text] [doi: 10.1016/j.brat.2016.09.011] [Medline: 27697671]
14. Sung JY, Mumper E, Schleider JL. Empowering anxious parents to manage child avoidance behaviors: randomized control trial of a single-session intervention for parental accommodation. JMIR Ment Health 2021 Jul 06;8(7):e29538 [FREE Full text] [doi: 10.2196/29538] [Medline: 34255718]
15. Schleider JL, Mullarkey MC, Fox KR, Dobias ML, Shroff A, Hart EA, et al. A randomized trial of online single-session interventions for adolescent depression during COVID-19. Nat Hum Behav 2022 Feb;6(2):258-268 [FREE Full text] [doi: 10.1038/s41562-021-01235-0] [Medline: 34887544]
16. Dobias ML, Chen S, Fox K, Schleider JL. Brief interventions for self-injurious thoughts and behaviors in young people: a systematic review. Open Science Framework. 2020 Mar 4. URL: https://osf.io/e4jwv/ [accessed 2022-03-20]
17. Cohen KA, Schleider JL. Adolescent dropout from brief digital mental health interventions within and beyond randomized trials. Internet Interv 2022 Mar;27:100496 [FREE Full text] [doi: 10.1016/j.invent.2022.100496] [Medline: 35257001]
18. Sorkin DH, Janio EA, Eikey EV, Schneider M, Davis K, Schueller SM, et al. Rise in use of digital mental health tools and technologies in the United States during the COVID-19 pandemic: survey study. J Med Internet Res 2021 Apr 16;23(4):e26994 [FREE Full text] [doi: 10.2196/26994] [Medline: 33822737]
19. Hwang WC, Myers HF, Chiu E, Mak E, Butner JE, Fujimoto K, et al. Culturally adapted cognitive-behavioral therapy for Chinese Americans with depression: a randomized controlled trial. Psychiatr Serv 2015 Oct;66(10):1035-1042 [FREE Full text] [doi: 10.1176/appi.ps.201400358] [Medline: 26129996]
20. Magaña S, Lopez K, Salkas K, Iland E, Morales MA, Garcia Torres M, et al. A randomized waitlist-control group study of a culturally tailored parent education intervention for Latino parents of children with ASD. J Autism Dev Disord 2020 Jan;50(1):250-262. [doi: 10.1007/s10803-019-04252-1] [Medline: 31606887]
21. Park AL, Rith-Najarian LR, Saifan D, Gellatly R, Huey SJ, Chorpita BF. Strategies for incorporating culture into psychosocial interventions for youth of color. Evid Based Pract Child Adolesc Ment Health (forthcoming) 2022 Jan 20:1-13. [doi: 10.1080/23794925.2022.2025629]
22. Ramos G, Ponting C, Labao JP, Sobowale K. Considerations of diversity, equity, and inclusion in mental health apps: a scoping review of evaluation frameworks. Behav Res Ther 2021 Dec;147:103990 [FREE Full text] [doi: 10.1016/j.brat.2021.103990] [Medline: 34715396]
23. U.S. Census Bureau quickfacts: United States. U.S. Census Bureau. 2020. URL: https://www.census.gov/quickfacts/US [accessed 2022-03-20]
24. Lee S, Held ML. Variation in mental health service use among U.S. Latinos by place of origin and service provider type. Psychiatr Serv 2015 Jan 01;66(1):56-64. [doi: 10.1176/appi.ps.201300533] [Medline: 25179526]
25. Krogstad JM, Stepler R, Lopez MH. English Proficiency on the Rise Among Latinos. Pew Research Center. 2015 May 12. URL: https://www.pewresearch.org/hispanic/2015/05/12/english-proficiency-on-the-rise-among-latinos/ [accessed 2022-03-22]
26. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. JMIR Mhealth Uhealth 2016 Apr 19;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]
27. Muñoz AO, Camacho E, Torous J. Marketplace and literature review of Spanish language mental health apps. Front Digit Health 2021 Feb 15;3:615366 [FREE Full text] [doi: 10.3389/fdgth.2021.615366] [Medline: 34713093]
28. Youth Data 2022. Mental Health America. URL: https://www.mhanational.org/issues/2022/mental-health-america-youth-data [accessed 2022-03-22]
29. MUA Find. HRSA Data Warehouse. URL: https://data.hrsa.gov/tools/shortage-area/mua-find [accessed 2022-08-17]
30. San Antonio, Texas Population 2022 (Demographics, Maps, Graphs). World Population Review. URL: https://worldpopulationreview.com/us-cities/san-antonio-tx-population [accessed 2022-03-25]

31. Schleider JL, Sung J, Bianco A, Gonzalez A, Vivian D, Mullarkey MC. Open pilot trial of a single-session consultation service for clients on psychotherapy wait-lists. Behav Ther 2021 Jan 15;44(1):8-15. [doi: 10.31234/osf.io/fdwqk]
32. Heim E, Kohrt BA. Cultural adaptation of scalable psychological interventions. Clin Psychol Eur 2019 Dec 17;1(4):1-22.
33. Schleider J, Weisz J. A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. J Child Psychol Psychiatry 2018 Feb;59(2):160-170. [doi: 10.1111/jcpp.12811] [Medline: 28921523]
34. Schleider JL, Sung J, Dobias ML, Mumper E, Mullarkey MC. Project YES. Open Science Framework. 2019 Oct 3. URL: https://osf.io/e52p3 [accessed 2023-02-08]
35. Project YES. Lab for Scalable Mental Health. URL: https://www.schleiderlab.org/yes.html [accessed 2023-02-08]
36. Schleider JL. Project Personality. Open Science Framework. 2019 Sep 12. URL: https://osf.io/259jv/ [accessed 2022-03-25]
37. Schleider JL. Sharing feelings project. Open Science Framework. 2019 Sep 23. URL: https://osf.io/swv5b/ [accessed 2023-02-06]
38. Schleider JL, Mullarkey MC, Mumper E, Sung J. The ABC project: action brings change. Open Science Framework. 2019 Sep 12. URL: https://osf.io/qj94c/ [accessed 2022-03-25]
39. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
40. Schleider JL, Burnette JL, Widman L, Hoyt C, Prinstein MJ. Randomized trial of a single-session growth mind-set intervention for rural adolescents' internalizing and externalizing problems. J Clin Child Adolesc Psychol 2020;49(5):660-672 [FREE Full text] [doi: 10.1080/15374416.2019.1622123] [Medline: 31219698]
41. Lundervold AJ, Hinshaw SP, Sørensen L, Posserud MB. Co-occurring symptoms of attention deficit hyperactivity disorder (ADHD) in a population-based sample of adolescents screened for depression. BMC Psychiatry 2016 Feb 25;16:46 [FREE Full text] [doi: 10.1186/s12888-016-0739-3] [Medline: 26915733]
42. Snyder CR, Sympson SC, Ybasco FC, Borders TF, Babyak MA, Higgins RL. Development and validation of the State Hope Scale. J Pers Soc Psychol 1996 Feb;70(2):321-335. [doi: 10.1037//0022-3514.70.2.321] [Medline: 8636885]
43. Perczel Forintos D, Rózsa S, Pilling J, Kopp M. Proposal for a short version of the Beck Hopelessness Scale based on a national representative survey in Hungary. Community Ment Health J 2013 Dec;49(6):822-830. [doi: 10.1007/s10597-013-9619-1] [Medline: 23756722]
44. Turnell AI, Fassnacht DB, Batterham PJ, Calear AL, Kyrios M. The Self-Hate Scale: development and validation of a brief measure and its relationship to suicidal ideation. J Affect Disord 2019 Feb 15;245:779-787. [doi: 10.1016/j.jad.2018.11.047] [Medline: 30448763]
45. Anvari F, Lakens D. Using anchor-based methods to determine the smallest effect size of interest. J Exp Soc Psychol 2021 Sep;96:104159. [doi: 10.1016/j.jesp.2021.104159]
46. Schleider JL, Mullarkey MC, Weisz JR. Virtual reality and web-based growth mindset interventions for adolescent depression: protocol for a three-arm randomized trial. JMIR Res Protoc 2019 Jul 09;8(7):e13368 [FREE Full text] [doi: 10.2196/13368] [Medline: 31290406]
47. Mullarkey M, Dobias M, Sung J, Ahuvia I, Shumake J, Beevers C, et al. Web-based single session intervention for perceived control over anxiety during COVID-19: randomized controlled trial. JMIR Ment Health 2022 Apr 12;9(4):e33473 [FREE Full text] [doi: 10.2196/33473] [Medline: 35230962]
48. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
49. Arias VB, Garrido LE, Jenaro C, Martínez-Molina A, Arias B. A little garbage in, lots of garbage out: assessing the impact of careless responding in personality survey data. Behav Res Methods 2020 Dec;52(6):2489-2505. [doi: 10.3758/s13428-020-01401-8] [Medline: 32462604]
50. DeSimone JA, DeSimone AJ, Harms PD, Wood D. The differential impacts of two forms of insufficient effort responding. Appl Psychol 2018 Apr;67(2):309-338. [doi: 10.1111/apps.12117]
51. Spanhel K, Balci S, Feldhahn F, Bengel J, Baumeister H, Sander LB. Cultural adaptation of internet- and mobile-based interventions for mental disorders: a systematic review. NPJ Digit Med 2021 Aug 25;4(1):128 [FREE Full text] [doi: 10.1038/s41746-021-00498-1] [Medline: 34433875]
52. Dobias ML, Morris RR, Schleider JL. Single-session interventions embedded within Tumblr: acceptability, feasibility, and utility study. JMIR Form Res 2022 Jul 25;6(7):e39004 [FREE Full text] [doi: 10.2196/39004] [Medline: 35877163]

Abbreviations

CIE: careless or invalid effort
IRB: institutional review board
LGBTQ+: lesbian, gay, bisexual, transgender, queer and more
PFS: Program Feedback Scale
SMFQ: Mood and Feelings Questionnaire—Short
SSI: single-session intervention

UTTH: University of Texas Teen Health
YES: Youth Empowerment and Support

Edited by J Torous, G Eysenbach; submitted 28.09.22; peer-reviewed by M Barnett, J Torous; comments to author 06.01.23; revised version received 09.01.23; accepted 09.01.23; published 14.02.23.

Please cite as:
Shroff A, Roulston C, Fassler J, Dierschke NA, Todd JSP, Ríos-Herrera Á, Plastino KA, Schleider JL
A Digital Single-Session Intervention Platform for Youth Mental Health: Cultural Adaptation, Evaluation, and Dissemination
JMIR Ment Health 2023;10:e43062
URL: https://mental.jmir.org/2023/1/e43062
doi:10.2196/43062
PMID:36787180

Original Paper

Associations Between Smartphone Keystroke Metadata and Mental Health Symptoms in Adolescents: Findings From the Future Proofing Study

Taylor A Braund[1,2], PhD; Bridianne O'Dea[1,2], PhD; Debopriyo Bal[2], BSc; Kate Maston[2], MPsych; Mark Larsen[1,2], DPhil; Aliza Werner-Seidler[1,2], PhD; Gabriel Tillman[3], PhD; Helen Christensen[1,2], PhD

[1]Faculty of Medicine and Health, University of New South Wales, Kensington, Australia

[2]Black Dog Institute, University of New South Wales, Randwick, Australia

[3]Institute of Health and Wellbeing, Federation University, Ballarat, Australia

Corresponding Author:
Taylor A Braund, PhD
Faculty of Medicine and Health
University of New South Wales
High St
Kensington, 2052
Australia
Phone: 61 290659255
Email: t.braund@blackdog.org.au

Abstract

Background: Mental disorders are prevalent during adolescence. Among the digital phenotypes currently being developed to monitor mental health symptoms, typing behavior is one promising candidate. However, few studies have directly assessed associations between typing behavior and mental health symptom severity, and whether these relationships differs between genders.

Objective: In a cross-sectional analysis of a large cohort, we tested whether various features of typing behavior derived from keystroke metadata were associated with mental health symptoms and whether these relationships differed between genders.

Methods: A total of 934 adolescents from the Future Proofing study undertook 2 typing tasks on their smartphones through the Future Proofing app. Common keystroke timing and frequency features were extracted across tasks. Mental health symptoms were assessed using the Patient Health Questionnaire-Adolescent version, the Children's Anxiety Scale-Short Form, the Distress Questionnaire 5, and the Insomnia Severity Index. Bivariate correlations were used to test whether keystroke features were associated with mental health symptoms. The false discovery rates of *P* values were adjusted to *q* values. Machine learning models were trained and tested using independent samples (ie, 80% train 20% test) to identify whether keystroke features could be combined to predict mental health symptoms.

Results: Keystroke timing features showed a weak negative association with mental health symptoms across participants. When split by gender, females showed weak negative relationships between keystroke timing features and mental health symptoms, and weak positive relationships between keystroke frequency features and mental health symptoms. The opposite relationships were found for males (except for dwell). Machine learning models using keystroke features alone did not predict mental health symptoms.

Conclusions: Increased mental health symptoms are weakly associated with faster typing, with important gender differences. Keystroke metadata should be collected longitudinally and combined with other digital phenotypes to enhance their clinical relevance.

Trial Registration: Australian and New Zealand Clinical Trial Registry, ACTRN12619000855123; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=377664&isReview=true

(JMIR Ment Health 2023;10:e44986) doi:10.2196/44986

KEYWORDS

adolescents; anxiety; depression; digital phenotype; keystroke dynamics; keystroke metadata; smartphone; students

Introduction

Mental disorders are prevalent among adolescents worldwide and have significantly increased over recent years [1,2]. Approximately half of the mental disorders emerge before the age of 18 years [3], with more than 1 in 8 young people (ie, 4-17 years) meeting criteria for a diagnosis of a mental disorder in the last 12 months [1]. Despite the existence of evidence-based treatments, few adolescents who seek professional help receive effective treatment [4]. Of those who seek professional help, up to 50% do not respond to treatment [5]. To address the rising rates of youth mental illness, digital phenotyping (ie, using smartphone data to build a rich, personalized digital picture of behavior) has been proposed as a novel solution to enhance all aspects of clinical care, including diagnosis, symptom monitoring, and predicting response to treatment [6-8].

Among the digital phenotypes currently being developed, typing behavior, such as keystroke dynamics presents, is one of the most relevant. Keystroke dynamics refers to the detailed timing information captured when pressing and releasing of keys on a keyboard [9]. Although keystroke dynamics have historically been used as a behavioral biometric to verify individuals' identities for security purposes (eg, phone banking and security systems), they have more recently been leveraged from individuals' smartphones to remotely diagnose mental disorders and monitor mental health symptoms (for a review, see Alfalahi et al [10]). Typing behavior is especially relevant to adolescents, given smartphones are now almost universally owned and where typing behavior, such as instant messaging is among the most frequently used features [11-13].

Recent studies applying keystroke dynamics to mental health diagnostics and symptom monitoring have shown promising findings. For example, Cao et al [14] and Huang et al [15] first applied deep learning models to keystroke metadata to predict depression and mania symptoms in participants with bipolar disorder from the BiAffect study, yielding high predictive accuracies (~90%). Using the same sample, Stange et al [16] showed that an increased instability of typing speed (ie, root-mean-square successive difference) predicted higher future depression but not mania symptoms, whereas Zulueta, Zulueta et al [17] found that increased average interkey delay (ie, the average time between keystrokes) predicted higher depression but not mania symptoms. Mastoras et al [18] found a random forest classifier was the best-performing model among common machine learning models using various keystroke features to predict depression status. Vesel et al [19] used a larger mixed clinical sample to show higher depression was associated with more variable typing speed, and Bennett et al [20] found a random forest classifier best predicted change in depression severity (ie, change of 4 or more on the Patient Health Questionnaire [PHQ]) using keystroke features with high accuracy (ie, 90%). Together, these studies suggest that some keystroke dynamics features can provide useful information for predicting symptoms of depression and mania and diagnosing depression status.

However, these studies have several limitations. First, most of the studies have used small sample sizes (ie, n≤25 [14-18]). Second, extraction and application of keystroke timing and frequency features vary greatly between studies. Third, samples have mostly been limited to adults with bipolar disorder and outcome measures of depression and mania symptoms. Finally, all studies (except for Mastoras et al [18]) have used keystroke metadata collected during participants' routine use of their smartphone using the same platform (ie, BiAffect [17]), limiting the generalizability to other data collection methodologies (eg, task-derived metadata) and other platforms. Together, these limitations impact the validity, replicability, and translatability of the findings.

Recently, the Future Proofing study [21,22] has collected digital data from n>6000 adolescents in a high school setting to facilitate prevention and early intervention of depression. Initial findings suggest that the sample is broadly representative of the Australian adolescent population, with digital data actively and passively collected continuously throughout the study [22]. Although not all participants were involved in typing data collection, data from this study provides an excellent opportunity to explore keystroke features associated with mental health symptom severity in a large sample using various keystroke features across a range of mental health measures.

This paper aims to examine the associations between common keystroke timing and frequency features and mental health symptoms in adolescence using data from the Future Proofing study [21,22]. Keystroke typing features were derived from 2 typing tasks completed by 934 participants in the custom-built Future Proofing app within 2 weeks of their baseline assessment. It was hypothesized that increased severity in depression symptoms would be associated with slower keystrokes and less frequent keystrokes due to increased psychomotor impairment [17,18]. It was also hypothesized that machine learning models using keystroke features would significantly predict symptoms of depression [14,15,18,20]. Relationships between keystroke features and other mental health symptoms (ie, anxiety, distress, and insomnia) were exploratory.

Methods

Design

This study was a secondary analysis of baseline data from a prospective cohort study with an embedded cluster randomized controlled trial. The trial protocol [21] and baseline characteristics [22] have been published elsewhere. Ethics approvals were obtained from the University of New South Wales Human Research Ethics Committee (HC180836), the State Education Research Applications Process for the New South Wales Department of Education (SERAP2019201), and relevant Catholic Schools Dioceses across Australia.

Setting

The Future Proofing study was conducted in 134 secondary schools located across Australia, including government and nongovernment (independent and Catholic) schools. Recruitment was conducted from March 2019 to March 2022. All New South Wales government and independent secondary schools and

eligible New South Wales Catholic secondary schools were invited to participate. Independent schools in capital cities from around Australia were also invited to participate. Consent was provided on the internet from a parent or guardian and the adolescent before study participation. Data collection took place across 3 separate year 8 cohorts (students aged 13-14 years): the 2019 Cohort from August to September 2019; the 2020 Cohort from August to November 2020; the 2021 Cohort from April-December 2021 (extended to March 2022 due to COVID-19).

Participants

Participants (N=6388) were secondary school adolescents from year 8 who attended participating schools. All adolescents enrolled in year 8 at each participating school were eligible to participate in the trial if they had a smartphone with iOS or Android operating system and an active phone number. All participants were invited to undertake the typing tasks, as part of an omnibus of digital tasks [22]. Participants were limited to those who completed the typing speed tasks and had their digital data available for analysis (1208/6388, 18.9%) and those who completed the tasks within the first 2 weeks of completing their baseline questionnaires (ie, 934/1208, 77.3%; for distributions of days between baseline assessment and typing task completion, see Figure S1 in Multimedia Appendix 1).

Procedure

Participants with parental consent completed the study questionnaires through a secure web-based portal, accessible using their phone number and a 1-time password sent through SMS text messaging. Following the completion of baseline questionnaires, participants were instructed to open the Future Proofing app and complete the typing speed tasks during class time and in their own time.

Typing Task

Participants completed 2 typing tasks through the Future Proofing app—the prose task and the composition task. During the prose task, participants were asked to copy (type) as much text from a given script as possible within 30 seconds. For the composition task, participants were asked to type between 200 and 250 characters on a randomly given topic with no time constraints. Scripts and topics were randomly allocated to participants from a pool of 8 (see Figure S2 and Text S1 in Multimedia Appendix 1 for task screenshots, scripts, and topics).

Measures

Depressive Symptoms

The PHQ for Adolescents (PHQ-A) is a validated modification of the PHQ-9 for adolescents, a 9-item self-administered depression severity screening and diagnostic tool based on Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) criteria [23]. The scale assesses the frequency of occurrence of 9 depression symptom criteria during the previous 2 weeks, with items rated on a 4-point scale ranging from 0 “Not at all” to 3 “Nearly every day.” Total scale scores on the PHQ-A depression scale can range from 0 to 27, with higher scores reflecting more severe depressive symptoms. The accepted clinical cutoff points are as follows: a score of 0-4 indicates nil to minimal symptoms, 5-9 indicates mild symptoms, 10-14 indicates moderate symptoms, 15-19 indicates moderately severe symptoms and 20-27 indicates severe symptoms. A threshold of ≥15, reflecting moderately severe symptoms was used to determine caseness. The internal consistency of the PHQ-A in this study was high (α=.88 [22]).

Anxiety Symptoms

The Spence Children’s Anxiety Scale Short-Form (CAS-8) is an 8-item brief measure of anxiety for children and adolescents, based on the Spence Children’s Anxiety Scale [24]. The Spence Children’s Anxiety Scale was designed to measure the severity of children's and adolescents’ anxiety symptoms based broadly on Diagnostic and Statistical Manual of Mental Disorders, fourth edition criteria for anxiety disorders [25]. Respondents rate the degree to which they experience each symptom on a 4-point frequency scale, ranging from 0 “Never” to 3 “Always.” Total scale scores on the 8-item CAS-8 can range from 0 to 24, with higher scores reflecting greater anxiety. A threshold of ≥14 was used to determine caseness. The internal consistency of the CAS-8 in this study was high (α=.88 [22]).

Distress

The Distress Questionnaire-5 (DQ-5; [26]) is a 5-item brief screening tool for identifying general psychological distress. Respondents rate each item on a 5-point scale, ranging from 1 “Never” to 5 “Always.” The total scores on the scale range from 5 to 25, with higher scores indicating greater psychological distress. A threshold of ≥14 was used to determine high distress caseness. The internal consistency of the DQ-5 in this study was high (α=.88 [22]).

Insomnia

The Insomnia Severity Index (ISI) is a psychometrically sound, 7-item self-report measure of insomnia symptoms over the previous 2 weeks [27]. Responses are reported on a Likert scale ranging from 0 “Not at all” to 4 “Very,” producing total scores of 0-28. Cutoff scores are as follows: 0-7 reflects no clinically significant insomnia, 8-14 indicates subthreshold insomnia, 15-21 suggests moderate severity insomnia, and 22-28 indicates severe insomnia. A threshold of ≥15, reflecting moderately severe symptoms was used to determine high insomnia caseness. The ISI was designed for use in adults but has been widely administered to, and validated in, adolescent samples [28,29].

Preprocessing and Feature Extraction

Data were restricted to participants who completed their first composition and prose tasks within 2 weeks from their baseline questionnaires. In total, 54 features were extracted (for a full list, see Table S1 in Multimedia Appendix 1). For the keystroke timing features of dwell (ie, the time interval between a key press and release of the same key), latency (ie, the time interval between a key press of a keystroke and key release of the following keystroke), interval (ie, the time interval between a key press of a keystroke and key release of the following key), up up (ie, the time interval between the key release of a keystroke and key release of the following keystroke), and down down (ie, the time interval between key press of a keystroke and key press of the following keystroke), extracted features included the median, mean, variance, minimum, maximum, Q1,

Q3, and skewness, kurtosis. Keystroke frequency features include total keystrokes, total backspace, total spaces, total nonalphanumeric, and total null keys (ie, unlabeled or undetermined keys), as well as the proportion of total backspace, total spaces, total nonalphanumeric, and total null keys to total keystrokes. Features were extracted using metadata from both typing tasks, as well as separately from the task to investigate whether the task impacted the relationship between keystroke features and mental health symptoms.

Statistical Analysis

Pearson correlations were used to test associations between keystroke features and mental health symptoms. The false discovery rate of *P* values was adjusted to *q* values using the Benjamini and Hochberg [30] approach. Correlations were also tested separately by gender (restricted to males and females due to limited numbers in other categories) and likely clinical caseness to investigate whether these groups differed in their patterns of association.

Various machine learning models including linear and nonlinear support vector machines, neural nets, random forest, extreme gradient boosting, and elastic net models were used to test whether all keystroke features and demographic variables could be combined to predict mental health symptoms. Machine learning models were chosen based on their performance and use among the keystroke dynamics literature [18,20,31] and more widely in precision psychiatry [32,33]. Highly correlated features (r>0.95) were first removed from the 54 keystroke features, leaving 34 features. Data were then split into training (80%) and test (20%) sets so models developed using the training set could be validated through hold-out cross-validation in the test set. Hyperparameters were manually optimized using logarithmically spaced values between 2^{-10} and 2^{10} where appropriate (except for random forest models, where the number of features was set between 1:34) in the training set using 10-fold cross-validation with 20 repeats based on the root mean standard error with and without recursive feature elimination. The best-performing models were then validated on the test set. Model performance was assessed by correlating the predicted values with the observed values in the test sets. Models were also developed with the demographic variables gender and age included among the pool of features to determine whether demographic variables could improve model performance beyond keystroke features alone.

All analyses were performed using R 4.2.1 [34]. Correlation plots were created using the “corrplot” package in R [35]. Machine learning models were trained using the “caret” package in R [36].

Results

Participants

Participant characteristics are presented in Table 1. The mean age was 13.8 (SD 0.6, range 11.1-16.3) years and 64.0% (598/934) were female. The mean PHQ-A score was 7.9 (SD 6.5), with 16.8% (157/934) meeting the clinical cut-off for depression (for distributions of mental health scores, see Figure S3 in Multimedia Appendix 1). Latency had the longest duration (mean 0.30, SD 0.08) and dwell had the shortest duration (mean 0.08, SD 0.01) among keystroke timing features. The mean total keystrokes was 98.5 (SD 37.1) with few nonalphanumeric (mean 1.7, SD 1.4) or null keys (mean 0.4, SD 1.3) used. Most participants self-reported that they typed with both hands (72.7%; 679/934) and 34.4% (321/934) used iPhones.

Table 1. Demographic and clinical characteristics of participants.

Characteristics	Total sample (N=934)
Age (years), mean (SD)	13.84 (0.56)
Depressive symptoms (PHQ-A[a]), mean (SD)	7.87 (6.5)
Anxiety symptoms (CAS-8[b]), mean (SD)	9.11 (5.6)
Distress symptoms, mean (SD)	11.65 (5.1)
Insomnia symptoms, mean (SD)	7.62 (5.86)
Keystroke timing features (median), mean (SD)	
Dwell	0.08 (0.01)
Latency	0.30 (0.07)
Interval	0.14 (0.06)
Up up	0.21 (0.07)
Down down	0.22 (0.07)
Keystroke frequency features (total), mean (SD)	
Keystrokes	337.97 (168.54)
Spaces	57.15 (31.93)
Backspaces	42.01 (36.75)
Nonalphanumeric	4.66 (7.39)
Null keys	0.92 (3.55)
Gender (female), n (%)	598 (64)
Typing hand, n (%)	
Left	8 (0.9)
Right	60 (6.4)
Both	679 (72.7)
Not reported	187 (20)
Device, n (%)	
iOS	321 (34.4)
Android	98 (10.5)
Not reported	515 (55.1)
Likely clinical case, n (%)	
Depressive symptoms (PHQ-A) (≥15)	157 (16.8)
Anxiety symptoms (CAS-8) (≥14)	208 (22.3)
Distress symptoms (DQ-5[c]) (≥14)	317 (34)
Insomnia symptoms (ISI[d]) (≥15)	133 (14.3)

[a]PHQ-A: Patient Health Questionnaire for Adolescents.

[b]CAS-8: Children's Anxiety Scale-Short Form.

[c]DQ-5: Distress Questionnaire-5.

[d]ISI: Insomnia Severity Index.

Associations Between Keystroke Features and Mental Health Symptoms

Associations between mental health symptoms and commonly reported keystroke features (ie, median timing and total frequency features) are presented in Figure 1 (for all features, see Figures S4 and S5 in Multimedia Appendix 1). Keystroke metadata was combined across tasks and prompts due to similar patterns of association being observed when assessing associations separately by task (see Figure S6 in Multimedia Appendix 1) and by emotionally valanced prompts (see Figure S7 in Multimedia Appendix 1). All measures of mental health symptoms (ie, depression, anxiety, distress, and insomnia) showed weak negative associations with all keystroke timing

features, indicating that increased mental health symptoms were associated with faster typing. However, after correcting for multiple comparisons, only the association between distress and latency remained significant (r=0.07, q=0.045). Measures of mental health symptoms showed a mix of weak positive and negative associations with keystroke frequency features, although none of these relationships were significant (all q>0.05).

Associations between mental health symptoms and keystroke features by gender are presented in Figure 2 (for all features, see Figures S8-S11 in Multimedia Appendix 1). In females, a similar pattern emerged between mental health symptoms and keystroking timing features, whereby increased mental health symptoms were associated with faster typing. Generally, positive associations were also seen between mental health symptoms and keystroke frequency features, whereby increased mental health symptoms were associated with more frequent keystrokes. In males, the opposite patterns emerged with mental health symptoms showing a weak positive association with keystroke timing features (except for dwell) and weak negative associations with timing features. However, all associations were nonsignificant after correcting for multiple comparisons (all q>0.05).

Associations between mental health symptoms and keystroke features by likely clinical caseness are presented in Figure 3 (for all features, see Figures S12-S16 in Multimedia Appendix 1). In people who did not meet PHQ-A caseness, all measures of mental health symptoms showed weak negative associations with all keystroke timing features, whereas people who did meet PHQ-A caseness showed the opposite pattern (except for depression). All measures of mental health symptoms generally showed weak positive associations with all keystroke frequency features in both people who did and did not meet PHQ-A caseness (with the exception of null keys). However, all associations were nonsignificant after correcting for multiple comparisons (all q>0.05).

In people who did not meet CAS-8 caseness, all measures of mental health symptoms showed weak negative associations with all keystroke timing features, with significant associations between distress and dwell (r=−0.08, q=0.048), latency (r=−0.09, q=0.026), and down down (r=–0.08, q=0.045), as well as insomnia and dwell (r=−0.08, q=0.042) and latency (r=−0.09, q=0.037). People who did meet CAS-8 caseness showed the opposite pattern (except for depression). All measures of mental health symptoms generally showed weak positive associations with all keystroke frequency features in people who did not meet CAS-8 caseness, whereas the opposite pattern was seen in people who did meet CAS-8 caseness. However, these associations were nonsignificant after correcting for multiple comparisons (all q>0.05).

In people who did not meet DQ-5 caseness, all measures of mental health symptoms showed weak negative associations with all keystroke features. In people who did meet DQ-5 caseness, all measures of mental health symptoms showed weak positive associations with all keystroke timing features and weak negative associations between mental health symptoms (except for insomnia) with keystroke frequency features. However, all associations were nonsignificant after correcting for multiple comparisons (all q>0.05).

In people who did not meet criteria for ISI caseness, all measures of mental health symptoms showed weak negative associations with all keystroke timing features, with significant associations between distress and latency (r=−0.09, q=0.023) and down down (r=−0.08, q=0.046), as well as insomnia and latency (r=−0.08, q=0.035). People who did meet ISI caseness showed the opposite pattern (except for depression). Both people who did and did not meet criteria for ISI caseness showed a mix of weak positive and negative associations between mental health symptoms and keystroke frequency features. However, these associations were nonsignificant after correcting for multiple comparisons (all q>0.05).

Figure 1. Associations between mental health symptoms and common keystroke features.

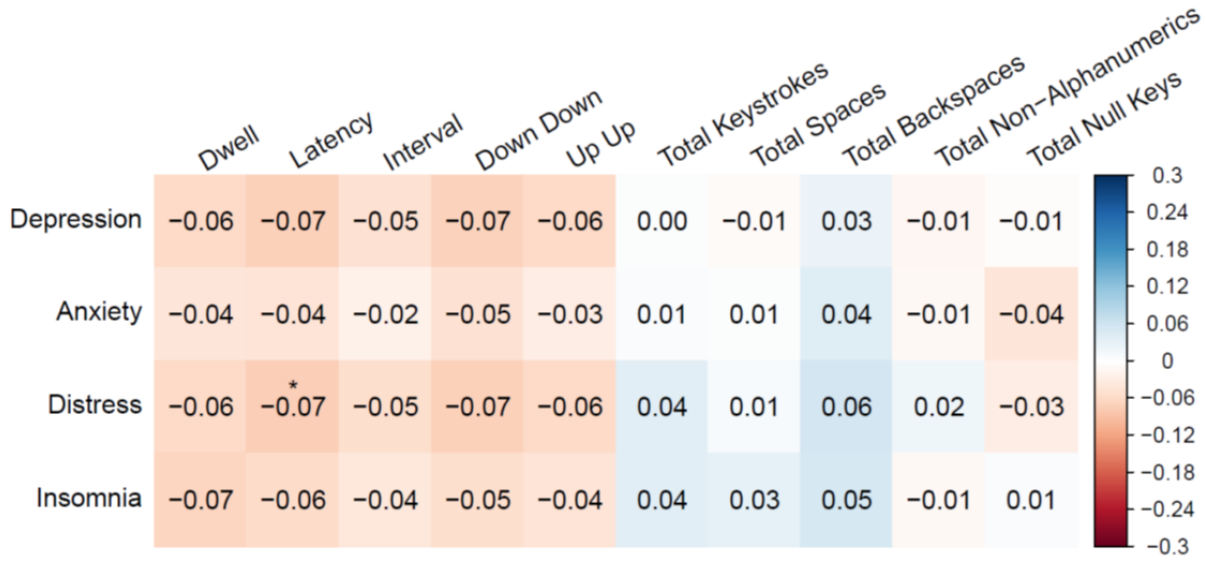


Figure 2. Associations between mental health symptoms and keystroke features for females and males.

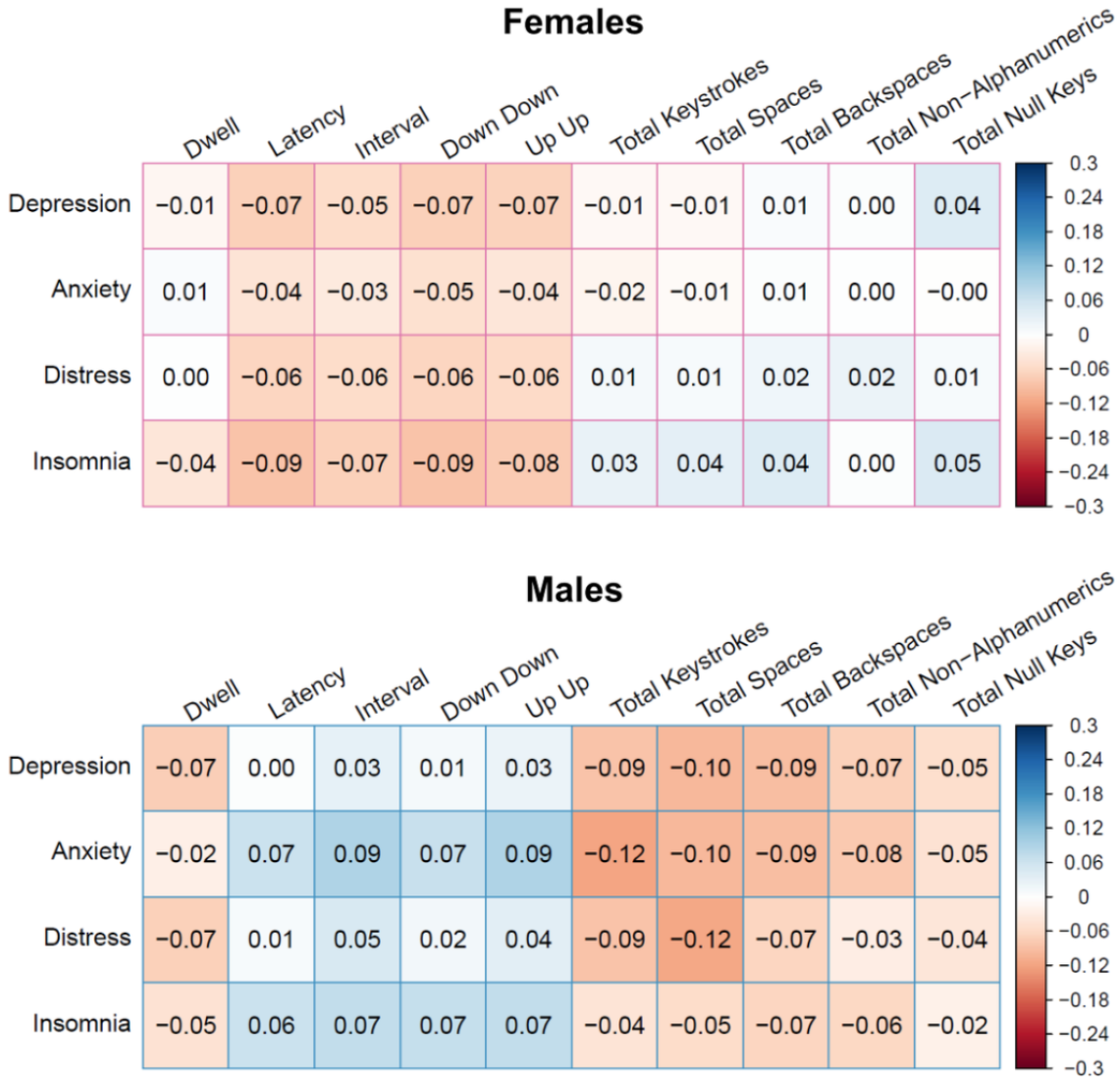
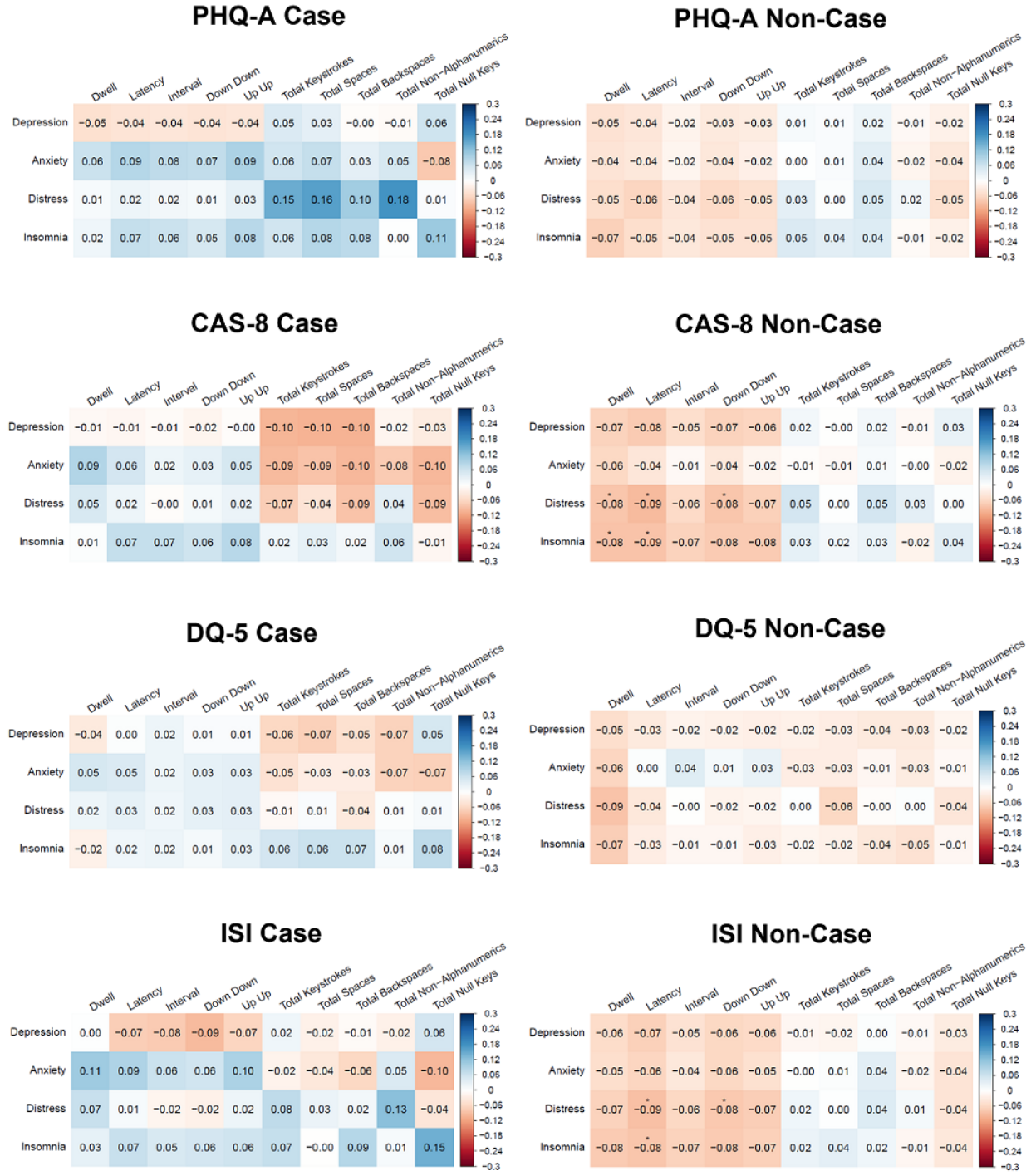


Figure 3. Associations between mental health symptoms and common keystroke features for people who do and do not meet likely clinical caseness. CAS-8: Children's Anxiety Scale-Short Form; DQ-5: Distress Questionnaire-5; ISI: Insomnia Severity Index; PHQ-A: Patient Health Questionnaire for Adolescents.



Keystroke Features Predicting Mental Health Symptoms

Results from machine learning models using keystroke features and demographics to predict mental health symptoms are presented in Table 2. The best-performing models for predicting mental health symptoms were elastic net models. No significant models could be produced using keystroke features alone (all $P>.05$). When using keystroke features and gender, predicted and observed values were correlated for depression ($r=0.30$, $P<.001$), anxiety ($r=0.38$, $P<.001$), distress ($r=.33$, $P<.001$), and insomnia ($r=0.19$, $P=.01$) in the test sets. However, keystroke features were minimally weighted compared to gender (see Figure S17 in Multimedia Appendix 1). Models were also developed using additional multiplicative features between gender and keystroke features to test whether including these additional features would improve prediction accuracy, but these models were similarly poor performing. No significant models could be produced when developed separately by gender (all $P>.05$).

Table 2. Results from machine learning models where predicted values significantly correlated with observed values in test sets to predict mental health symptoms.

Mental health symptoms	Model	Hyperparameters, α^a, λ^b	Performance metrics		Correlation between observed and predicted test values	
			RMSE[c]	MAE[d]	*r*	*P* value
Depression	Elastic Net	1, 0.50	6.22	5.10	0.30	<.001
Anxiety	Elastic Net	8, 0.25	5.23	4.25	0.38	<.001
Distress	Elastic Net	1, 0.25	4.75	3.93	0.33	<.001
Insomnia	Elastic Net	125, 0.5	5.68	4.65	0.19	.01

[a]α: alpha.

[b]λ: lambda.

[c]RMSE: root-mean-square error.

[d]MAE: mean absolute error.

Supplemental Analysis of Total Keystrokes and Total Spaces Split by Task

Although most keystroke features showed similar patterns across tasks, we performed additional supplemental analyses of the frequency features "total keystrokes" and "total spaces" due to their stronger associations with mental health symptoms in the prose task when compared to the composition task. Although females show similar patterns of association between composition and prose tasks for total keystrokes and total spaces, males showed different patterns. However, none of the associations split by gender and task were significant (see Figure S18 in Multimedia Appendix 1). Noncases showed similar patterns of association between composition and prose tasks for total keystrokes and total spaces. However, cases showed different patterns between tasks, with a significant association between insomnia and total spaces for the composition task (r=0.22, q=0.032). No other associations split by caseness and task were significant (see Figure S19 in Multimedia Appendix 1).

Discussion

Overview

This study investigated associations between common keystroke timing and frequency features and mental health symptoms in a large cohort of adolescents. We found increased mental health symptoms (ie, depression, anxiety, distress, and insomnia) were weakly associated with faster keystroke timing features. When split by gender, females with higher mental health symptoms exhibited faster keystroke timing features and fewer keystroke frequency features, whereas males showed opposite patterns. When split by likely clinical caseness, various patterns of association were found. However, most associations were no longer significant following corrections to multiple comparisons. Machine learning models using keystroke features alone also did not predict mental health symptoms and contributed minimally when models included gender.

Few studies have directly assessed correlations between keystroke features and mental health symptoms [16,20]. Of those studies, only Bennett et al [20] provided a correlation matrix assessing associations between a range of keystroke features and depression severity (measured using the PHQ-8) and similarly found a mix of weak positive and negative associations (r=0.16 to 0.10). Our finding of increased mental health symptoms being associated with faster keystroke timing features is in line with suggestions that increased psychomotor agitation could lead to a general speeding up of behaviors, including typing speed [17]. However, increased mental health symptoms, such as depression can also lead to psychomotor impairment, which will produce slower typing speeds. The diverse symptom profiles experienced by people with mental disorders and differences in associations between various subgroups seen in this study highlight the need for more personalized approaches to assessment and treatment [6,7,37-39]. The relative weakness of cross-sectional associations between keystroke features and mental health symptoms also highlights the importance of needing to combine various digital phenotyping features in prediction models.

Previous studies leveraging keystroke features to predict mental health symptoms using machine learning and deep learning models have all yielded high predictive accuracies (~90%) [14,15,18,20]. The poor predictive performance yielded from the models in our study could be attributed to several factors. First, all previous studies (except for Bennett et al [20]) used small sample sizes (ie, n≤25 [14,15,18]). And while Bennett et al [20] used a larger sample size with methods to rebalance the data, no previous study used independent data sets to evaluate their models performance, which can strongly bias performance estimates [40]. Second, our models used cross-sectional task-derived keystroke metadata and measures of mental health symptoms, whereas previous studies have used session-level keystroke data collected during routine use to predict future mental health symptoms [14,15,18,20], which provides a richer pool of predictive features. Third, most studies dichotomized mental health symptoms rather than treating the measures as continuous (eg, Hamilton Depression Rating Scale scores ≤7 as negative depression and >7 as positive [14]), losing important information and ignoring the variation within each group [41]. Lastly, most studies leveraged data sets made up of a single clinical population [14,15,18], increasing uniformity and impacting generalizability. Together, our findings suggest that keystroke features provide minimal use in predicting mental health symptoms cross-sectionally.

Limitations

This study includes several limitations. Keystroke metadata was collected from typing tasks completed within 2 weeks of participants' baseline assessments and metadata collected temporally closer to their baseline assessment may provide more accurate reflections of mental health symptoms. However, most participants completed the typing task on the same day as their baseline assessments (see Figure S1 in Multimedia Appendix 1). Metadata related to the distance between keys was not collected, so additional features previously calculated (eg, speed [18]) could not be replicated. We combine metadata across typing tasks and prompts. Although patterns of association were generally similar across tasks and prompts, some differences in patterns of association were observed for total keystrokes and total spaces when split by task. Future studies should consider how these and other methodological considerations (eg, text length and typing accuracy) may impact associations between keystroke features and mental health symptoms. As the Future Proofing study is still being completed, longitudinal typing and mental health data were not yet available for analysis.

Conclusions

In conclusion, our findings suggest keystroke features are weakly associated with mental health symptoms with differing patterns for males and females. Keystroke features also provide minimal clinical use in predicting mental health symptoms cross-sectionally. Future research should focus on collecting keystroke metadata longitudinally and combining these data with other digital phenotypes being developed to enhance their potential for aiding clinical care.

Acknowledgments

The authors would like to thank the young people, their families, and school communities who are participating in this study. The authors would also like to thank the well-being staff from our partner schools, including deputy principals, year advisors, school counselors, and student support officers among many others. This study would not have been possible without the committed Black Dog Institute volunteers who have given up their time to support and facilitate the implementation of this study. Finally, we would like to thank the Black Dog Institute Education, Marketing and Communications, and IT teams, and the Deakin University team. This project was funded by a NHMRC project grant awarded to Helen Christensen (GNT1138405), a Ramsay Health Philanthropic Grant awarded to Helen Christensen and Aliza Werner-Seidler, a donation made by General Pants Co, a NHMRC Emerging Leader Fellowship awarded to Aliza Werner-Seidler (GNT1197074), and a SPRF NHMRC Fellowship to Helen Christensen (GNT 1155614).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary text, figures, and tables.
[DOCX File , 5557 KB - mental_v10i1e44986_app1.docx]

References

1. Lawrence D, Johnson S, Hafekost J, de Haan KB, Sawyer M, Ainley J, et al. The Mental Health of Children and Adolescents: Report on the Second Australian Child and Adolescent Survey of Mental Health and Wellbeing. Canberra: Department of Health (Australia); 2015.
2. Shorey S, Ng ED, Wong CHJ. Global prevalence of depression and elevated depressive symptoms among adolescents: a systematic review and meta-analysis. Br J Clin Psychol 2022;61(2):287-305. [doi: 10.1111/bjc.12333] [Medline: 34569066]
3. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022;27(1):281-295 [FREE Full text] [doi: 10.1038/s41380-021-01161-7] [Medline: 34079068]
4. Harvey AG, Gumport NB. Evidence-based psychological treatments for mental disorders: modifiable barriers to access and possible solutions. Behav Res Ther 2015;68:1-12. [doi: 10.1016/j.brat.2015.02.004] [Medline: 25768982]
5. Curry J, Silva S, Rohde P, Ginsburg G, Kratochvil C, Simons A, et al. Recovery and recurrence following treatment for adolescent major depression. Arch Gen Psychiatry 2011;68(3):263-269 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2010.150] [Medline: 21041606]
6. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA 2017;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
7. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
8. Braund TA, Zin MT, Boonstra TW, Wong QJJ, Larsen ME, Christensen H, et al. Smartphone sensor data for identifying and monitoring symptoms of mood disorders: a longitudinal observational study. JMIR Ment Health 2022;9(5):e35549 [FREE Full text] [doi: 10.2196/35549] [Medline: 35507385]

9. Kołakowska A, Szwoch W, Szwoch M. A review of emotion recognition methods based on data acquired via smartphone sensors. Sensors (Basel) 2020;20(21):6367 [FREE Full text] [doi: 10.3390/s20216367] [Medline: 33171646]
10. Alfalahi H, Khandoker AH, Chowdhury N, Iakovakis D, Dias SB, Chaudhuri KR, et al. Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic review and meta-analysis. Sci Rep 2022;12(1):7690 [FREE Full text] [doi: 10.1038/s41598-022-11865-7] [Medline: 35546606]
11. Smith A. US smartphone use in 2015. Pew Research Center. 2015. URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2023-04-21]
12. Levine M. Australian Teenagers and their Mobile Phones. Roy Morgan. 2016. URL: http://www.roymorgan.com/findings/6929-australian-teenagers-and-their-mobile-phones-june-2016-201608220922#:~:text=Just%20over%20one%20million%20Australian,' [accessed 2023-04-21]
13. Mehner M. WhatsApp, WeChat, Facebook Messenger, and more: global usage of messaging apps and statistics. Sinch Engage. URL: https://engage.sinch.com/blog/global-messenger-apps-usage-statistics/#Global [accessed 2023-04-29]
14. Cao B, Zheng L, Zhang C, Yu PS, Piscitello A, Zulueta J, et al. DeepMood: modeling mobile phone typing dynamics for mood detection. 2017 Presented at: KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13 - 17, 2017; Halifax NS, Canada p. 747. [doi: 10.1145/3097983.3098086]
15. Huang H, Cao B, Yu PS, Wang CD, Leow AD. dpMood: exploiting local and periodic typing dynamics for personalized mood prediction. 2018 Presented at: 2018 IEEE International Conference on Data Mining (ICDM); November 17-20, 2018; Singapore p. 157-166. [doi: 10.1109/icdm.2018.00031]
16. Stange JP, Zulueta J, Langenecker SA, Ryan KA, Piscitello A, Duffecy J, et al. Let your fingers do the talking: passive typing instability predicts future mood outcomes. Bipolar Disord 2018;20(3):285-288 [FREE Full text] [doi: 10.1111/bdi.12637] [Medline: 29516666]
17. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a BiAffect digital phenotyping study. J Med Internet Res 2018;20(7):e241 [FREE Full text] [doi: 10.2196/jmir.9775] [Medline: 30030209]
18. Mastoras R, Iakovakis D, Hadjidimitriou S, Charisis V, Kassie S, Alsaadi T, et al. Touchscreen typing pattern analysis for remote detection of the depressive tendency. Sci Rep 2019;9(1):13414 [FREE Full text] [doi: 10.1038/s41598-019-50002-9] [Medline: 31527640]
19. Vesel C, Rashidisabet H, Zulueta J, Stange JP, Duffecy J, Hussain F, et al. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: A BiAffect iOS study. J Am Med Inform Assoc 2020;27(7):1007-1018. [doi: 10.1093/jamia/ocaa057] [Medline: 32467973]
20. Bennett CC, Ross MK, Baek E, Kim D, Leow AD. Predicting clinically relevant changes in bipolar disorder outside the clinic walls based on pervasive technology interactions via smartphone typing dynamics. Pervasive Mob Comput 2022;83:101598. [doi: 10.1016/j.pmcj.2022.101598]
21. Werner-Seidler A, Huckvale K, Larsen ME, Calear AL, Maston K, Johnston L, et al. A trial protocol for the effectiveness of digital interventions for preventing depression in adolescents: the future proofing study. Trials 2020;21(1):2 [FREE Full text] [doi: 10.1186/s13063-019-3901-7] [Medline: 31898512]
22. Werner-Seidler A, Maston K, Calear AL, Batterham PJ, Larsen ME, Torok M, et al. The future proofing study: design, methods and baseline characteristics of a prospective cohort study of the mental health of Australian adolescents. Int J Methods Psychiatr Res 2022:e1954. [doi: 10.1002/mpr.1954] [Medline: 36444163]
23. Johnson JG, Harris ES, Spitzer RL, Williams JBW. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolesc Health 2002;30(3):196-204. [Medline: 11869927]
24. Spence SH, Barrett PM, Turner CM. Psychometric properties of the Spence Children's Anxiety Scale with young adolescents. J Anxiety Disord 2003;17(6):605-625. [Medline: 14624814]
25. Spence SH. A measure of anxiety symptoms among children. Behav Res Ther 1998;36(5):545-566. [Medline: 9648330]
26. Batterham PJ, Sunderland M, Carragher N, Calear AL, Mackinnon AJ, Slade T. The distress questionnaire-5: population screener for psychological distress was more accurate than the K6/K10. J Clin Epidemiol 2016;71:35-42. [doi: 10.1016/j.jclinepi.2015.10.005] [Medline: 26464194]
27. Bastien CH, Vallières A, Morin CM. Validation of the insomnia severity index as an outcome measure for insomnia research. Sleep Med 2001;2(4):297-307. [Medline: 11438246]
28. Alvaro PK, Roberts RM, Harris JK. The independent relationships between insomnia, depression, subtypes of anxiety, and chronotype during adolescence. Sleep Med 2014;15(8):934-941. [doi: 10.1016/j.sleep.2014.03.019] [Medline: 24958244]
29. Chung K, Kan KK, Yeung W. Assessing insomnia in adolescents: comparison of insomnia severity index, athens insomnia scale and sleep quality index. Sleep Med 2011;12(5):463-470. [doi: 10.1016/j.sleep.2010.09.019] [Medline: 21493134]
30. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 2018;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
31. Zulueta J, Demos AP, Vesel C, Ross M, Piscitello A, Hussain F, et al. The effects of bipolar disorder risk on a mobile phone keystroke dynamics based biomarker of brain age. Front Psychiatry 2021;12:739022 [FREE Full text] [doi: 10.3389/fpsyt.2021.739022] [Medline: 35002792]

32. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018;3(3):223-230. [doi: 10.1016/j.bpsc.2017.11.007] [Medline: 29486863]
33. Braund TA, Breukelaar IA, Griffiths K, Tillman G, Palmer DM, Bryant R, et al. Intrinsic functional connectomes characterize neuroticism in major depressive disorder and predict antidepressant treatment outcomes. Biol Psychiatry Cogn Neurosci Neuroimaging 2022;7(3):276-284. [doi: 10.1016/j.bpsc.2021.07.010] [Medline: 34363999]
34. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2022. URL: https://www.R-project.org/ [accessed 2023-04-21]
35. Wei T, Simko V. Visualization of a correlation matrix (Version 0.92). Package 'corrplot'. 2021. URL: https://github.com/taiyun/corrplot [accessed 2023-04-21]
36. Kuhn M. caret: classification and regression training. R package version 6.0-93. 2022. URL: https://CRAN.R-project.org/package=caret [accessed 2023-04-21]
37. Zimmerman M, Ellison W, Young D, Chelminski I, Dalrymple K. How many different ways do patients meet the diagnostic criteria for major depressive disorder? Compr Psychiatry 2015;56:29-34. [doi: 10.1016/j.comppsych.2014.09.007] [Medline: 25266848]
38. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. Proc Natl Acad Sci U S A 2018;115(27):E6106-E6115 [FREE Full text] [doi: 10.1073/pnas.1711978115] [Medline: 29915059]
39. Braund TA, Palmer DM, Williams LM, Harris AW. Characterising anxiety in major depressive disorder and its use in predicting antidepressant treatment outcome: An iSPOT-D report. Aust N Z J Psychiatry 2019;53(8):782-793. [doi: 10.1177/0004867419835933] [Medline: 30880405]
40. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One 2019;14(11):e0224365 [FREE Full text] [doi: 10.1371/journal.pone.0224365] [Medline: 31697686]
41. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332(7549):1080 [FREE Full text] [doi: 10.1136/bmj.332.7549.1080] [Medline: 16675816]

Abbreviations

CAS-8: eight-item Children's Anxiety Scale
DQ-5: Distress Questionnaire-5
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, fourth edition
ISI: Insomnia Severity Index
PHQ-A: Patient Health Questionnaire for Adolescents

Edited by J Torous; submitted 13.12.22; peer-reviewed by J Zulueta, D Zarate; comments to author 09.01.23; revised version received 24.02.23; accepted 22.03.23; published 15.05.23.

Please cite as:
Braund TA, O'Dea B, Bal D, Maston K, Larsen M, Werner-Seidler A, Tillman G, Christensen H
Associations Between Smartphone Keystroke Metadata and Mental Health Symptoms in Adolescents: Findings From the Future Proofing Study
JMIR Ment Health 2023;10:e44986
URL: https://mental.jmir.org/2023/1/e44986
doi:10.2196/44986
PMID:37184904

Original Paper

Effects of a Self-Guided Transdiagnostic Smartphone App on Patient Empowerment and Mental Health: Randomized Controlled Trial

André Kerber[1], Dipl Psych; Ina Beintner[2], PhD; Sebastian Burchert[1], Dipl Psych; Christine Knaevelsrud[1], Prof Dr

[1]Department of Clinical-Psychological Intervention, Freie Universität Berlin, Berlin, Germany

[2]MindDoc Health GmbH, Munich, Germany

Corresponding Author:
André Kerber, Dipl Psych
Department of Clinical-Psychological Intervention
Freie Universität Berlin
Habelschwerdter Allee 45
Berlin, 14195
Germany
Phone: 49 30 838 63093
Email: andre.kerber@fu-berlin.de

Abstract

Background: Mental disorders impact both individuals and health systems. Symptoms and syndromes often remain undetected and untreated, resulting in chronification. Besides limited health care resources, within-person barriers such as the lack of trust in professionals, the fear of stigmatization, or the desire to cope with problems without professional help contribute to the treatment gap. Self-guided mental health apps may support treatment seeking by reducing within-person barriers and facilitating mental health literacy. Digital mental health interventions may also improve mental health related self-management skills and contribute to symptom reduction and the improvement of quality of life.

Objective: This study aims to investigate the effects of a self-guided transdiagnostic app for mental health on help seeking, reduced stigma, mental health literacy, self-management skills, mental health symptoms, and quality of life using a randomized controlled design.

Methods: Overall, 1045 participants (recruited via open, blinded, and web-based recruitment) with mild to moderate depression or anxiety-, sleep-, eating-, or somatization-related psychopathology were randomized to receive either access to a self-guided transdiagnostic mental health app (MindDoc) in addition to care as usual or care as usual only. The core features of the app were regular self-monitoring, automated feedback, and psychological courses and exercises. The coprimary outcomes were mental health literacy, mental health–related patient empowerment and self-management skills (MHPSS), attitudes toward help seeking, and actual mental health service use. The secondary outcomes were psychopathological symptom burden and quality of life. Data were collected at baseline and 8 weeks and 6 months after randomization. Treatment effects were investigated using analyses of covariance, including baseline variables as predictors and applying multiple imputation.

Results: We found small but robust between-group effects for MHPSS (Cohen *d*=0.29), symptoms burden (Cohen *d*=0.28), and quality of life (Cohen *d*=0.19) 8 weeks after randomization. The effects on MHPSS were maintained at follow-up. Follow-up assessments also showed robust effects on mental health literacy and preliminary evidence for the improvement of help seeking. Predictors of attrition were lower age and higher personality dysfunction. Among the non-attritors, predictors for deterioration were less outpatient treatment and higher initial symptom severity.

Conclusions: A self-guided transdiagnostic mental health app can contribute to lasting improvements in patient empowerment. Symptoms of common mental disorders and quality of life improved faster in the intervention group than in the control group. Therefore, such interventions may support individuals with symptoms of 1 or more internalizing disorders, develop health-centered coping skills, prevent chronification, and accelerate symptom improvement. Although the effects for individual users are small and predictors of attrition and deterioration need to be investigated further, the potential public health impact of a self-guided intervention can be large, given its high scalability.

Trial Registration: German Clinical Trials Register DRKS00022531; https://drks.de/search/de/trial/DRKS00022531

(JMIR Ment Health 2023;10:e45068) doi:10.2196/45068

KEYWORDS

patient empowerment; mental health–related self-management skills; help-seeking attitude; mental health literacy; internet-based interventions; unguided; self-guided; transdiagnostic mental health app

Introduction

Mental Health Problems Come With High Costs for Individuals and Society

More than 300 million people have depression and more than 250 million people are affected by an anxiety disorder around the world [1]. Population-wide studies in Germany have revealed that approximately 1 in 10 people meets the diagnostic criteria for depression. Approximately 1 in 5 lives with an anxiety disorder, and 1 in 20 people deals with chronic pain [2] or insomnia [3]. Depression and anxiety are among the top 10 contributors to health loss, accounting for 7.5% and 4.5% of years lived with a disability, respectively [1,4]. Mental disorders result in high direct and indirect costs, estimated at more than €60 (US $63) billion per year in Germany alone [5,6].

Digital Mental Health Interventions May Reduce Symptom Burden

The efficacy of digital interventions has been shown for both the reduction of symptoms of common mental disorders and improvement of quality of life [7]. Most digital interventions address singular disorder categories and are structured similarly to disorder-specific treatment manuals [8]. These interventions consist of several subsequent sessions or modules. Interventions targeting the same disorder tend to be similar in terms of their components and content.

Guided digital interventions often yield larger effects than self-guided interventions (eg, the studies by Koelen et al [9], Lakhtakia and Torous [10], Moshe et al [11], and Schröder et al [12]), but a recent meta-analysis revealed that overall, the difference in effect sizes may not be very substantial [13], at least concerning anxiety disorders. Besides within-intervention guidance, guidance-related aspects of the study design such as the conduct of clinical interviews [14], use of automated reminders [15], or treatment setting in which the intervention is used [16] seem to be associated with greater effects. Nevertheless, although the scalability of web-based interventions with guidance is higher than that of traditional psychotherapy, it is limited because staff resources are required, even if the time allotted for guidance is limited. Self-guided interventions, although associated with only small to medium effects on individuals, can reach a larger number of patients and may thus have a larger public health impact [17].

Comorbidity and Transdiagnostic Interventions

Comorbidity among mental disorders is high and has been shown to be largely associated with common causal pathways in several large studies [18,19]. Therefore, contemporary models of psychopathology postulate dimensional spectra (eg, internalizing, thought disorder, and externalizing) comprising multiple mental health syndromes with co-occurring genetic, neurobiological, environmental, and behavioral indicators [20]. Many pharmacological and psychological treatments yield transdiagnostic effects on multiple mental disorders [21-23]. Therefore, recent treatment protocols for mental disorders, such as the Unified Protocol [24] and the Common Elements Treatment Approach [25], replace disorder- and symptom-specific interventions with interventions that address common causal factors and have been shown to be transdiagnostically effective.

Most people with mental illness in Germany receive care exclusively from primary care physicians; 83% of patients with affective disorders (F3, International Classification of Diseases, 10th Revision) and 91% of patients with neurotic, stress, and somatoform disorders (F4, International Classification of Diseases, 10th Revision) do not receive treatment from a mental health specialist [26]. Differentiating between different mental disorders can be a challenge for primary care providers, especially given the high rate of comorbidity. The presence of emotional problems is recognized in most cases during contact with a primary care provider, but an accurate diagnosis is made much less frequently [27,28]. Thus, low-threshold transdiagnostic interventions may be more suitable than disorder-specific interventions in a primary care setting.

Digital interventions with a transdiagnostic approach have yet been less well researched than disorder-specific interventions, but the evidence base is growing (eg, the studies by Newby et al [29], Newby et al [30], and Păsărelu et al [31]). However, based on the findings discussed earlier, we expected a transdiagnostic digital intervention to impact a range of mental disorder symptoms, including anxiety and depression, as well as quality of life.

Digital Mental Health Interventions May Reduce Treatment Barriers

Timely treatment for mental disorders is impeded by structural barriers, such as limited availability and high cost, but within-person attitudinal barriers may constitute an even stronger obstacle for treatment seeking. Wanting to handle the problem on one's own, low perceived need for care, stigma, low knowledge about mental health services, and fear about the act of help seeking or the source of help itself have been shown to be the largest treatment barriers by far [32-34]. These factors reduce the chances of timely intervention and increase the risk of long-term symptom deterioration and chronification [35]. Low-threshold digital interventions that involve no personal contact and can be used anonymously can counteract at least some of these internal barriers and simplify help seeking.

An underinvestigated research area concerns the reduction of within-person barriers through digital mental health interventions. Some older studies evaluating digital mental health interventions for depression and anxiety have shown a decrease in self-stigmatization [36-38]. Effects on help-seeking attitudes and actual help seeking have been detected in some

randomized controlled trials (RCTs) [38-41]. These effects seem to be linked to changes in health literacy [36,38]. In addition, a recent review on digital interventions found an increase in self-management behavior to be an important mediator of treatment effects [42]. However, research on the effects of transdiagnostic interventions on attitudinal barriers, help seeking, and self-management skills is lacking. A major advantage of transdiagnostic digital interventions in this respect is their potential to be truly low threshold. Unlike disorder-specific interventions, transdiagnostic interventions can be applied before diagnosis and, in a first step, help users determine whether they have a mental health problem at all, what their problem is, and whether they may need help from a mental health professional. Help seeking can then be actively encouraged by providing information and correcting unhelpful and false assumptions about mental health care.

This Study

On the basis of the findings from previous research, we expected that the use of a self-guided transdiagnostic self-management app for mental health, in addition to care as usual (CAU), would lead to significant improvements in mental health literacy and variables that reflect patient empowerment, such as help seeking, reduced stigma, and self-management skills. If self-guided mental health apps have these effects, they would constitute a low-cost public health impact if made available for people with mental health problems in addition to CAU. Furthermore, we aimed to explore whether such an intervention leads to a greater reduction in symptoms of common mental disorders and a stronger improvement in quality of life than CAU only.

Thus, the aim of our study was to investigate whether the use of a self-guided transdiagnostic app for mental health is associated with improvements in mental health literacy and variables that reflect patient empowerment, such as help seeking, reduced stigma, and self-management skills. Furthermore, the intervention's effects on symptoms of common mental disorders and quality of life were explored.

Methods

Ethical Considerations

This trial was registered in the German Clinical Trials Register (DRKS00022531**)**, and the local ethical committee of Freie Universität Berlin approved the protocol (AZ 039/2020).

Design

To examine the effects of a transdiagnostic mental health app (MindDoc), we conducted a single-center RCT with 3 assessments [43]. We assigned participants to 2 groups in a 1:1 ratio. The intervention group (IG) received immediate access to the MindDoc app in addition to current care (CAU); however, a limitation exists in that we recruited participants without outpatient or inpatient psychotherapy at the start of the trial (see *Recruitment Strategy* section). In the control condition, participants were not given any guidance or encouragement to modify their current care and were informed that they had the option to receive access to the MindDoc app after the 6-month study period. Essentially, this created a waitlist control condition for the use of the MindDoc app. However, it is important to note that participants were allowed to use any treatments that were accessible to them during the trial.

Health literacy, patient empowerment, help-seeking attitudes, health service use, symptom distress, and quality of life were assessed before randomization (baseline assessment), 8 weeks after randomization (postintervention assessment), and 6 months after randomization (follow-up assessment).

The participants in the control group (CG) received access to the MindDoc app after completing their follow-up assessment.

Intervention

The users in the IG received immediate access to the MindDoc app. The MindDoc app is a self-guided transdiagnostic intervention designed for individuals who want to take care of their mental health. It can be used across the mental health care spectrum, including for (indicated) prevention, early recognition, treatment, and aftercare. The core features of the app encompass regular self-monitoring and automated feedback and psychological courses and exercises.

The self-monitoring feature consists of an adaptive system of daily multiple-choice questions based on the Hierarchical Taxonomy of Psychopathology (HiTOP) [20] as well as regular mood ratings. On the basis of their entries in the self-monitoring feature, users receive regular automated feedback on relevant symptoms, problem areas, and in-app resources. Users also receive biweekly feedback on their overall mental health as well as encouragement to seek help depending on their health status.

The courses and exercises in the app provide information on common mental disorders and their treatment (psychoeducation) and teach self-management skills to support users in coping with symptoms and problems. Although some courses are disorder specific, most follow a transdiagnostic approach, considering the heterogeneity and high comorbidity in mental illness. The learning goals of the courses include, for example, identifying and gradually changing unhelpful thought patterns and basic assumptions, clarifying personal goals and values, promoting functional stress management behaviors, and fostering the ability to relax. All the courses and exercises are based on the fundamentals of cognitive behavioral therapy and its derivatives (eg, acceptance and commitment therapy and mindfulness-based stress reduction).

All the app's content was developed by or under the supervision of licensed clinical psychological psychotherapists based on established approaches and guidelines for identifying and treating mental illness. The MindDoc app underwent no major changes in content or functioning during the intervention phase of the trial. Updates to the app that occurred during the research study were bug fixes and performance improvements and would not have impacted the therapeutic approach or usability of the app.

The app can send push notifications to a user's phone every time a new question block is ready to be answered (3 times a day) and when automated feedback (insights) has been generated. Push notifications can be turned on and off by each user for question blocks, insights, or both. We did not monitor whether the trial participants used this feature. The notification

settings in the RCT did not differ from those in routine application. There were no cointerventions, except for support by the study coordinator in installing the app.

The app contains detailed information on how to access mental health care. Participants who report a high symptom burden or functional impairment within the monitoring function of the app will be prompted to consult a health care professional in the automated feedback. Furthermore, users are repeatedly reminded that study participation does not substitute for diagnosis, counseling, or treatment by a licensed physician or psychotherapist.

The MindDoc app is a commercial product with both free and premium (paid) features. The study participants had free access to all the features. A more detailed description of the app is provided in the Multimedia Appendix 1. Descriptions of the app in this section and in Multimedia Appendix 1 correspond to the version of the app used in the research study and may not be exactly apply to the currently available version of the app.

Recruitment Strategy

The participants were openly recruited via press releases and social media as well as health insurance member magazines and websites in Germany. Recruitment took place over a period of approximately 7 months (December 2020 to June 2021). Participation in the study was anonymous, but the participants were required to provide an (anonymous) email address through which they could be contacted.

Participants and Procedures

Assessment Procedure

All assessments related to the trial were carried out outside the app via web-based surveys (self-assessment) on a web-based platform (Unipark/EFS Survey; Questback GmbH).

After receiving detailed written information about the study procedures and data processing, participants provided electronic informed consent. Participants were screened according to the predefined inclusion and exclusion criteria (see the subsequent sections). Eligible individuals then received access to the baseline assessment. Those who completed baseline assessments were randomly assigned to either the IG or CG in a 1:1 ratio using an algorithm provided by the assessment platform (Unipark/EFS Survey). They were immediately informed about the result of the assignment on the assessment platform and via email.

The participants in the IG received access to the MindDoc app and were recommended to use it for at least 8 weeks, although they had full access to it for 6 months. To this end, they received individual codes that unlocked the content of the app after downloading the app from the app store. The participants of both groups received an email invitation to the postintervention and follow-up assessments 8 weeks and 6 months after the baseline assessment, respectively. Participants who did not complete the postintervention or follow-up assessment were reminded 1 week later.

The participants were not financially compensated for participating in the study. However, participants who had completed the postintervention and follow-up assessments took part in a monthly raffle, where they could win a universal €50 (US $52.66) voucher that can be redeemed in a number of web-based stores.

Inclusion Criteria

We included adults with clinically relevant symptoms of internalizing disorders indicated by scoring above the cutoff for mild symptoms on one or more of the following scales: Patient Health Questionnaire–9 (PHQ-9) score>4 [44], Generalized Anxiety Disorder–7 (GAD-7) score>4 [45], Mini-Social Phobia Inventory score>6 [46], Patient Health Questionnaire–15 (PHQ-15) score>4 [47], Regensburg Insomnia Scale (RIS) score>12 [48], binge eating or compensatory behaviors>once/wk, BMI<18.5 kg/m^2 or critical weight loss, and weight and shape concern.

In addition, participants needed to have full legal capacity (self-disclosure), have access to a smartphone (iOS [Apple Inc] or Android [Google LLC]) and the internet, and live in Germany.

Exclusion Criteria

We excluded individuals with severe symptoms of internalizing disorders (PHQ-9 score>19, GAD-7 score>15, or PHQ-15 score>14) and severely underweight individuals (BMI<15 kg/m^2). We also excluded individuals who reported acute suicidality and individuals who reported a history of bipolar disorder, psychotic disorder, or substance use disorder. Participants who met these exclusion criteria were provided with detailed information on treatment options.

To ensure that the effects we discovered in the trial were attributable to the use of the app and not to other specific and intensive treatments, we excluded individuals with current or planned outpatient psychotherapy or inpatient treatment for a mental disorder. However, participants were allowed to use or initiate any treatment during the study period.

Outcomes and Measures

Primary Outcomes

We assessed 4 coprimary outcomes in the trial: mental health literacy, mental health–related patient empowerment and self-management skills (MHPSS), attitudes toward help seeking (after 8 wk), and actual mental health service use (after 6 months).

Mental health literacy was assessed using the Mental Health Literacy Questionnaire, which is a 29-item scale with 4 dimensions (knowledge of mental health problems, erroneous beliefs or stereotypes, help-seeking and first-aid skills, and self-help strategies). The measure differentiates well between individuals with more experience with mental health and individuals with less experience with mental health and has good internal consistency (Cronbach α=.84) for the total score [49].

We used the Assessment of Mental Health Related Patient Empowerment and Self-Management-Skills questionnaire, which was constructed based on a systematic review on

self-management skills for depression [50], a Delphi consensus study on self-help strategies for depression [51], 2 studies on useful self-management skills for mood [52] and anxiety [53] disorders from the patient perspective, and a conceptual framework for patient choice and empowerment in northern European health systems [54]. The questionnaire consists of 10 items in a statement format assessing patient empowerment based on how much patients agree or disagree on a 5-point Likert scale, for example, "I know well about the treatment options for my disease," and 18 items in a question format assessing the frequency of self-management skills on a 5-point Likert scale, for example, "In the last 8 weeks, how often have you engaged in activities that gave you a feeling of achievement?" The complete Assessment of Mental Health Related Patient Empowerment and Self-Management-Skills can be retrieved in Multimedia Appendix 2.

Attitudes toward help seeking were assessed using the Inventory of Attitudes Toward Seeking Mental Health Services, which is a 24-item scale assessing 3 internally consistent within-person barriers to seeking mental health services: psychological openness, help-seeking propensity, and indifference to stigma. Internal consistency (Cronbach α=.87) and the validity of the assessment could be confirmed in separate samples [55].

Actual seeking of outpatient psychotherapeutic or psychiatric treatment was assessed via 2 questions asking whether these services were used in the last 6 months.

Secondary Outcomes

The PHQ-9 is the depression module of the self-administered version of the Primary Care Evaluation of Mental Disorders diagnostic instrument for common mental disorders. It scores each of the 9 Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) diagnostic criteria from 0 (not at all) to 3 (nearly every day). The PHQ-9 is a reliable (Cronbach α=.89) and valid measure of depression severity [44]. Higher scores indicate a higher symptom load.

The GAD-7 is a 1D instrument designed to detect symptoms of generalized anxiety disorder as defined in the DSM-5. The item scores range from 0 (not at all) to 3 (nearly every day). The GAD-7 is a valid and efficient tool for screening for anxiety disorders and assessing their severity in clinical practice and research [56]. Higher scores indicate a higher symptom load.

The PHQ-15 is the module for assessing the severity of somatic symptoms of the self-administered version of the Primary Care Evaluation of Mental Disorders diagnostic instrument for common mental disorders. It comprises 15 somatic symptoms from the PHQ, with each symptom scored from 0 ("not bothered at all") to 2 ("bothered a lot"). The PHQ-15 is a reliable (Cronbach α=.80) and valid screening tool for somatization [47]. Higher scores indicate a higher symptom load.

The RIS [48] is a self-rating scale with 10 items for assessing the cognitive, emotional, and behavioral aspects of psychophysiological insomnia. It has good internal consistency with Cronbach α=.89 and distinguishes well between controls and patients with psychophysiological insomnia. Higher scores indicate a higher symptom load.

The Personality Inventory for DSM-5, Brief Form Plus is a short form of the Personality Inventory for DSM-5 with 34 items, which is compatible with the dimensional assessment of maladaptive personality expressions in the International Classification of Diseases, 11th Revision. The Operationalized Psychodynamic Diagnosis-Structure Questionnaire Short is a short 12-item measure for assessing the severity of personality dysfunction. Dimensional assessment of the severity and style of personality dysfunction according to DSM-5 and International Classification of Diseases, 11th Revision are important predictors of treatment course, adherence, and response and general psychopathology [57]. Both the Operationalized Psychodynamic Diagnosis-Structure Questionnaire Short (Cronbach α=.89) and the Personality Inventory for DSM-5, Brief Form Plus (average McDonald ω=0.81) are validated and reliable measures [58-60]. Higher scores indicate higher personality dysfunction.

Quality of life was assessed using the Assessment of Quality of Life-8 Dimensions, which is a 35-item self-assessment scale designed to evaluate health services that impact the psychosocial aspects of quality of life. It assesses 3 physical and 5 psychosocial domains of functioning. It has good reliability (Cronbach α=.96) and convergent and predictive validity [61]. Higher scores indicate a lower quality of life.

Secondary outcomes were symptoms of common mental disorders and quality of life 8 weeks and 6 months after baseline assessments.

To determine the overall burden of symptoms of common mental disorders, we calculated a composite score from the sum scores of the PHQ-9, GAD-7, PHQ-15, and RIS, divided by the respective scale span. The composite score calculated in this manner can take values between 0 and 1, with higher values indicating higher symptom burden.

All assessments were tested before fielding the trial; the web-based survey contained, on average, 12 items per page, and there was no adaptive testing. The questionnaires at the 3 assessment time points contained between 20 and 25 pages of questions, and all items were mandatory. We used no cookies or IP check but identified multiple entries from the same users through either multiple code redemptions on the same device or multiple participant code or email entries. We applied no weighting of the items, and there were no incomplete questionnaires, as all items were mandatory.

Use data were recorded directly in the MindDoc app. Data from the 2 sources (MindDoc app and study survey) were matched via a personalized download link, which users in the IG received after randomization. A day of use was recorded when a user actively engaged with the app on that day, such as answering questions or engaging with an exercise. An exercise was considered engaged with once it was opened.

Statistical Analyses

Assumptions for the appropriate statistical tests were checked for normality through histograms, skewness, and Kolmogorov-Smirnov test; sphericity was assessed through Mauchly test; and the assumption of equality of

variance-covariance matrices was investigated through Box test and Levene test.

If participants entered assessments several times, only the first assessment was used in the consecutive analyses. Participants were excluded if they fulfilled the exclusion criteria or were missing data from their baseline assessments. Missing data from the postintervention and follow-up assessments of nonbinary primary outcomes (mental health literacy, help-seeking attitudes, and MHPSS) were imputed using baseline scores on symptom severity, mental health literacy, patient empowerment, help-seeking attitudes, quality of life, and severity of personality dysfunction and demographic information by applying an iterative Markov Chain Monte Carlo method based on the initial treatment assignment. We calculated imputations for 10, 50, 100, 150, 200, 300, 400, and 500 iterations. Then, we calculated the fraction of missing information (FMI) index for all multiply imputed data sets. The FMI ranges from 0 to 1 (with 1 meaning that 100% of the data necessary for the planned inferences are missing) and is a reliable indicator of the validity of inferences based on imputed data sets [62]. For the following analyses, we chose the number of imputations that yielded no further decline in the FMI compared with the previous ones. To further investigate the robustness of the findings and address potential bias due to nonrandom missing outcome observations, we calculated Random Forest Lee Bounds (RFLBs) for all nonbinary primary and secondary outcomes as a second procedure to account for missing data and potential nonrandom missingness [63].

In addition, a per-protocol (PP) analysis was performed excluding participants who made the following protocol violations: (1) failure to download the app and complete the onboarding process, (2) use of the app before the randomization date, (3) reporting of the use of the MindDoc app during the intervention and follow-up periods in the waitlist condition, and (4) noncompletion of postintervention or follow-up assessment.

To further investigate the robustness of the results and avoid potential bias due to nonrandom missing outcome observations, RFLBs were determined based on the PP sample for all nonbinary primary and secondary outcomes. In this procedure, the first step uses the random forest (RF) procedure to determine variables based on the baseline data that are related to missing data at later measurement time points. On the basis of this, upper and lower bounds were calculated for the mean differences in actual values between the IG and CG on the respective measure under investigation. If the rate of missing values differs across the study arms, the RFLB procedure trims the outcome distribution of the group with the lower dropout rate by removing observations from the lower (upper) end of the distribution using RFs trained on baseline variables that are predictive of dropout so that an upper and a lower bound adjusted for potential nonrandom missing outcome observations are estimated for the treatment effect. Upper and lower bounds that do not include 0 indicate that the measured difference between the groups persists, that is, is robust, even under the assumption of systematic differences in dropouts (missing not at random) between the IG and CG. Baseline variables identified to be predictive of dropout using the RFLB procedure were further investigated for systematic differences between participants who did drop out at the postintervention assessment point and those who did.

To compare the intervention effects on all non–count-based primary and secondary outcomes, we applied analysis of covariance (ANCOVA) between the groups at the posttreatment and follow-up time points, adjusting for baseline scores both on the multiply imputed intention-to-treat (ITT) and the PP data set. Differences in mental health–related health service use between the IG and CG were assessed using chi-square tests for the available data and the PP sample at follow-up. Multiple imputation on zero-inflated count data, such as physician visits, yields unreliable estimates [64] especially if missing data rates are high. Concerns of multiple testing error for the primary outcomes were addressed through Bonferroni correction. Differences in count-based measures of health service use between the groups were assessed using chi-square tests in the PP sample.

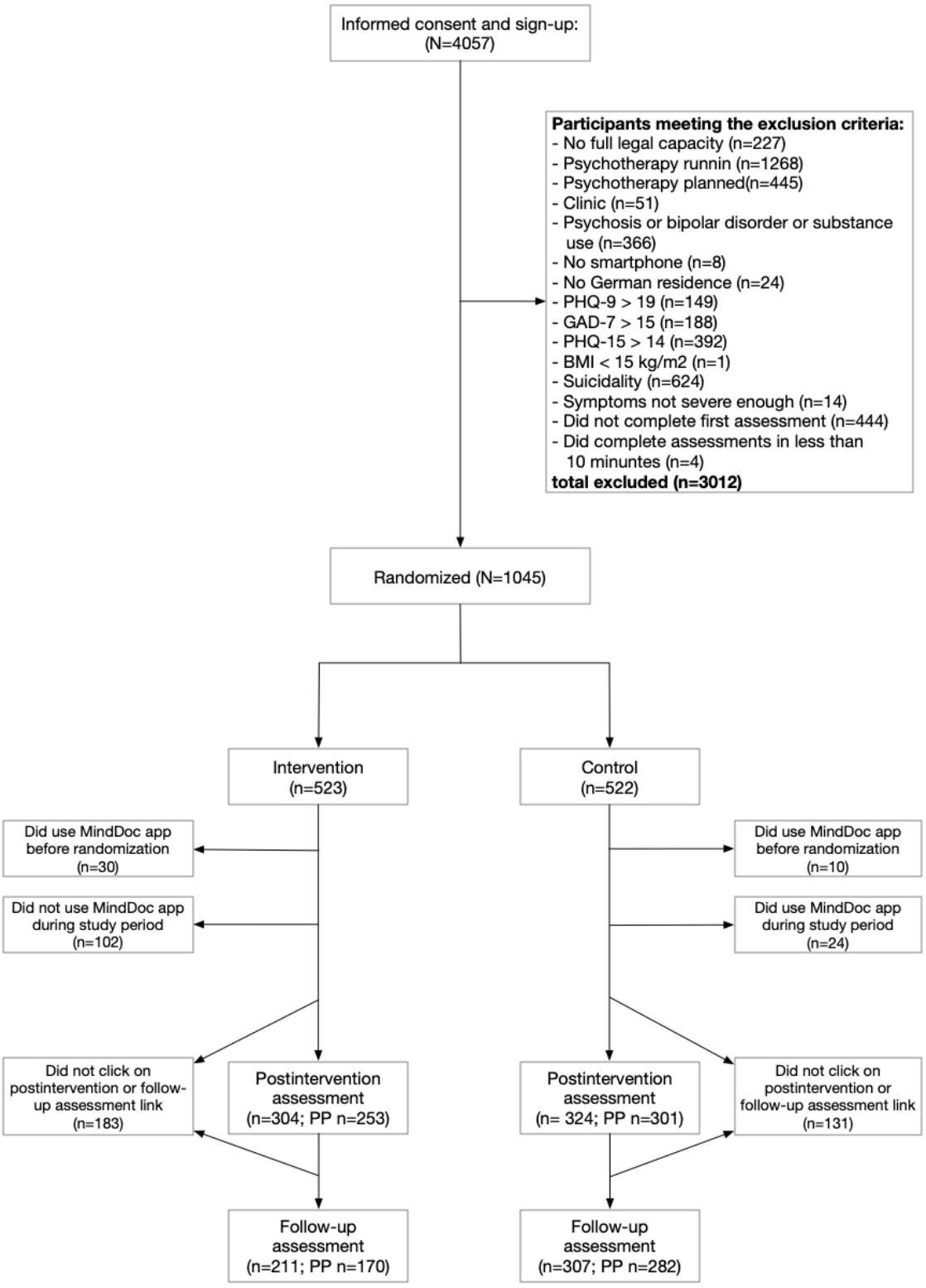
To investigate the predictors of study dropout and adverse events, such as deterioration, baseline variables that were more predictive than a random variable in RF models were identified and investigated using 2-tailed *t* tests.

Results

Participant Flow

Out of the 4057 individuals who provided informed consent to participate in the study, 1045 (25.76%) were randomized (Figure 1). Study dropout rates in the IG and CG were 41.9% (219/523) and 37.9% (198/522) at the postintervention assessment and increased to 59.7% (312/523) and 41.2% (215/522) at the follow-up assessment, respectively. Assessment data from participants without protocol violations (PP sample) was available of 48.4% (253/523) of the participants in the IG and 57.7% (301/522) of the participants in the CG at the postintervention assessment, decreasing to 32.5% (170/523) and 54% (282/522), respectively, at the follow-up assessment.

Figure 1. Participant flow. Note that owing to the possibility of only having 1 assessment after baseline, numbers after baseline assessments are not additive. GAD-7: Generalized Anxiety Disorder–7; PHQ-15: Patient Health Questionnaire–15; PHQ-9: Patient Health Questionnaire–9; PP: per-protocol sample.



Protocol Violations

The MindDoc app was available from the Apple App Store and Google Play Store during the study period. To decrease the risk of app use in the CG, we did not disclose which app was the subject of the trial until after randomization (IG) or follow-up (CG). Nevertheless, protocol violations related to app use did occur in the sample: of the 1045 participants, 40 (3.8%; IG: 30/523, 5.7%; CG: 10/522, 1.9%) used the app before randomization. This information was obtained based on self-report or the use logs of the app manufacturer. An additional 24 (4.6%) participants in the CG used the app during the study period. In the IG, 102 (19.5%) participants did not download the app or complete the app onboarding. The postintervention or follow-up assessment was not completed by 183 (35%)

participants in the IG and 131 (25.1%) participants in the CG. Of the 1045 participants, 10 (1%) did multiple assessments, of which we only used the first assessment respectively for further analyses.

Sample Characteristics

Table 1 presents the baseline characteristics of the participants. The mean age of the overall sample was 38.3 (SD 11.19; range 18-77) years. Most participants (769/1045, 73.59%) were female, and most (865/1045, 82.78%) had completed upper secondary education ("Abitur," European Qualifications Framework level 4) or higher education.

Although current psychotherapeutic treatment was an exclusion criterion for the trial, almost 2 out of 3 (636/1045, 60.86%) participants reported previous psychiatric or psychotherapeutic treatment. The symptom burden was evenly distributed between mild and moderate in most of the individual symptom domains, and most participants (924/1045, 88.42%) reported elevated symptoms in ≥3 domains.

Table 1. Participant characteristics.

Characteristics	Full sample (N=1045)	Intervention (n=523)	Control (n=522)
Sociodemographic characteristics			
Age (y), mean (SD; range)	38.3 (11.19; 18-77)	38.5 (11.65; 18-77)	38.21 (10.72; 18-71)
Gender (women), n (%)	769 (73.6)	393 (75.1)	376 (72)
Education, n (%)			
Higher education	371 (35.5)	188 (35.9)	183 (35.1)
Upper secondary education	494 (47.3)	240 (45.9)	254 (48.7)
Previous psychiatric or psychotherapeutic treatment, n (%)			
None	409 (39.1)	205 (39.2)	204 (39.1)
One treatment provider	331 (31.7)	170 (32.5)	161 (30.8)
Multiple treatment providers	305 (29.2)	148 (28.3)	157 (30.1)
Inpatient treatment	234 (22.4)	118 (22.6)	116 (22.2)
Depression			
PHQ-9[a] total score, mean (SD)	10.62 (3.71)	10.62 (3.83)	10.62 (3.59)
No or minimal symptoms (up to 4 points), n (%)	34 (3.3)	23 (4.4)	11 (2.1)
Mild symptoms (5-9 points), n (%)	392 (37.5)	192 (36.7)	200 (38.3)
Moderate symptoms (10-14 points), n (%)	448 (42.9)	216 (41.3)	232 (44.4)
Moderately severe symptoms (15-19 points), n (%)	171 (16.4)	92 (17.6)	79 (15.1)
Anxiety			
GAD-7[b] total score, mean (SD)	8.26 (3.24)	8.25 (3.17)	8.27 (3.31)
No or minimal symptoms (up to 4 points), n (%)	158 (15.1)	77 (14.7)	71 (13.6)
Mild symptoms (5-9 points), n (%)	537 (51.4)	272 (52)	265 (50.8)
Moderate symptoms (10-15 points), n (%)	350 (33.5)	174 (33.3)	176 (33.7)
Somatic symptoms			
PHQ-15[c] total score, mean (SD)	9.63 (3.24)	9.57 (3.24)	9.69 (3.25)
No or minimal symptoms (up to 4 points), n (%)	139 (13.3)	68 (13)	71 (13.6)
Mild symptoms (5-9 points), n (%)	428 (41)	220 (42.1)	208 (39.8)
Moderate symptoms (10-14 points), n (%)	478 (45.7)	235 (44.9)	243 (46.6)
Insomnia			
RIS[d] total score, mean (SD)	14.04 (5.49)	13.98 (5.52)	14.10 (5.47)
No or few insomnia symptoms (up to 12 points), n (%)	419 (40.1)	216 (41.3)	203 (38.9)
Marked insomnia symptoms (≥12 points), n (%)	626 (59.9)	307 (58.7)	319 (61.1)
Disordered eating, n (%)			
Binge eating at least once per wk in the last 3 mo	111 (10.6)	58 (11.1)	53 (10.2)
Compensatory behaviors at least once per wk in the last 3 mo	49 (4.7)	27 (5.2)	22 (4.2)
Underweight or significant weight loss in combination with weight and shape concerns	10 (1)	7 (1.3)	3 (0.6)
Comorbidity, n (%)			
Elevated symptoms in 1 domain	35 (3.3)	20 (3.8)	15 (2.9)
Elevated symptoms in 2 domains	86 (8.2)	45 (8.6)	41 (7.9)
Elevated symptoms in ≥3 domains	924 (88.4)	458 (87.6)	466 (89.3)

[a]PHQ-9: Patient Health Questionnaire–9 (depression module).

[b]GAD-7: Patient Health Questionnaire–7 (anxiety module).

[c]PHQ-15: Patient Health Questionnaire–15 (somatic symptom module).

[d]RIS: Regensburg Insomnia Scale.

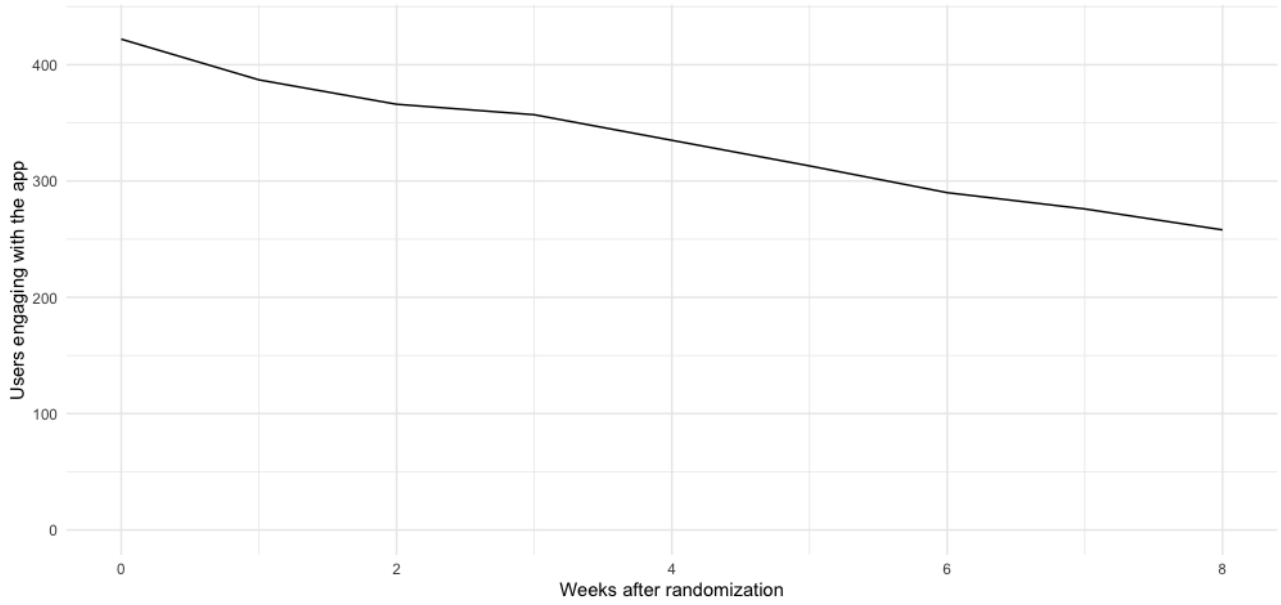
App Use

The participants in the IG were recommended to use the MindDoc app for at least 8 weeks, but they were given unlimited access to the app for 6 months. Table 2 shows the use metrics of the participants in the IG. Figure 2 shows the number of active users per week during the intervention period. During this 8-week period, they used the app, on average, for 2 out of 3 days, and during the 6-month period, they used it for 4 out of 10 days. Although engagement with the questions continued well after the 8-week period, engagement with the courses and exercises subsided over time.

Table 2. App use metrics.

	All the participants in the intervention group who downloaded the app (n=423), mean (SD)	Participants in the intervention group without the protocol violation "app use before randomization" (n=393), mean (SD)
Days of use (within 8 wk)	38.4 (18.9)	38.5 (18.7)
Days of use (within 6 mo)	76.8 (61.6)	76.9 (61)
Question blocks answered (within 8 wk)	88.9 (53.3)	89.3 (53.2)
Question blocks answered (within 6 mo)	177 (159.4)	175 (157.8)
Exercises engaged with (within 8 wk)	16.2 (19.2)	16.3 (19.3)
Exercises engaged with (within 6 mo)	19.8 (24.5)	19.9 (24.6)

Figure 2. User engagements during study period.



Outcome Analysis

Primary Outcome Measures

The results of the efficacy analyses with respect to the 4 coprimary outcomes 8 weeks after baseline assessments are summarized in Table 3.

Intervention effects on the improvement in MHPSS could be confirmed 8 weeks after baseline assessments. ANCOVAs including baseline assessments as predictors showed significant (P<.001) results in both the multiply imputed ITT (100 imputations) and the PP analyses. Effect sizes for the between-group comparison for ITT were Cohen d=0.29 and Cohen d=0.28 for PP analyses 8 weeks after the baseline assessment.

Mental health literacy in the IG did not significantly exceed that in the CG 8 weeks after baseline assessments. Positive help-seeking attitudes in the IG did not significantly exceed those in the CG 8 weeks after baseline assessments. The proportion of participants who actually sought outpatient psychotherapeutic or psychiatric treatments was 35.5% (75/211; available data) or 33.5% (57/170; PP sample) in the IG and 25.7% (79/307; available data) or 25.5% (72/282; PP sample) in the CG. Results of between-group chi-square tests were F_1=5.30 (P=.02) using available data and F_1=2.95 (P=.09) only

using data from the PP sample, indicating a trend toward higher outpatient treatment seeking in the IG, but the results were not significant in the PP sample.

Of the 4 coprimary end points, only the between-group differences in MHPSS yield a *P* value below the Bonferroni corrected α value of .013.

Table 3. Results of analyses of covariance (ANCOVAs) and Cohen d for the nonbinary coprimary outcomes (intention-to-treat [ITT] and per-protocol [PP] analyses) 8 weeks after baseline assessment[a].

	Baseline assessment (IG[b]: n=523; CG[c]: n=522), mean (SD)	Postintervention assessment observed (PP; IG: n=253; CG: n=301), mean (SD)	Postintervention assessment estimated, mean (SD)	ANCOVA (ITT, multiple imputation)			ANCOVA (PP)			RFLB[d]	
				F test (*df*)	*P* value	Cohen *d* (95% CI)	*F* test (*df*)	*P* value	Cohen *d* (95% CI)	Lower bound[e] (SD)	Upper bound[e] (SD)
Mental health literacy (MHLq[f])				0.57 (99, 1130.75)	.45	0.01 (−0.13 to 0.16)	0 (1, 551)	.95	0.02 (−0.14 to 0.19)	−1.23 (0.69)	0.89 (0.6)
Intervention	121.53 (9.12)	121.85 (10.72)	121.87 (11.64)								
Control	121.02 (9.17)	122.06 (8.48)	122.03 (8.71)								
Mental health–related patient empowerment and self-management skills (AMHPSS[g])				12.84 (99, 774.22)	<.001	0.29 (0.15 to 0.43)	13.29 (1, 155)	<.001	0.28 (0.11 to 0.44)	2.96 (1.86)	8.44 (1.56)
Intervention	68.44 (18.44)	75.72 (19.05)	75.44 (19.54)								
Control	66.37 (18.95)	70.41 (19.48)	69.62 (20.46)								
Attitudes toward help seeking (IASMHS[h])				2.26 (99, 1012.95)	.13	0.12 (−0.02 to 0.26)	2.58 (1, 551)	.11	0.12 (−0.05 to 0.29)	−0.27 (1.05)	2.40 (0.73)
Intervention	49.68 (9.19)	51.93 (9.59)	51.97 (9.59)								
Control	49.18 (9.95)	50.73 (10.14)	50.77 (10.42)								

[a]The *F* test results were pooled using the D_2 statistic [65].

[b]IG: intervention group.

[c]CG: control group.

[d]RFLB: Random Forest Lee Bound.

[e]These are lower and upper bounds of the Random Forest Lee Bound procedure.

[f]MHLq: Mental Health Literacy Questionnaire.

[g]AMHPSS: Assessment of Mental Health Related Patient Empowerment and Self-Management-Skills.

[h]IASMHS: Inventory of Attitudes Toward Seeking Mental Health Services.

Secondary Outcomes and Other Measures

For all secondary outcomes and other measures, including the mental health composite score and the measures for depression, anxiety, somatization, insomnia, and quality of life, the results showed significant between-group effects at the postintervention assessment point, controlling for pretreatment scores both in ITT and PP analyses (Table 4). Intervention effects ranged from Cohen *d*=0.19 (Cohen *d*=0.23 in PP sample) for quality of life to Cohen *d*=0.34 (Cohen *d*=0.39 in PP sample) for depression.

Table 4. Results of analyses of covariance (ANCOVAs) and Cohen d for the secondary and additional outcomes (intention-to-treat [ITT] and per-protocol [PP] analyses) at the postintervention assessment point[a].

	Baseline assessment (IG[b]: n=523; CG[c]: n=522), mean (SD)	Postintervention assessment observed (PP; IG: n=253; CG: n=301), mean (SD)	Postintervention assessment estimated, mean (SD)	ANCOVA (ITT, multiple imputation)			ANCOVA (PP)			RFLB[d]	
				F test (*df*)	*P* value	Cohen *d* (95% CI)	*F* test (*df*)	*P* value	Cohen *d* (95% CI)	Lower bound[e] (SD)	Upper bound[e] (SD)
Symptom burden (composite score)				13.75 (99, 633.44)	<.001	0.28 (0.13-0.43)	17.55 (1, 551)	<.001	0.34 (0.17-0.51)	0.02 (0.01)	−0.04 (0.01)
Intervention	0.43 (0.07)	0.41 (0.09)	0.41 (0.09)								
Control	0.43 (0.07)	0.44 (0.09)	0.44 (0.10)								
Quality of life (AQoL-8D[f])				5.49 (99, 672.22)	.02	0.19 (0.05-0.34)	5.22 (1, 551)	.02	0.23 (0.06-0.40)	1.72 (1.60)	6.46 (1.26)
Intervention	88.13 (13.77)	84.09 (16.36)	84.22 (16.67)								
Control	89.19 (14.49)	87.90 (16.92)	87.46 (16.92)								
Depression (PHQ-9[g])				21.54 (99, 690.83)	<.001	0.34 (0.19-0.49)	25.22 (1, 551)	<.001	0.39 (0.22-0.56)	1.13 (0.45)	2.32 (0.30)
Intervention	10.62 (3.83)	8.96 (4.44)	9.10 (4.55)								
Control	10.63 (3.59)	10.71 (4.58)	10.65 (4.60)								
Anxiety (GAD-7[h])				8.73 (99, 693.01)	.003	0.22 (0.07-0.37)	11.34 (1, 551)	.001	0.28 (0.11-0.45)	0.68 (0.34)	1.61 (0.24)
Intervention	8.25 (3.17)	7.31 (4.06)	7.39 (4.13)								
Control	8.27 (3.31)	8.46 (4.13)	8.30 (4.14)								
Somatic symptoms (PHQ-15[i])				7.07 (99, 837.46)	.008	0.20 (0.05-0.35)	2.14 (1, 551)	.008	0.24 (0.08-0.41)	0.46 (0.41)	1.62 (0.33)
Intervention	9.57 (3.24)	9.02 (4.13)	9.07 (4.18)								
Control	9.69 (3.25)	10.05 (9.92)	9.92 (4.41)								
Insomnia (RIS[j])				5.31 (99, 672.22)	.02	0.19 (0.05-0.34)	5.22 (1, 551)	.02	0.23 (0.06-0.40)	0.33 (0.54)	1.89 (0.44)
Intervention	13.98 (5.52)	12.65 (5.77)	12.64 (5.84)								
Control	14.10 (5.47)	13.71 (5.72)	13.52 (5.84)								

[a]The *F* test results were pooled using the D_2 statistic [65].

[b]IG: intervention group.

[c]CG: control group.

[d]RFLB: Random Forest Lee Bound.

[e]These are lower and upper bounds of the Random Forest Lee Bound procedure.

[f]AqOL-8D: Assessment of Quality of Life–8 Dimensions.

[g]PHQ-9: Patient Health Questionnaire–9 (depression module).

[h]GAD-7: Generalized Anxiety Disorder–7 (anxiety module).

[i]PHQ-15: Patient Health Questionnaire–15 (somatic symptom module).

[j]RIS: Regensburg Insomnia Scale.

Sensitivity Analyses

To further investigate the robustness of the results and avoid potential bias due to nonrandom missing outcome observations, RFLBs were determined for all nonbinary primary and secondary outcomes. Using this method, baseline variables that are predictive of dropout were identified, and participants from the lower and upper distributions on a specific outcome measure were excluded based on a RF decision tree trained on the group with more dropouts.

Differences between the groups in patient empowerment and self-management skills showed robustness with RFLBs excluding 0 for both postintervention and follow-up assessments.

For mental health literacy and attitudes toward help seeking, upper and lower bounds included 0 for the postintervention assessment, whereas RFLBs for the follow-up assessment did not include 0. Thus, it is highly unlikely that group differences in mental health literacy and help-seeking attitudes at follow-up are solely attributable to nonrandom missing outcome observations.

All secondary and additional outcomes showed robust treatment effects corrected for potential nonrandom missingness using the RFLB procedures at the postintervention assessment point.

Predictors of Dropout

Variable importance plots from the RF models indicating baseline variables that are predictive of attrition can be found in Multimedia Appendix 3. In the IG, 21 baseline variables were more predictive of dropout than a random variable. In the CG, 19 baseline variables were more predictive of dropout than a random variable. The 5 most predictive variables for dropout in the RF models in the IG were lower age; lower personality functioning; higher detachment; and lower mental health–related self-management skills, attitudes, and literacy. In the CG, lower mental health–related self-management skills, attitudes, and literacy; lower quality of life; lower age; and higher disinhibition were the most predictive variables in the RF models. We further investigated all variables identified to be predictive of dropout in the RF models using *t* tests for systematic between-group differences between participants who did and did not drop out at the postintervention assessment point. Here, we found small but significant differences in personality functioning (Cohen *d*=0.12), disinhibition (Cohen *d*=0.24), psychoticism (Cohen *d*=0.16), and age (Cohen *d*=0.15). That is, participants who dropped out were, on average, younger and had higher personality dysfunction scores.

Follow-Ups

Tables 5 and 6 show the results of ITT and PP ANCOVAs as well as RFLB analyses for all outcomes at follow-up. We found a robust group difference in mental health literacy (Cohen *d*=0.28, 95% CI 0.09-0.46) in addition to the difference in MHPSS that was maintained from the postintervention assessment point (Cohen *d*=0.34, 95% CI 0.18-0.55). The differences became apparent in the ITT, PP, and RFLB analyses. Attitudes toward help seeking yielded a small between-group effect (Cohen *d*=0.22, 95% CI 0.02-0.42) with RFLBs not containing 0, whereas ITT and PP ANCOVAs revealed a trend toward a group difference (ITT, *P*=.053; PP, *P*=.06). The mental health composite score showed a small between-group effect in the PP sample at follow-up (Cohen *d*=0.20, 95% CI 0.01-0.39), which could not be corroborated in RFLB analyses and ITT and PP ANCOVAs. From the measures of symptoms of common mental disorders, anxiety (GAD-7) showed a significant difference in the ANCOVA of the PP sample at follow-up with RFLBs not including 0, but the ANCOVA in the ITT sample did not reach significance (*P*=.34). Quality of life, depression, somatization, and sleep symptoms showed no significant between-group differences at follow-up (all *P*>.05).

Table 5. Results of analyses of covariance (ANCOVAs) and Cohen d for the nonbinary coprimary outcomes (intention-to-treat [ITT] and per-protocol [PP] analyses) at follow-up[a].

	Baseline (IG[b]: n=523; CG[c]: n=522), mean (SD)	Follow-up assessment observed (PP; IG: n=170; CG: n=282), mean (SD)	Follow-up assessment estimated, mean (SD)	ANCOVA (ITT, multiple imputation)			ANCOVA (PP)			RFLB[d]	
				F test (*df*)	*P* value	Cohen *d* (95% CI)	*F* test (*df*)	*P* value	Cohen *d* (95% CI)	Lower bound[e] (SD)	Upper bound[e] (SD)
Mental health literacy (MHLq[f])				8.11 (99, 266.16)	.005	0.28 (0.09-0.46)	7.7 (1, 449)	.006	0.22 (0.03-0.42)	0.24 (1.40)	3.77 (1.25)
Intervention	121.53 (9.12)	123.73 (8.81)	124.13 (9.95)								
Control	121.02 (9.17)	121.66 (9.48)	121.38 (9.95)								
Mental health–related patient empowerment and self-management skills (AMHPSS[g])				12.13 (99, 0255.15)	.001	0.37 (0.18-0.55)	16.08 (1, 449)	<.001	0.38 (0.19-0.58)	3.51 (3.10)	11.33 (2.91)
Intervention	68.44 (18.44)	78.61 (19.31)	78.53 (21.64)								
Control	66.37 (18.95)	70.96 (20.33)	70.70 (21.29)								
Attitudes toward help seeking (IASMHS[h])				3.78 (99, 206.32)	.053	0.22 (0.02-0.42)	3.53 (1, 449)	.06	0.27 (0.07-0.46)	0.56 (1.56)	3.93 (1.56)
Intervention	49.68 (9.19)	53.05 (9.56)	52.51 (10.55)								
Control	49.18 (9.95)	50.56 (9.31)	50.34 (9.57)								

[a]The *F* tests results were pooled using the D_2 statistic [65].

[b]IG: intervention group.

[c]CG: control group.

[d]RFLB: Random Forest Lee Bound.

[e]These are lower and upper bounds of the Random Forest Lee Bound procedure.

[f]MHLq: Mental Health Literacy Questionnaire.

[g]AMHPSS: Assessment of Mental Health Related Patient Empowerment and Self-Management-Skills.

[h]IASMHS: Inventory of Attitudes Toward Seeking Mental Health Services.

Table 6. Results of analyses of covariance (ANCOVAs) and Cohen d for the secondary and additional outcomes (intention-to-treat [ITT] and per-protocol [PP] analyses) at follow-up[a].

	Baseline (IG[b]: n=523; CG[c]: n=522), mean (SD)	Follow-up assessment observed (PP; IG: n=170; CG: n=282), mean (SD)	Follow-up assessment estimated, mean (SD)	ANCOVA (ITT, multiple imputation)			ANCOVA (PP)			RFLB[d]	
				F test (*df*)	*P* value	Cohen *d* (95% CI)	*F* test (*df*)	*P* value	Cohen *d* (95% CI)	Lower bound[e] (SD)	Upper bound[e] (SD)
Symptom burden (composite score)				0.96 (99, 448.81)	.33	0.05 (−0.17 to 0.26)	17.55 (1, 449)	.06	0.20 (0.01 to 0.39)	0 (0.01)	0.04 (0.00)
Intervention	0.43 (0.07)	0.40 (0.10)	0.42 (0.11)								
Control	0.43 (0.07)	0.42 (0.09)	0.42 (0.10)								
Quality of life (AQoL-8D[f])				0.71 (99, 371.03)	.40	0.06 (−0.15 to 0.26)	0.00 (1, 449)	.98	.06 (−0.13 to 0.25)	−2.73 (2.74)	4.49 (2.60)
Intervention	88.13 (13.77)	84.32 (17.21)	84.12 (18.95)								
Control	89.19 (14.49)	85.39 (18.66)	85.21 (19.02)								
Depression (PHQ-9[g])				0.56 (99, 462.45)	.46	0.04 (−0.15 to 0.23)	2.09 (1, 449)	.15	1.15 (−0.04 to 0.34)	−0.20 (0.65)	1.64 (0.60)
Intervention	10.62 (3.83)	8.95 (5.17)	9.47 (5.85)								
Control	10.63 (3.59)	9.72 (4.90)	9.66 (7.43)								
Anxiety (GAD-7[h])				0.89 (99, 448,48)	.34	0.06 (−0.13 to 0.26)	2.51 (1, 449)	.11	0.18 (−0.01 to 0.38)	0.03 (0.50)	1.37 (0.40)
Intervention	8.25 (3.17)	6.99 (4.24)	7.43 (5.04)								
Control	8.27 (3.31)	7.76 (4.11)	7.72 (4.35)								
Somatic symptoms (PHQ-15[i])				0.98 (99, 401.19)	.32	0.07 (−0.13 to 0.27)	2.14 (1, 449)	.14	0.10 (−0.09 to 0.29)	−0.38 (0.59)	1.21 (0.58)
Intervention	9.57 (3.24)	9.01 (4.06)	9.12 (4.51)								
Control	9.69 (3.25)	9.44 (4.40)	9.44 (4.64)								
Insomnia (RIS[j])				1.00 (99, 431.29)	.32	0.07 (−0.12 to 0.26)	0.06 (1, 449)	.80	0.02 (−0.01 to 0.21)	−0.93 (0.80)	1.24 (0.80)
Intervention	13.98 (5.52)	13.25 (5.60)	13.05 (6.27)								
Control	14.10 (5.47)	13.39 (5.97)	13.49 (6.36)								

[a]The *F* test results were pooled using the D_2 statistic [65].

[b]IG: intervention group.

[c]CG: control group.

[d]RFLB: Random Forest Lee Bound.

[e]These are lower and upper bounds of the Random Forest Lee Bound procedure.

[f]AqOL-8D: Assessment of Quality of Life–8 Dimensions.

[g]PHQ-9: Patient Health Questionnaire–9 (depression module).

[h]GAD-7: Generalized Anxiety Disorder–7 (anxiety module).

[i]PHQ-15: Patient Health Questionnaire–15 (somatic symptom module).

[j]RIS: Regensburg Insomnia Scale.

Adverse Events

Using the follow-up assessment, adverse events were recorded for individuals with available data (IG: 211/523, 40.3%; CG: 307/522, 58.8%). Suicidal ideations were reported by 48 (22.7%) participants in the IG and 75 (24.4%) participants in the CG (χ^2_1=0.1, *P*=.74). Symptom deterioration, that is, an increase in the scores of the GAD-7, PHQ-9, or PHQ-15 larger than minimally clinical important differences (MCIDs) that were available from previous validation studies [66-68] was detected in 19 (9%) participants in the IG and 31 (10.1%) participants in the CG for the PHQ-9 (χ^2_1=0.1, *P*=.79), 34 (16.1%) participants in the IG and 46 (15%) participants in the CG for the GAD-7 (χ^2_1=0.1, *P*=.82), and 22 (10.4%) participants in the IG and 33 (10.7%) participants in the CG for the PHQ-15 (χ^2_1=0.0, *P*>.99). Deterioration rates of the PHQ-9 scores after 8 weeks were 7.9% based on MCIDs (27.6% worsening of any size) in the IG and 13% based on MCIDs (44.4% worsening of any size) in the CG, with χ^2_1=3.8 and *P*=.05 for the group difference based on the MCID method and χ^2_1=18.5 and *P*<.001 for group differences based on any size of deterioration.

Variable importance plots from the RF models indicating baseline variables that were predictive of deterioration can be found in the Multimedia Appendix 3. In these models, 21 baseline variables were more predictive of deterioration than a random variable. Further investigation of all variables identified using *t* tests for systematic between-group differences between participants who did deteriorate at follow-up and participants who did not yielded significant differences in the GAD-7 (Cohen *d*=0.53), PHQ-15 (Cohen *d*=0.28), PHQ-9 (Cohen *d*=0.23), RIS (Cohen *d*=0.12), outpatient psychiatric or psychotherapeutic treatment (Cohen *d*=−0.31), other medical treatment (Cohen *d*=−0.24), and higher education (Cohen *d*=0.21). That is, participants who deteriorated had, on average, more severe symptoms at baseline; received less outpatient treatment before baseline; and had a higher education.

Inpatient treatment was reported by 43 (20.4%) participants in the IG and 55 (17.9%) participants in the CG (χ^2_1=0.4, *P*=.56). Severe health issues were reported by 29 (13.7%) participants in the IG and 45 (14.7%) participants in the CG (χ^2_1=0.0, *P*=.87). There were no serious adverse events that required a report to the institutional review board.

Discussion

General Findings

This study was the first to investigate the effects of a transdiagnostic self-guided mental health app on mental health–related patient empowerment, attitudes, self-management skills, and literacy in a large sample. Overall, the intervention yielded small but lasting effects on patient empowerment and self-management skills. It was also shown that improvements in symptom burden and quality of life can be accelerated by using the app. During the follow-up period, improvements in mental health literacy were detected, which were related to access to the intervention. Attitudes toward help seeking and use of outpatient psychiatric or psychotherapeutic services showed a trend toward change through intervention 6 months after baseline assessments. Although the effect sizes are small, the public health impact of the self-guided intervention can still be substantial given its scalability and comparatively low cost [17].

Patient Empowerment, Mental Health Literacy, and Help Seeking

Previous research has shown that self-management skills play a key role in recovery from severe mental disorders [69-72]. Improving self-management skills and patient empowerment, especially in people with comorbid mental disorders, is an important factor contributing to mental health–related recovery processes [73]. Besides leading to more favorable treatment outcomes [74], improving self-management skills may lead to a sense of empowerment and responsibility as well as a sense of partnership between patients and clinicians [75].

The application of self-management skills may constitute an important component of interventions in multiple settings [76]. In an RCT with patients with chronic and treatment-resistant anxiety or depressive disorders, an intervention targeting self-management in addition to face-to-face treatment led to significant group differences in patient empowerment and social relations, and a larger proportion of patients were able to reduce treatment intensity [77]. A recent systematic review found self-help strategies to be an important mediator of outcomes in web-based interventions for depression [42]. Further, improving mental health–related self-management skills may shift patients' approach to mental health problems from being disease centered to being more health centered [78].

Concerning attitudes toward help seeking and the use of outpatient psychiatric or psychotherapeutic services, we found some indications for intervention effects 6 months after baseline but no robust effects. As a substantial proportion of our participants (636/1045, 60.86%) had previous experience with psychiatric or psychotherapeutic treatment, internal barriers to treatment seeking were likely less prevalent in our sample than in the general population or other samples. To date, evidence on changes in help-seeking attitudes or behavior through web-based interventions is scarce. Although a more recent study found promising effects in a very small sample [38], findings from older studies with larger samples were mixed [37,39]. All these studies had a rather short follow-up period, especially given the fact that mental disorders can persist over years until those affected seek treatment. A longer intervention and follow-up period may be necessary to induce and measure changes in trust among mental health providers. A similar explanation may hold true for our findings on mental health literacy. Previous studies investigating changes in mental health

literacy using follow-up measurements 6 months after baseline are scarce and not comparable in terms of sample size. Although a study on schizophrenia-related health literacy in a sample of 31 participants [79] found effects 6 months after a web-based psychoeducational intervention, another study [36] found comparable effects on depression and cognitive behavioral therapy literacy at the postintervention assessment point and 6-month follow-up. As participants in this study already had high mental health literacy scores at baseline, an interpretation of our findings may be that our participants were already quite mental health literate at baseline, and changes beyond this high average level of mental health literacy may take time.

Symptom Burden and Quality of Life

Exploratory analyses revealed a small effect of the intervention on both symptoms of common mental disorders and quality of life at the postintervention assessment point. Symptom burden improved modestly during the intervention period in the IG but was slightly exacerbated in the CG. This pattern was observed in 3 (depression, anxiety, and somatic symptoms) out of 4 domains of symptoms, and the between-group difference was reflected in small effect sizes. Insomnia symptoms improved in both groups during the intervention period, but the improvement in the IG was greater. At follow-up, the results were mixed. On average, participants in the CG had caught up with those in the IG regarding improvements in both symptoms of common mental disorders and quality of life, although the mental health composite score and anxiety measure showed a trend toward greater change in the IG. Overall, both groups achieved comparable improvements during the combined intervention and follow-up period, but participants in the IG achieved these improvements earlier.

User Engagement and Adverse Events

In participants in the IG who downloaded the MindDoc app, the average period of use went beyond the recommended 8 weeks. In addition, users frequently engaged with the intervention, on average, by accessing the app every other day.

There were no differences between the IG and CG in the frequency of adverse events (suicidal ideations, symptom deterioration, inpatient treatment, and severe health issues) 6 months after randomization. These findings are in line with those from other recent trials on digital mental health interventions that reported on adverse events (eg, the studies by MacLean et al [80], Baumeister et al [81], Oehler et al [82], Axelsson et al [83] and Reins et al [84]). Suicidal ideations were reported by 22.7% (48/211) and 24.4% (75/307) of the IG and CG participants who reported adverse events, respectively, during the follow-up period. Given the fact that more than half of the participants reported moderate or moderately severe symptoms of depression at baseline, this is an expected finding and corresponds with the prevalence of suicidal ideation in adults with major depression [85,86]. There was no indication that the adverse events were associated with the intervention and study participation. Symptom deterioration measured using the PHQ-9 after 8 weeks was significantly (P=.05 using MCIDs and P<.001 using any size of deterioration) higher in the CG and comparable with that in a meta-analysis from 2018 [86] investigating deterioration rates at the postintervention measurement point in RCTs of digital interventions for depression, although these investigated RCTs had more participant contact, such as weekly phone calls. Insignificant group differences in deterioration rates after 6 months and significant group differences at the postintervention measurement point were also found in a previous meta-analysis on digital interventions [87].

However, the IG in our study was associated with a faster improvement in symptoms of common mental disorders and quality of life than the CG. The MindDoc app showed effects that were comparable with findings from previous meta-analyses on the efficacy of self-guided app-based interventions for quality of life, depression, and anxiety [7]. Previous evidence on self-help interventions for somatization showed slightly higher effects [88,89]. Early symptom improvement is an important predictor of long-term outcome in mental disorders. In a study based on data of more than half a million health care users in England, the time passed without intervention was a strong and reliable predictor of later chronicity and nonresponse to treatment [90]. From this perspective, small but early improvement of symptoms through a self-guided web-based intervention such as the MindDoc app may prevent chronification and support remission and recovery.

As is common in studies investigating self-guided interventions [10], especially with web-based recruitment [91], dropout rates in our study were high. Moreover, the dropout rate in the IG was higher than that in the CG. This may be related to the fact that participants in the CG received access to the MindDoc app after the completion of the follow-up assessment, which could have represented an additional incentive that was absent in the IG. The most predictive variables for trial discontinuation in the IG were younger age and higher personality dysfunction. In particular, the latter is a relevant predictor of the amount of guidance required in psychological interventions [92]. Tackling this issue may require, besides human-delivered guidance, adjusting app behavior and content according to dimensional measures of personality functioning or other dimensional conceptualizations of psychopathology such as HiTOP [20]. The strongest predictors of deterioration within participants that did not drop out were more severe symptomatology and less outpatient treatment at baseline, which also point to the need for more guidance in cases with more severe psychopathology. Symptom severity was previously found to be an important indicator for more guidance in internet-based interventions [93,94] and higher treatment intensity in stratified care [95].

Limitations

Nevertheless, this study had a number of limitations. First, dropout rates were high for both postintervention and follow-up assessments, which may have biased the findings, especially those of the ITT analyses using multiple imputation. However, provided sufficient auxiliary variables, multiple imputation can provide valid estimations even in situations with large proportions of missing data [62]. The dropout rates in our study were comparable with those in previous studies on self-guided internet-based mental health interventions [96], especially in studies with entirely web-based recruitment methods [91]. Further, there were no systematic differences in any primary or

secondary outcome variable at baseline between participants who dropped out and those who did not drop out 8 weeks or 6 months after baseline assessments.

There was no active CG within the study design; however, participants were allowed to use any treatment during the study period, and a significant proportion of our participants stated that they had used mental health specialist care during the trial period.

Another limitation concerns the inclusion of participants based solely on self-report. Although this constitutes a very low threshold and a cost- and time-saving approach, it very likely resulted in the inclusion of a substantial proportion of participants with subclinical symptoms, which may have resulted in floor effects. Therefore, future research on the intervention should include clinical interviews upon the inclusion of study participants to confirm diagnoses.

Another limitation concerns the sample. Most participants were recruited through social media advertisements. This resulted in a study sample in which more than half of the participants had previous treatment experience. Therefore, mental health literacy at baseline was likely to be higher in this sample than in the general population or a primary care sample, and positive attitudes toward help seeking were likely to be more prevalent. Both may have resulted in floor effects on the respective outcomes. Therefore, future research on the effect of the intervention on mental health literacy and attitudes toward help seeking should specifically address treatment-naive participants and be conducted in a primary care setting.

Conclusions

Using a nonguided transdiagnostic mental health app not only improved patient empowerment and self-management skills but also accelerated the improvement of symptoms of common mental disorders and quality of life. Thus, the intervention supported patients dealing with the symptoms of one or multiple internalizing disorders in developing more health-centered coping skills, preventing chronification, and facilitating recovery.

The effect sizes were small, but given the high scalability of a fully automated intervention, the impact on public mental health may still be considerable. With effect sizes comparable with those of antidepressants [97] and a possibly lower risk for unwanted somatic side effects, a self-guided app-based intervention can be a low-threshold, low-cost addition to mental health care services across the treatment spectrum.

Future studies are needed to specifically address the effects of self-guided mental health apps in various health care settings, including primary care, blended treatment, and aftercare, and differential effects in various groups of users, including those with subthreshold symptoms and those with full-syndrome mental disorders.

Another focus for future research may be the further customization of app behavior depending on dimensional measures of personality functioning or psychopathology, for example, adjusting the amount of automated in-app guidance and the number of automated in-app reminders, tailoring the user experience to meet the needs of users with different degrees of impairment, or providing more options for customization based on personal preferences.

Acknowledgments

This study was funded by MindDoc Health GmbH (Leopoldstraße) within the framework of a scientific cooperation agreement between Freie Universität Berlin and MindDoc Health GmbH. AK, SB, and CK have ultimate authority over the study design; collection, management, analysis, and interpretation of data; writing of the report; and decision to submit the report for publication.

Authors' Contributions

AK and IB conceptualized and designed the study, wrote the first draft of the protocol, and developed the statistical analysis plan. AK performed the statistical analyses. AK and IB drafted the manuscript. AK, SB, and CK reviewed the manuscript; AK and IB revised the manuscript based on these reviews and created the final version for submission. The manuscript was written according to the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guidelines.

Conflicts of Interest

IB is the chief science and clinical officer of MindDoc Health GmbH, the app manufacturer. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Detailed description of the digital mental health app MindDoc.
[PDF File (Adobe PDF File), 1798 KB - mental_v10i1e45068_app1.pdf]

Multimedia Appendix 2
Assessment of Mental Health Related Patient Empowerment and Self-Management-Skills.
[PDF File (Adobe PDF File), 86 KB - mental_v10i1e45068_app2.pdf]

Multimedia Appendix 3
Variables predictive of dropout or deterioration identified by random forest models.
[PDF File (Adobe PDF File), 460 KB - mental_v10i1e45068_app3.pdf]

Multimedia Appendix 4
CONSORT EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (version 1.6.1).
[PDF File (Adobe PDF File), 450 KB - mental_v10i1e45068_app4.pdf]

References

1. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://iris.who.int/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2022-12-15]
2. Jacobi F, Höfler M, Strehle J, Mack S, Gerschler A, Scholl L, et al. [Mental disorders in the general population : study on the health of adults in Germany and the additional module mental health (DEGS1-MH)]. Nervenarzt 2014 Jan;85(1):77-87. [doi: 10.1007/s00115-013-3961-y] [Medline: 24441882]
3. Schlack R, Hapke U, Maske U, Busch M, Cohrs S. Frequency and distribution of sleep problems and insomnia in the adult population in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1). Article in German. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2013 May 25;56(5-6):740-748. [doi: 10.1007/s00103-013-1689-2] [Medline: 23703493]
4. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health 2013 Mar 18;34(1):119-138 [FREE Full text] [doi: 10.1146/annurev-publhealth-031912-114409] [Medline: 23514317]
5. Kamp L, Pickshaus K. Regelungslücke psychische Belastungen schließen. Hans-Böckler Siftung. 2011. URL: https://klaus-pickshaus.de/wp-content/uploads/2014/10/JahrbuchKritMed_Psyche.pdf [accessed 2022-12-15]
6. Direkte Kosten psychischer Erkrankungen in Deutschland nach Krankheitsart in den Jahren 2002 bis 2012. Statista. URL: https://de.statista.com/statistik/daten/studie/246590/umfrage/direkte-kosten-psychischer-erkrankungen-in-deutschland-nach-krankheitsart/ [accessed 2023-10-09]
7. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct 09;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
8. Blasko K, Ciulla R, Cavanagh R, Kinn J, Armstrong C, Micheel L, et al. Internet-based mental health interventions: evidence, practical considerations, and future directions. In: Technology and Mental Health. Milton Park, UK: Routledge; 2020.
9. Koelen JA, Vonk A, Klein A, de Koning L, Vonk P, de Vet S, et al. Man vs. machine: a meta-analysis on the added value of human support in text-based internet treatments ("e-therapy") for mental disorders. Clin Psychol Rev 2022 Aug;96:102179 [FREE Full text] [doi: 10.1016/j.cpr.2022.102179] [Medline: 35763975]
10. Lakhtakia T, Torous J. Current directions in digital interventions for mood and anxiety disorders. Curr Opin Psychiatry 2022 Mar 01;35(2):130-135. [doi: 10.1097/YCO.0000000000000772] [Medline: 34966117]
11. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
12. Schröder J, Berger T, Westermann S, Klein JP, Moritz S. Internet interventions for depression: new developments. Dialogues Clin Neurosci 2022 Apr 01;18(2):203-212 [FREE Full text] [doi: 10.31887/dcns.2016.18.2/jschroeder]
13. Pauley D, Cuijpers P, Papola D, Miguel C, Karyotaki E. Two decades of digital interventions for anxiety disorders: a systematic review and meta-analysis of treatment effectiveness. Psychol Med 2023 Jan 28;53(2):567-579 [FREE Full text] [doi: 10.1017/S0033291721001999] [Medline: 34047264]
14. Titov N, Fogliati VJ, Staples LG, Gandy M, Johnston L, Wootton B, et al. Treating anxiety and depression in older adults: randomised controlled trial comparing guided v. self-guided internet-delivered cognitive-behavioural therapy. BJPsych Open 2016 Jan;2(1):50-58 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002139] [Medline: 27703754]
15. Titov N, Dear BF, Johnston L, Lorian C, Zou J, Wootton B, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. PLoS One 2013;8(7):e62873 [FREE Full text] [doi: 10.1371/journal.pone.0062873] [Medline: 23843932]
16. Massoudi B, Holvast F, Bockting CL, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: a systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:728-743. [doi: 10.1016/j.jad.2018.11.050] [Medline: 30447572]
17. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
18. Caspi A, Houts RM, Ambler A, Danese A, Elliott ML, Hariri A, et al. Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the Dunedin birth cohort study. JAMA Netw Open 2020 Apr 01;3(4):e203221 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3221] [Medline: 32315069]

19. Kessler RC, Ormel J, Petukhova M, McLaughlin KA, Green JG, Russo LJ, et al. Development of lifetime comorbidity in the World Health Organization world mental health surveys. Arch Gen Psychiatry 2011 Jan 03;68(1):90-100 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2010.180] [Medline: 21199968]
20. Kotov R, Krueger RF, Watson D, Cicero DC, Conway CC, DeYoung CG, et al. The hierarchical taxonomy of psychopathology (HiTOP): a quantitative nosology based on consensus of evidence. Annu Rev Clin Psychol 2021 May 07;17(1):83-108. [doi: 10.1146/annurev-clinpsy-081219-093304] [Medline: 33577350]
21. Barlow DH, Sauer-Zavala S, Carl JR, Bullis JR, Ellard KK. The nature, diagnosis, and treatment of neuroticism: back to the future. Clin Psychol Sci 2013 Oct 14;2(3):344-365. [doi: 10.1177/2167702613505532]
22. Gosmann NP, Costa MA, Jaeger MB, Motta LS, Frozi J, Spanemberg L, et al. Selective serotonin reuptake inhibitors, and serotonin and norepinephrine reuptake inhibitors for anxiety, obsessive-compulsive, and stress disorders: a 3-level network meta-analysis. PLoS Med 2021 Jun;18(6):e1003664 [FREE Full text] [doi: 10.1371/journal.pmed.1003664] [Medline: 34111122]
23. Weitz E, Kleiboer A, van Straten A, Cuijpers P. The effects of psychotherapy for depression on anxiety symptoms: a meta-analysis. Psychol Med 2018 Jan 24;48(13):2140-2152. [doi: 10.1017/s0033291717003622]
24. Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, et al. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. JAMA Psychiatry 2017 Sep 01;74(9):875-884 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.2164] [Medline: 28768327]
25. Murray LK, Dorsey S, Haroz E, Lee C, Alsiary MM, Haydary A, et al. A common elements treatment approach for adult mental health problems in low- and middle-income countries. Cogn Behav Pract 2014 May;21(2):111-123 [FREE Full text] [doi: 10.1016/j.cbpra.2013.06.005] [Medline: 25620867]
26. Gaebel W, Kowitz S, Fritze J, Zielasek J. Use of health care services by people with mental illness: secondary data from three statutory health insurers and the German Statutory Pension Insurance Scheme. Dtsch Arztebl Int 2013 Nov 22;110(47):799-808 [FREE Full text] [doi: 10.3238/arztebl.2013.0799] [Medline: 24314623]
27. Kostyuk GP, Masyakin AV, Burygina LA, Reverchuk IV. The problem of the diagnosis and treatment of depression by primary care physicians. Neurosci Behav Physi 2019 Dec 04;50(1):45-50. [doi: 10.1007/s11055-019-00867-0]
28. Wittchen HU, Kessler RC, Beesdo K, Krause P, Höfler M, Hoyer J. Generalized anxiety and depression in primary care: prevalence, recognition, and management. J Clin Psychiatry 2002;63 Suppl 8:24-34. [Medline: 12044105]
29. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
30. Newby JM, Twomey C, Yuan Li SS, Andrews G. Transdiagnostic computerised cognitive behavioural therapy for depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2016 Jul 15;199:30-41. [doi: 10.1016/j.jad.2016.03.018] [Medline: 27060430]
31. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
32. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2013 Aug 09;44(6):1303-1317. [doi: 10.1017/s0033291713001943]
33. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
34. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
35. Halfin A. Depression: the benefits of early and appropriate treatment. Am J Manag Care 2007 Nov;13(4 Suppl):S92-S97 [FREE Full text] [Medline: 18041868]
36. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Web-based cognitive behavior therapy for depression with and without telephone tracking in a national helpline: secondary outcomes from a randomized controlled trial. J Med Internet Res 2012 Jun 27;14(3):e68 [FREE Full text] [doi: 10.2196/jmir.1859] [Medline: 22738715]
37. Griffiths KM, Christensen H, Jorm AF, Evans K, Groves C. Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatising attitudes to depression: randomised controlled trial. Br J Psychiatry 2004 Oct;185:342-349. [doi: 10.1192/bjp.185.4.342] [Medline: 15458995]
38. Taylor-Rodgers E, Batterham PJ. Evaluation of an online psychoeducation intervention to promote mental health help seeking attitudes and intentions among young adults: randomised controlled trial. J Affect Disord 2014 Oct;168:65-71. [doi: 10.1016/j.jad.2014.06.047] [Medline: 25038293]

39. Christensen H, Leach LS, Barney L, Mackinnon AJ, Griffiths KM. The effect of web based depression interventions on self reported help seeking: randomised controlled trial [ISRCTN77824516]. BMC Psychiatry 2006 Apr 05;6(1). [doi: 10.1186/1471-244x-6-13]
40. Klein B, Richards JC, Austin DW. Efficacy of internet therapy for panic disorder. J Behav Ther Exp Psychiatry 2006 Sep;37(3):213-238. [doi: 10.1016/j.jbtep.2005.07.001] [Medline: 16126161]
41. Kravitz RL, Franks P, Feldman MD, Tancredi DJ, Slee CA, Epstein RM, et al. Patient engagement programs for recognition and initial treatment of depression in primary care: a randomized trial. JAMA 2013 Nov 06;310(17):1818-1828 [FREE Full text] [doi: 10.1001/jama.2013.280038] [Medline: 24193079]
42. Domhardt M, Steubl L, Boettcher J, Buntrock C, Karyotaki E, Ebert DD, et al. Mediators and mechanisms of change in internet- and mobile-based interventions for depression: a systematic review. Clin Psychol Rev 2021 Feb;83:101953. [doi: 10.1016/j.cpr.2020.101953] [Medline: 33422841]
43. Kerber A, Beintner I, Burchert S, Knaevelsrud C. Does app-based unguided self-management improve mental health literacy, patient empowerment and access to care for people with mental health impairments? study protocol for a randomised controlled trial. BMJ Open 2021 Jul 15;11(7):e049688 [FREE Full text] [doi: 10.1136/bmjopen-2021-049688] [Medline: 34266843]
44. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
45. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
46. Fogliati VJ, Terides MD, Gandy M, Staples LG, Johnston L, Karin E, et al. Psychometric properties of the Mini-Social Phobia Inventory (Mini-SPIN) in a large online treatment-seeking sample. Cogn Behav Ther 2016 Apr 04;45(3):236-257. [doi: 10.1080/16506073.2016.1158206] [Medline: 27046641]
47. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [doi: 10.1097/00006842-200203000-00008] [Medline: 11914441]
48. Crönlein T, Langguth B, Popp R, Lukesch H, Pieh C, Hajak G, et al. Regensburg Insomnia Scale (RIS): a new short rating scale for the assessment of psychological symptoms and sleep in insomnia; study design: development and validation of a new short self-rating scale in a sample of 218 patients suffering from insomnia and 94 healthy controls. Health Qual Life Outcomes 2013 Apr 22;11:65 [FREE Full text] [doi: 10.1186/1477-7525-11-65] [Medline: 23601161]
49. Dias P, Campos L, Almeida H, Palha F. Mental health literacy in young adults: adaptation and psychometric properties of the mental health literacy questionnaire. Int J Environ Res Public Health 2018 Jun 23;15(7):1318 [FREE Full text] [doi: 10.3390/ijerph15071318] [Medline: 29937509]
50. Houle J, Gascon-Depatie M, Bélanger-Dumontier G, Cardinal C. Depression self-management support: a systematic review. Patient Educ Couns 2013 Jun;91(3):271-279. [doi: 10.1016/j.pec.2013.01.012] [Medline: 23414831]
51. Morgan AJ, Jorm AF. Self-help strategies that are helpful for sub-threshold depression: a Delphi consensus study. J Affect Disord 2009 May;115(1-2):196-200. [doi: 10.1016/j.jad.2008.08.004] [Medline: 18799220]
52. van Grieken RA, Kirkenier AC, Koeter MW, Nabitz UW, Schene AH. Patients' perspective on self-management in the recovery from depression. Health Expect 2015 Oct 02;18(5):1339-1348 [FREE Full text] [doi: 10.1111/hex.12112] [Medline: 23910797]
53. Villaggi B, Provencher H, Coulombe S, Meunier S, Radziszewski S, Hudon C, et al. Self-management strategies in recovery from mood and anxiety disorders. Glob Qual Nurs Res 2015 Sep 21;2:2333393615606092 [FREE Full text] [doi: 10.1177/2333393615606092] [Medline: 28462317]
54. Saltman RB. Patient choice and patient empowerment in northern European health systems: a conceptual framework. Int J Health Serv 1995 Jan 01;24(2):201-229. [doi: 10.2190/8wmp-rr2k-abm7-nvnh]
55. Mackenzie CS, Knox VJ, Gekoski WL, Macaulay HL. An adaptation and extension of the attitudes toward seeking professional psychological help scale. J Appl Social Pyschol 2004 Nov;34(11):2410-2433. [doi: 10.1111/j.1559-1816.2004.tb01984.x]
56. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
57. Zimmermann J, Kerber A, Rek K, Hopwood CJ, Krueger RF. A brief but comprehensive review of research on the alternative DSM-5 model for personality disorders. Curr Psychiatry Rep 2019 Aug 13;21(9):92. [doi: 10.1007/s11920-019-1079-z] [Medline: 31410586]
58. Bach B, Kerber A, Aluja A, Bastiaens T, Keeley JW, Claes L, et al. International assessment of DSM-5 and ICD-11 personality disorder traits: toward a common nosology in DSM-5.1. PsyArXiv Preprints. Preprint posted online January 11, 2020 2023 [FREE Full text] [doi: 10.31234/osf.io/qcwsb]
59. Ehrenthal JC, Dinger U, Schauenburg H, Horsch L, Dahlbender RW, Gierk B. Development of a 12-item version of the OPD-Structure Questionnaire (OPD-SQS). Article in German. Z Psychosom Med Psychother 2015 Sep 01;61(3):262-274. [doi: 10.13109/zptm.2015.61.3.262] [Medline: 26388057]

60. Kerber A, Schultze M, Müller S, Rühling RM, Wright AG, Spitzer C, et al. Development of a short and ICD-11 compatible measure for DSM-5 maladaptive personality traits using ant colony optimization algorithms. PsyArXiv Preprints. Preprint posted online on November 17, 2019 2023 [FREE Full text] [doi: 10.31234/osf.io/rsw54]
61. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. Patient 2014 Nov 23;7(1):85-96 [FREE Full text] [doi: 10.1007/s40271-013-0036-x] [Medline: 24271592]
62. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol 2019 Jun;110:63-73 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.02.016] [Medline: 30878639]
63. Cornelisz I, Cuijpers P, Donker T, van Klaveren C. Addressing missing data in randomized clinical trials: a causal inference perspective. PLoS One 2020 Jul 6;15(7):e0234349 [FREE Full text] [doi: 10.1371/journal.pone.0234349] [Medline: 32628678]
64. Kleinke K, Reinecke J. Multiple imputation of incomplete zero-inflated count data. Stat Neerl 2013 May 17;67(3):311-336. [doi: 10.1111/stan.12009]
65. Grund S, Lüdtke O, Robitzsch A. Pooling ANOVA results from multiply imputed datasets. Methodology 2016 Jul;12(3):75-88. [doi: 10.1027/1614-2241/a000111]
66. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). J Affect Disord 2020 Mar 15;265:395-401. [doi: 10.1016/j.jad.2020.01.032] [Medline: 32090765]
67. Toussaint A, Kroenke K, Baye F, Lourens S. Comparing the patient health questionnaire - 15 and the somatic symptom scale - 8 as measures of somatic symptom burden. J Psychosom Res 2017 Oct;101:44-50. [doi: 10.1016/j.jpsychores.2017.08.002] [Medline: 28867423]
68. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [doi: 10.1097/00005650-200412000-00006] [Medline: 15550799]
69. Deegan PE. The importance of personal medicine: a qualitative study of resilience in people with psychiatric disabilities. Scand J Public Health Suppl 2005 Oct 25;66(66_suppl):29-35. [doi: 10.1080/14034950510033345] [Medline: 16214720]
70. Russell SJ, Browne JL. Staying well with bipolar disorder. Aust N Z J Psychiatry 2005 Mar;39(3):187-193. [doi: 10.1080/j.1440-1614.2005.01542.x] [Medline: 15701069]
71. Todd NJ, Jones SH, Lobban FA. "Recovery" in bipolar disorder: how can service users be supported through a self-management intervention? a qualitative focus group study. J Ment Health 2012 Apr;21(2):114-126. [doi: 10.3109/09638237.2011.621471] [Medline: 22142324]
72. Young SL, Ensing DS. Exploring recovery from the perspective of people with psychiatric disabilities. Psychiatr Rehabil J 1999;22(3):219-231. [doi: 10.1037/h0095240]
73. Lean M, Fornells-Ambrojo M, Milton A, Lloyd-Evans B, Harrison-Stewart B, Yesufu-Udechuku A, et al. Self-management interventions for people with severe mental illness: systematic review and meta-analysis. Br J Psychiatry 2019 May;214(5):260-268 [FREE Full text] [doi: 10.1192/bjp.2019.54] [Medline: 30898177]
74. Morgan AJ, Mackinnon AJ, Jorm AF. Behavior change through automated e-mails: mediation analysis of self-help strategy use for depressive symptoms. Behav Res Ther 2013 Feb;51(2):57-62 [FREE Full text] [doi: 10.1016/j.brat.2012.11.002] [Medline: 23262114]
75. Morton E, Michalak EE, Hole R, Buzwell S, Murray G. 'Taking back the reins' - a qualitative study of the meaning and experience of self-management in bipolar disorder. J Affect Disord 2018 Mar 01;228:160-165. [doi: 10.1016/j.jad.2017.12.018] [Medline: 29248822]
76. Terides MD, Dear BF, Fogliati VJ, Gandy M, Karin E, Jones MP, et al. Increased skills usage statistically mediates symptom reduction in self-guided internet-delivered cognitive-behavioural therapy for depression and anxiety: a randomised controlled trial. Cogn Behav Ther 2018 Jan;47(1):43-61. [doi: 10.1080/16506073.2017.1347195] [Medline: 28724338]
77. Zoun MH, Koekkoek B, Sinnema H, van der Feltz-Cornelis CM, van Balkom AJ, Schene AH, et al. Effectiveness of a self-management training for patients with chronic and treatment resistant anxiety or depressive disorders on quality of life, symptoms, and empowerment: results of a randomized controlled trial. BMC Psychiatry 2019 Jan 28;19(1):46 [FREE Full text] [doi: 10.1186/s12888-019-2013-y] [Medline: 30691421]
78. Sterling EW, von Esenwein SA, Tucker S, Fricks L, Druss BG. Integrating wellness, recovery, and self-management for mental health consumers. Community Ment Health J 2010 Apr 23;46(2):130-138. [doi: 10.1007/s10597-009-9276-6] [Medline: 20033488]
79. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010 Nov;61(11):1099-1105. [doi: 10.1176/ps.2010.61.11.1099] [Medline: 21041348]
80. MacLean S, Corsi DJ, Litchfield S, Kucharski J, Genise K, Selaman Z, et al. Coach-facilitated web-based therapy compared with information about web-based resources in patients referred to secondary mental health care for depression: randomized controlled trial. J Med Internet Res 2020 Jun 09;22(6):e15001 [FREE Full text] [doi: 10.2196/15001] [Medline: 32515740]

81. Baumeister H, Paganini S, Sander LB, Lin J, Schlicker S, Terhorst Y, et al. Effectiveness of a guided internet- and mobile-based intervention for patients with chronic back pain and depression (WARD-BP): a multicenter, pragmatic randomized controlled trial. Psychother Psychosom 2021;90(4):255-268. [doi: 10.1159/000511881] [Medline: 33321501]
82. Oehler C, Görges F, Rogalla M, Rummel-Kluge C, Hegerl U. Efficacy of a guided web-based self-management intervention for depression or dysthymia: randomized controlled trial with a 12-month follow-up using an active control condition. J Med Internet Res 2020 Jul 14;22(7):e15361 [FREE Full text] [doi: 10.2196/15361] [Medline: 32673233]
83. Axelsson E, Andersson E, Ljótsson B, Björkander D, Hedman-Lagerlöf M, Hedman-Lagerlöf E. Effect of internet vs face-to-face cognitive behavior therapy for health anxiety: a randomized noninferiority clinical trial. JAMA Psychiatry 2020 Sep 01;77(9):915-924 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.0940] [Medline: 32401286]
84. Reins JA, Boß L, Lehr D, Berking M, Ebert DD. The more I got, the less I need? efficacy of internet-based guided self-help compared to online psychoeducation for major depressive disorder. J Affect Disord 2019 Mar 01;246:695-705. [doi: 10.1016/j.jad.2018.12.065] [Medline: 30611913]
85. Han B, McKeon R, Gfroerer J. Suicidal ideation among community-dwelling adults in the United States. Am J Public Health 2014 Mar;104(3):488-497. [doi: 10.2105/AJPH.2013.301600] [Medline: 24432951]
86. Karyotaki E, Kemmeren L, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, et al. Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? an individual participant data meta-analysis. Psychol Med 2018 Nov 15;48(15):2456-2466 [FREE Full text] [doi: 10.1017/S0033291718000648] [Medline: 29540243]
87. Ebert DD, Donkin L, Andersson G, Andrews G, Berger T, Carlbring P, et al. Does internet-based guided-self-help for depression cause harm? an individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. Psychol Med 2016 Oct;46(13):2679-2693 [FREE Full text] [doi: 10.1017/S0033291716001562] [Medline: 27649340]
88. Ho FY, Chan CS, Lo WY, Leung JC. The effect of self-help cognitive behavioral therapy for insomnia on depressive symptoms: an updated meta-analysis of randomized controlled trials. J Affect Disord 2020 Mar 15;265:287-304. [doi: 10.1016/j.jad.2020.01.062] [Medline: 32090753]
89. van Gils A, Schoevers RA, Bonvanie IJ, Gelauff JM, Roest AM, Rosmalen JG. Self-help for medically unexplained symptoms: a systematic review and meta-analysis. Psychosom Med 2016;78(6):728-739. [doi: 10.1097/PSY.0000000000000325] [Medline: 27187850]
90. Clark DM, Canvin L, Green J, Layard R, Pilling S, Janecka M. Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. Lancet 2018 Feb 17;391(10121):679-686 [FREE Full text] [doi: 10.1016/S0140-6736(17)32133-5] [Medline: 29224931]
91. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
92. Kerber A, Schaeuffele C, Krieger T, Urech A, Riper H, Berger T, et al. Differential effects of psychological interventions in online and face-to-face settings on DSM-5 and ICD-11 maladaptive trait domains: an exploratory pilot study. Front Psychiatry 2021 Jun 14;12:648367 [FREE Full text] [doi: 10.3389/fpsyt.2021.648367] [Medline: 34194347]
93. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
94. Furukawa TA, Karyotaki E, Suganuma A, Pompoli A, Ostinelli EG, Cipriani A, et al. Dismantling, personalising and optimising internet cognitive-behavioural therapy for depression: a study protocol for individual participant data component network meta-analysis. BMJ Open 2019 Feb 22;8(11):e026137 [FREE Full text] [doi: 10.1136/bmjopen-2018-026137] [Medline: 30798295]
95. Delgadillo J, Ali S, Fleck K, Agnew C, Southgate A, Parkhouse L, et al. Stratified care vs stepped care for depression: a cluster randomized clinical trial. JAMA Psychiatry 2022 Feb 01;79(2):101-108 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.3539] [Medline: 34878526]
96. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726 [FREE Full text] [doi: 10.1017/S0033291715000665] [Medline: 25881626]
97. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Focus (Am Psychiatr Publ) 2018 Oct;16(4):420-429 [FREE Full text] [doi: 10.1176/appi.focus.16407] [Medline: 32021580]

Abbreviations

ANCOVA: analysis of covariance
CAU: care as usual

CG: control group
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
FMI: fraction of missing information
GAD-7: Generalized Anxiety Disorder–7
HiTOP: Hierarchical Taxonomy of Psychopathology
IG: intervention group
ITT: intention-to-treat
MCID: minimally clinical important difference
MHPSS: mental health–related patient empowerment and self-management skills
PHQ-15: Patient Health Questionnaire–15
PHQ-9: Patient Health Questionnaire–9
PP: per-protocol
RCT: randomized controlled trial
RF: random forest
RFLB: Random Forest Lee Bound
RIS: Regensburg Insomnia Scale

Edited by J Torous; submitted 15.12.22; peer-reviewed by J Bullis, N Titov, I Bell; comments to author 07.06.23; revised version received 01.08.23; accepted 04.08.23; published 06.11.23.

Please cite as:
Kerber A, Beintner I, Burchert S, Knaevelsrud C
Effects of a Self-Guided Transdiagnostic Smartphone App on Patient Empowerment and Mental Health: Randomized Controlled Trial
JMIR Ment Health 2023;10:e45068
URL: https://mental.jmir.org/2023/1/e45068
doi:10.2196/45068
PMID:37930749

Original Paper

Suicide Risk Assessments Through the Eyes of ChatGPT-3.5 Versus ChatGPT-4: Vignette Study

Inbar Levkovich[1], PhD; Zohar Elyoseph[2,3], PhD

[1]Oranim Academic College, Faculty of Graduate Studies, Kiryat Tivon, Israel

[2]Department of Psychology and Educational Counseling, The Center for Psychobiological Research, Max Stern Yezreel Valley College, Emek Yezreel, Israel

[3]Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, United Kingdom

Corresponding Author:
Zohar Elyoseph, PhD
Department of Psychology and Educational Counseling
The Center for Psychobiological Research
Max Stern Yezreel Valley College
Hatena 14b Kiryat Tivon
Emek Yezreel, 3650414
Israel
Phone: 972 54 783 6088
Email: Zohare@yvc.ac.il

Abstract

Background: ChatGPT, a linguistic artificial intelligence (AI) model engineered by OpenAI, offers prospective contributions to mental health professionals. Although having significant theoretical implications, ChatGPT's practical capabilities, particularly regarding suicide prevention, have not yet been substantiated.

Objective: The study's aim was to evaluate ChatGPT's ability to assess suicide risk, taking into consideration 2 discernable factors—perceived burdensomeness and thwarted belongingness—over a 2-month period. In addition, we evaluated whether ChatGPT-4 more accurately evaluated suicide risk than did ChatGPT-3.5.

Methods: ChatGPT was tasked with assessing a vignette that depicted a hypothetical patient exhibiting differing degrees of perceived burdensomeness and thwarted belongingness. The assessments generated by ChatGPT were subsequently contrasted with standard evaluations rendered by mental health professionals. Using both ChatGPT-3.5 and ChatGPT-4 (May 24, 2023), we executed 3 evaluative procedures in June and July 2023. Our intent was to scrutinize ChatGPT-4's proficiency in assessing various facets of suicide risk in relation to the evaluative abilities of both mental health professionals and an earlier version of ChatGPT-3.5 (March 14 version).

Results: During the period of June and July 2023, we found that the likelihood of suicide attempts as evaluated by ChatGPT-4 was similar to the norms of mental health professionals (n=379) under all conditions (average *Z* score of 0.01). Nonetheless, a pronounced discrepancy was observed regarding the assessments performed by ChatGPT-3.5 (May version), which markedly underestimated the potential for suicide attempts, in comparison to the assessments carried out by the mental health professionals (average *Z* score of –0.83). The empirical evidence suggests that ChatGPT-4's evaluation of the incidence of suicidal ideation and psychache was higher than that of the mental health professionals (average *Z* score of 0.47 and 1.00, respectively). Conversely, the level of resilience as assessed by both ChatGPT-4 and ChatGPT-3.5 (both versions) was observed to be lower in comparison to the assessments offered by mental health professionals (average *Z* score of –0.89 and –0.90, respectively).

Conclusions: The findings suggest that ChatGPT-4 estimates the likelihood of suicide attempts in a manner akin to evaluations provided by professionals. In terms of recognizing suicidal ideation, ChatGPT-4 appears to be more precise. However, regarding psychache, there was an observed overestimation by ChatGPT-4, indicating a need for further research. These results have implications regarding ChatGPT-4's potential to support gatekeepers, patients, and even mental health professionals' decision-making. Despite the clinical potential, intensive follow-up studies are necessary to establish the use of ChatGPT-4's capabilities in clinical practice. The finding that ChatGPT-3.5 frequently underestimates suicide risk, especially in severe cases, is particularly troubling. It indicates that ChatGPT may downplay one's actual suicide risk level.

(JMIR Ment Health 2023;10:e51232) doi:10.2196/51232

KEYWORDS

artificial intelligence; ChatGPT; diagnosis; psychological assessment; psychological; suicide risk; risk assessment; text vignette; NLP; natural language processing; suicide; suicidal; risk; assessment; vignette; vignettes; assessments; mental; self-harm

Introduction

Background

Large language models (LLMs), a subset of natural language processing (NLP) models, are trained with ample textual data to generate advanced language predictions [1]. Recently, a ChatGPT-based agent from OpenAI in California gained significant internet attention due to its ability to produce human-like text from varied prompts [2,3]. In fact, since its launch in November 2022, ChatGPT swiftly gained millions of users [1]. Its ability to handle complex tasks and generate human-like language marks a breakthrough in artificial intelligence (AI) and natural language processing [3,4]. However, despite extensive study in academia, its applications in applied psychology [5-7], particularly its efficacy in addressing critical mental health issues such as suicide prevention, remain unclear.

Suicide represents a significant global health issue and a leading cause of mortality [8,9]. Despite numerous comprehensive studies over the past few decades, the issue of suicide risk assessment remains unresolved in the mental health field [10,11]. The currently used questionnaires and clinical evaluations have yet to overcome notable psychometric challenges [12]. Another difficulty is the lack of communities' sufficient access to suicide risk assessment [13]. In an extensive review of linguistic markers linked to suicidal tendencies, 75 studies involving 279,032 individuals were analyzed. Suicidal ideation was associated with increased use of intensifiers and superlatives, while suicidal actions correlated with more pronouns, varying verb usage, and other specific linguistic patterns [14]. Fernandes et al [15] used a machine learning algorithm on electronic health records to detect "suicidal thoughts" or "suicide attempt" mentions. Although successful, the algorithm's ability to predict future suicidal actions is yet uncertain. Aladağ et al [16] used machine learning on 10,000 forum posts classified by a clinician for suicidal content. Using metrics from the Linguistic Inquiry and Word Count software, they achieved high prediction accuracy, but did not specify the key metrics. Similarly, Tadesse et al [17] analyzed Reddit posts using NLP, comparing those with and without suicidal content without clinical validation.

A previous study evaluated the potential of ChatGPT-3.5 (March 14, 2023, version), a linguistic AI model, versus mental health professionals in assessing suicide risk [6]. The results showed that ChatGPT-3.5 (March 14 version) generally underestimated the risk of suicide, which raised concerns about its reliability for such assessments. Thus, our previous study indicated that until further evidence could support its accuracy, mental health professionals should consider ChatGPT's assessments to be nonprofessional [6]. However, considering recent developments, including the launch of ChatGPT-4 and the release of a new version of ChatGPT-3.5 (May 24, 2023, versions of both), we aimed in the previous study to examine whether these new versions exhibited improvements over their predecessors.

In recent times, the potential of AI to augment mental health services has been the subject of intense scrutiny. The envisaged applications range from aiding diagnostics [18], to streamlining administrative tasks that would afford clinicians more patient time [19], to enhancing social motivation and attentional performance via AI-powered games that foster mental health [20]. A recent review [21] shed light on the prospective utility of AI-driven chatbots in the mental health sphere.

In terms of linguistic diversity, ChatGPT-4 exhibits enhanced multilingual abilities [22,23] compared to ChatGPT-3.5. Specifically, it represents a substantial advance in model progression, with amplified capabilities such as multilingual expertise, extended context length, and image processing, thus presenting intriguing possibilities across diverse fields [24]. Yet the cost of ChatGPT-4 and its inherent limitations underline the necessity of scrutinizing a specific application before choosing it [22]. It should be noted that ChatGPT-4 has been acknowledged for its superior performance relative to its antecedents [25]. In a study evaluating ChatGPT-4's performance on the Ophthalmic Knowledge Assessment Program examination versus its predecessor, ChatGPT-3.5, the results demonstrated that ChatGPT-4 notably outperformed ChatGPT-3.5 (81% vs 57%), indicating progress in medical knowledge evaluation [26]. Another study contrasting the performance of ChatGPT-3.5 and ChatGPT-4 on the Japanese Medical Licensing Examination revealed that ChatGPT-4 surpassed ChatGPT-3.5 in terms of accuracy, particularly with regard to general, clinical, and clinical sentence queries [27].

This Study

The capacity of clinicians to identify signs of potential suicide is paramount for the administration of proper crisis management and suicide intervention tactics, particularly during times of severe crisis [28]. Given the seriousness of the issue, a substantial commitment toward evaluating suicide risk is vital [29-31].

AI, in theory, could assist gatekeepers in their decision-making processes and enhance the efficacy of formal psychometric tools and clinical evaluations in predicting suicidal behavior. Current methods often fall short in their predictive capabilities [29,32,33].

In this study, we evaluated the ability of ChatGPT-3.5 and ChatGPT-4 (May 24 versions) to identify the risk of suicide based on the interpersonal theory of suicide (ITS), an established and empirically backed theoretical model proposed by Joiner et al [34,35] for assessing suicide risk and identifying associated factors. We performed these evaluations on the same day. We specifically evaluated how the 2 key dimensions of ITS—perceived burdensomeness and thwarted belongingness—influence therapists' perceptions and assessments of suicidal ideation and behavior. Perceived

burdensomeness refers to a psychological state wherein individuals feel their existence imposes a burden on their family, friends, or society. This perception leads individuals to believe their death might be more beneficial than their continued existence [34]. Thwarted belongingness, as defined by Van Orden et al [35], alludes to feelings of alienation from others and emphasizes the distressing sensation of exclusion from one's family, friends, or other valued groups. This construct plays a critical role in suicide, as belongingness represents a fundamental psychological need, and failure to meet this need can lead to profound distress [36]. In a previous study, the tendency of ChatGPT-3.5 (March 14 version) to consistently underestimate the risk of suicide attempts, compared to estimates by professionals in the mental health field, was unveiled [6]. This trend was maintained irrespective of the influence of resilience, perceived burdensomeness, or thwarted belongingness on the evaluation of suicidal ideation and the potential risk of a suicide attempt. Alarmingly, we observed that the extent of underestimation of suicide risk by ChatGPT-3.5 (March 14 version) was at its highest when the severity of the case was at its maximum. This striking discovery poses a significant point of concern.

The Objectives of This Study

We aimed to (1) assess the ability of ChatGPT-3.5 and ChatGPT-4 (May 24 versions) to evaluate suicidal behavior and risk factors across 2 identifiable variables—perceived burdensomeness and thwarted belongingness—compared to the ability of mental health professionals and to the earlier version of ChatGPT-3.5 (March 14 version) and (2) evaluate whether ChatGPT-4 (May 24 version) evaluates suicide risk more accurately than ChatGPT-3.5 (May 24 version).

Methods

AI Procedure

Using ChatGPT-4 and the free version of ChatGPT-3.5 (both May 24), we conducted 3 evaluations during the period of June and July 2023 to examine the ability of these two versions of ChatGPT to evaluate various aspects of suicide risk compared to the performance of mental health professionals and to the previous version of ChatGPT-3.5 (March 14).

Input Source

In this investigation, we adopted the text vignette methodology as elucidated by Bresin et al [37] to examine the observer's perspective on suicidal acts. A specific vignette from Levi-Belz and Gamliel [38] was entered into the ChatGPT interface. The chosen vignette contained general background details and 2 additional paragraphs; one paragraph portrayed either low or high perceived burdensomeness and the other either low or high thwarted belongingness, for a total of 4 possible combinations and 4 vignettes (Multimedia Appendix 1). Each of the vignettes was presented 3 times to both the ChatGPT-4 and ChatGPT-3.5 models, each time in a new tab.

Measures

After each of the 4 vignette conditions, that is, (1) low burdensomeness and low belongingness, (2) low burdensomeness and high belongingness, (3) high burdensomeness and low belongingness, and (4) high burdensomeness and high belongingness, we submitted 6 questions to ChatGTP-4 and 3.5 (May 24 versions), taken from Levi-Belz and Gamliel [38]. Each question was categorized into one of four subscales: (1) psychache, that is, psychological pain (Q1: "What do you think is Jane's level of distress?" and Q2: "What do you think is Jane's level of mental pain?"), (2) suicidal ideation (Q3: "What is the likelihood Jane will have suicidal thoughts?"), (3) risk of suicide attempt (Q4: "What is the likelihood Jane will attempt suicide?" and Q5: "What is the likelihood Jane will make a serious suicide attempt?"), and (4) resilience (Q6: "How do you assess Jane's level of resilience?"). All 6 questions used a 7-point Likert scale, estimating intensity (Q1, Q2, Q6) and likelihood (Q3-Q5), ranging from 1 (very slight) to 7 (very high). Multimedia Appendix 2 shows examples of ChatGPT answers to the questions.

Scoring

The performance of ChatGPT-4 and ChatGPT-3.5 (May 24 versions) was scored and compared to the results of professional mental health professionals, which were previously published in accordance with Levi-Belz and Gamliel [38]. Subsequently, we compared the ChatGPT scores to the norms of 379 mental health professionals (n=80, 21% male; mean age 36, SD 8.8 years). Of these, 53 were graduate students in psychology, 266 held a master's degree, and 60 held a PhD. Regarding their professional roles, 43 participants were supervisors, 108 were certified experts, 128 were interns, and 100 either had not begun their internship or were in professions that did not require an internship. The majority of the sample (n=318, 84%) were currently practicing, whereas the remaining participants had previously worked in the mental health field but were currently not working [38].

Statistical Analysis

The data are presented as the mean (SD) scores of the first, second, and third evaluations of ChatGPT. Percentage and *Z* scores were used to evaluate the differences between the different ChatGPT versions' performance and the norms of the mental health professionals reported by Levi-Belz and Gamliel [38].

Results

Overview

Table 1 depicts the ChatGPT-4 and 3.5 (May 24 versions) mean (SD) scores for all four conditions, that is, (1) low burdensomeness and low belongingness, (2) low burdensomeness and high belongingness, (3) high burdensomeness and low belongingness, and (4) high burdensomeness and high belongingness, for the four dependent variables: (1) psychache, (2) suicidal ideation, (3) risk of suicide attempt, and (4) resilience, compared to the norms of the health professionals reported by Levi-Belz and Gamliel [38] and the results of ChatGPT-3.5 (March 14 version) reported by Elyoseph and Levkovich [6].

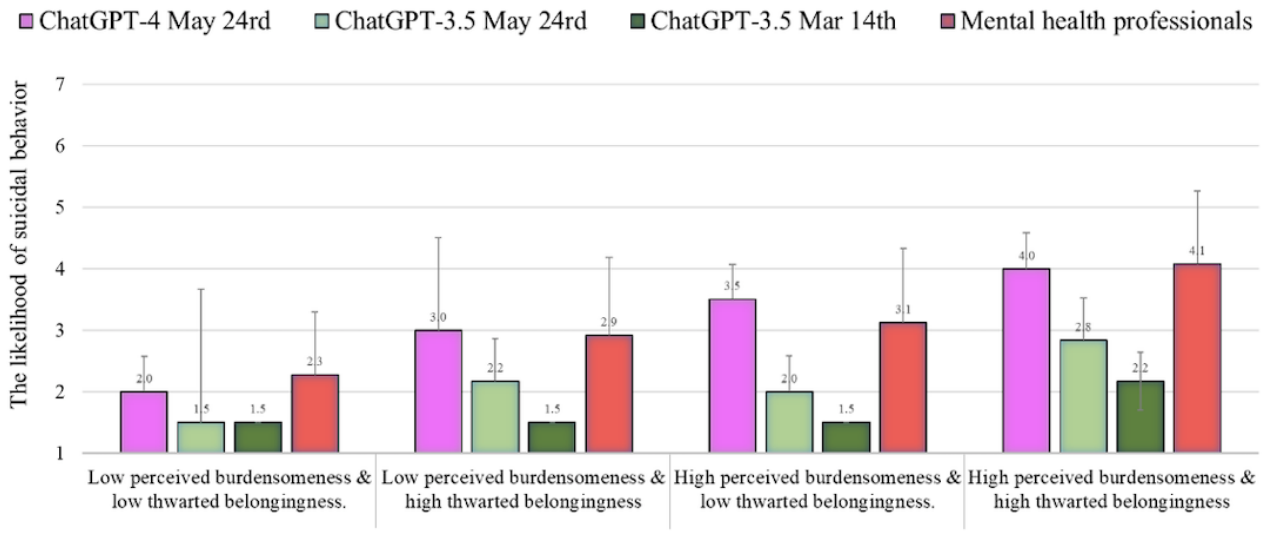
Table 1. Descriptive statistics of the mental health professionals (reported by Levi-Belz and Gamlie [38]), ChatGTP-4 (May 24), ChatGTP-3.5 (May 24), and ChatGTP-3.5 (March 14) for the 4 dependent variables (risk for suicide attempt, suicidal ideation, psychache, and resilience) in the 4 conditions.

Dependent variables	Low perceived burdensomeness and low thwarted belongingness, mean score (SD)	Low perceived burdensomeness and high thwarted belongingness, mean score (SD)	High perceived burdensomeness and low thwarted belongingness, mean score (SD)	High perceived burdensomeness and high thwarted belongingness, mean score (SD)
Risk for suicide attempt				
Mental health professionals	2.3 (1.0)	2.9 (1.3)	3.1 (1.2)	4.1 (1.2)
ChatGPT 4 (May)	2.0 (0.6)	3.0 (1.5)	3.5 (0.6)	4.0 (0.6)
ChatGPT 3.5 (May)	1.5 (0.8)	2.2 (0.7)	2.0 (0.6)	2.8 (0.7)
ChatGPT 3.5 (March)	1.5 (0.0)	1.5 (0.0)	1.5 (0.0)	2.7 (0.5)
Suicidal ideation				
Mental health professionals	3.6 (1.3)	4.3 (1.4)	5.0 (1.3)	5.4 (1.1)
ChatGPT 4 (May)	4.0 (0.0)	4.6 (0.5)	6.0 (0.0)	6.0 (0.0)
ChatGPT 3.5 (May)	3.3 (0.5)	4.7 (0.5)	4.3 (0.5)	5.7 (0.5)
ChatGPT 3.5 (March)	4.0 (0.0)	4.0 (0.0)	4.0 (0.0)	5.3 (0.5)
Psychache				
Mental health professionals	5.5 (0.8)	5.9 (0.6)	5.9 (0.7)	6.2 (0.7)
ChatGPT 4 (May)	6.0 (0.0)	6.3 (0.5)	6.8 (0.4)	7.0 (0.0)
ChatGPT 3.5 (May)	6.0 (0.0)	6.0 (0.0)	6.2 (0.4)	6.3 (0.5)
ChatGPT 3.5 (March)	6.0 (0.0)	6.0 (0.0)	6.0 (0.0)	6.3 (0.2)
Resilience				
Mental health professionals	5.1 (0.8)	4.5 (1.0)	4.2 (1.0)	3.4 (1.2)
ChatGPT 4 (May)	4.7 (0.5)	4.0 (0.0)	2.7 (0.5)	2.3 (0.5)
ChatGPT 3.5 (May)	4.7 (0.5)	4.0 (0.0)	2.6 (0.5)	2.3 (0.5)
ChatGPT 3.5 (March)	3.7 (0.5)	3.3 (0.0)	3.0 (0.0)	3.0 (0.0)

Risk of Suicide Attempt

Figure 1 shows the level of risk of suicide attempts as assessed by ChatGPT-4 and 3.5 (May 24 versions) as compared to the norms of the health professionals reported by Levi-Belz and Gamliel [38] and the results of ChatGPT-3.5 (March 14 version) reported by Elyoseph and Levkovich [6]. The level of risk of suicide attempts evaluated by ChatGPT-4 was similar to that of the mental health professionals in all conditions (t_{86}=0.13-0.57, P=.56-.9; average Z score=+0.01, average Z score in absolute value=0.17). In contrast, ChatGPT-3.5 (May 24 version) provided a significantly lower assessment of the level of risk of suicide attempts than did the mental health professionals (percentile range 15-27; average Z score=–0.83; average Z score in absolute value=0.83). This underestimation is similar to what was found by the ChatGPT-3.5 March 14 version (percentile range 5-23; average Z score=–1.21; average Z score in absolute value=1.21), but to a lesser extent.

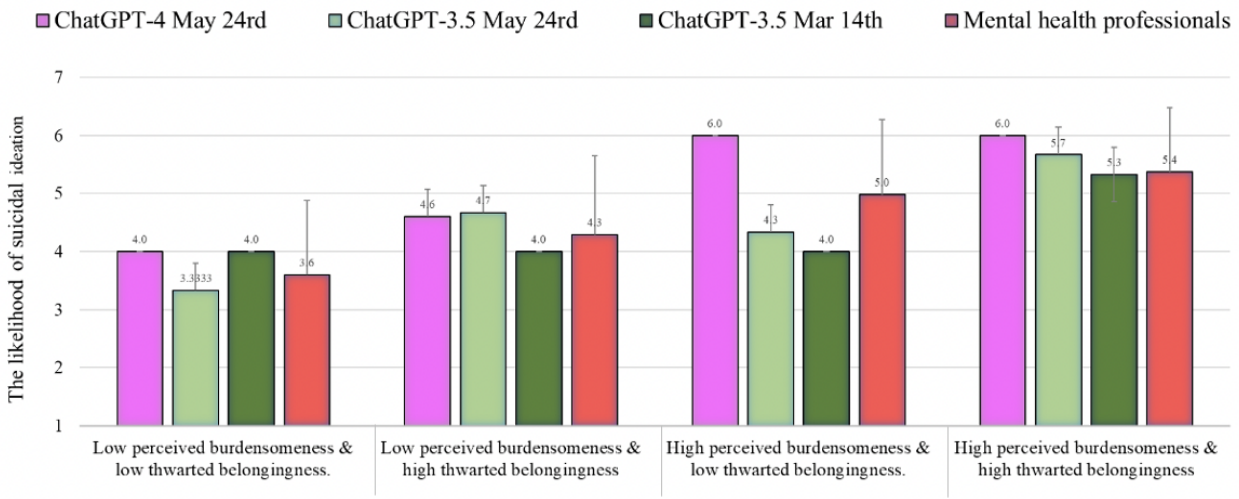
Figure 1. The level of risk of suicide attempts (mean, SD) assessed by ChatGPT-4 and 3.5 (May 24 versions) as compared to the norms of the health professionals and ChatGPT-3.5 (March 14 version).



Suicidal Ideation

Figure 2 shows the level of risk of suicidal ideation as assessed by ChatGPT-4 and 3.5 (May 24 versions) as compared to the norms of the health professionals reported by Levi-Belz and Gamliel [38] and the results of ChatGPT-3.5 (March 14 version) reported by Elyoseph and Levkovich [6]. ChatGPT-4 evaluated the likelihood of suicidal ideation higher than the mental health professionals in all conditions (percentile range 59-78; average *Z* score=0.47, average SD in absolute value=0.47). In contrast, ChatGPT-3.5 (May 24 version) assessed the likelihood of suicidal ideation quite similarly to the way in which the mental health professionals did (percentile range 31-61; average *Z* score=–0.04; average *Z* score in absolute value=0.31). This estimation was also quite similar to the results of the ChatGPT-3.5 March 14 version (percentile range 22-62; average *Z* score=–0.17; average *Z* score in absolute value=0.33).

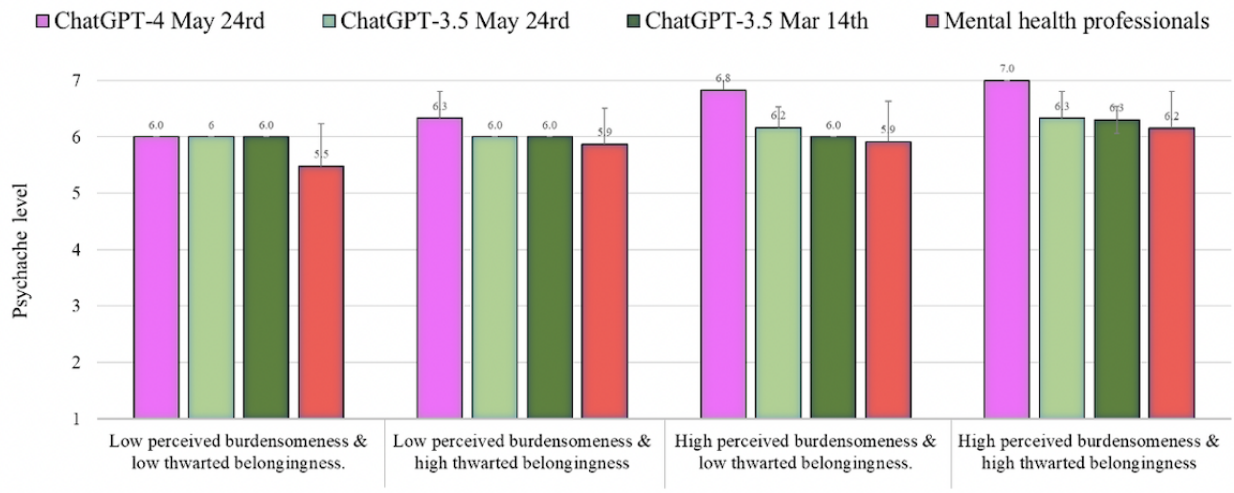
Figure 2. The likelihood of suicidal ideation (mean, SD) assessed by ChatGPT-4 and 3.5 (May 24 versions) as compared to the norms of the health professionals and ChatGPT-3.5 (March 14 version).



Psychache

Figure 3 shows the level of psychache as assessed by ChatGPT-4 and 3.5 (May 24 versions) as compared to the norms of the health professionals reported by Levi-Belz and Gamliel [38] and the results of ChatGPT-3.5 (March 14 version) reported by Elyoseph and Levkovich [6]. The level of psychache evaluated by ChatGPT-4 was higher than the level evaluated by the mental health professionals in all conditions (percentile range 76-90; average *Z* score=1.00, average SD in absolute value=1.00). In contrast, ChatGPT-3.5 (May 24 version) assessed the level of psychache quite similarly to how it was assessed by the mental health professionals (percentile range 55-76; average *Z* score=0.38; average *Z* score in absolute value=0.38). This estimation was also quite similar to the results of the ChatGPT-3.5 March 14 version (percentile range 22-62; average *Z* score=–0.17; average *Z* score in absolute value=0.31).

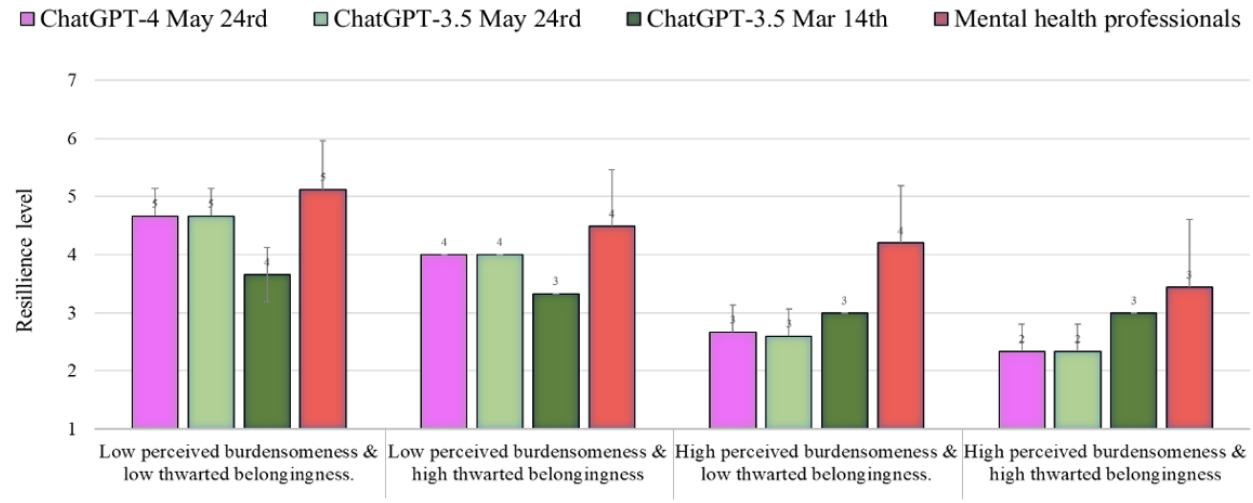
Figure 3. The level of psychache (mean, SD) assessed by ChatGPT-4 and 3.5 (May 24 versions) as compared to the norms of the health professionals and ChatGPT-3.5 (March 14 version).



Resilience

Figure 4 shows the level of resilience as assessed by ChatGPT-4 and 3.5 (May 24 versions) as compared to the norms of the health professionals reported by Levi-Belz and Gamliel [38] and the results of ChatGPT-3.5 (March 14 version) reported by Elyoseph and Levkovich [6]. The level of resilience evaluated by ChatGPT-4 was lower than the level evaluated by the mental health professionals in all conditions (percentile range 6-31; average *Z* score=–0.89, average SD in absolute value=0.89). Similarly, ChatGPT-3.5 (May 24 version) provided a lower assessment of resilience than did the mental health professionals (percentile range 5-31; average *Z* score=–0.90; average *Z* score in absolute value=0.90). This estimation was quite similar to the results of the ChatGPT-3.5 March 14 version (percentile range 4-30; average *Z* score=–1.13; average *Z* score in absolute value=1.13).

Figure 4. The level of resilience (mean, SD) assessed by ChatGPT-4 and 3.5 (May 24 versions) as compared to the norms of the health professionals and ChatGPT-3.5 (March 14 version).



Discussion

Principal Findings

In this study, we aimed to examine the efficacy of ChatGPT in conducting suicide risk assessments, with a particular focus on 2 discernible factors—sense of perceived burdensomeness and feelings of thwarted belongingness—across a period of 2 months. In addition, we wished to compare and analyze the precision of suicide risk assessment between two AI models, ChatGPT-4 and ChatGPT-3.5 (both the March 14 and May 24 versions), to ascertain which demonstrated superior performance. Parallels were found between the propensity for suicide attempts as assessed by ChatGPT-4 and as assessed by a sample of mental health professionals, under all conditions. However, there was a notable disparity between the evaluations made by ChatGPT-3.5 (both the March 14 and May 24 versions), which discernibly underestimated the risk of suicide attempts, and the assessments conducted by the mental health professionals.

This particular discovery carries substantial weight given the widespread and unrestricted use of AI chatbot technology by the general public. In a cross-sectional investigation, it was ascertained that 78.4% of the respondents exhibited a willingness to use ChatGPT for the purpose of self-diagnosis [39]. In a

variety of circumstances—including mental health evaluations, therapeutic consultations, medication management, and patient education—it has been substantiated that ChatGPT effectively delivers pertinent information and support to patients [40]. A recent review scrutinized the benefits conferred by ChatGPT and analogous LLMs in enhancing medical education, refining clinical decision-making processes, and propelling superior patient outcomes [41]. The findings of this study constitute preliminary evidence that ChatGPT-4 can provide an assessment similar to that provided by professionals on critical matters such as predicting the likelihood of suicidal behavior. The implications are vast; namely, ChatGPT-4 could serve as a decision-making support tool for clinicians and possibly offer professionals a second opinion. All of these possibilities necessitate continued research and development.

That said, we found that compared to the estimations made by mental health professionals across all conditions, ChatGPT-4 overestimated suicidal ideation. Surprisingly, the evaluation of suicidal ideation by ChatGPT-3.5 (May 24 version) was similar to the assessments made by the mental health professionals in all conditions as compared to the assessments of ChatGPT-4. Notably, these assessments by the May 24 version of ChatGPT-3.5 were consistent with the findings obtained from the March 14 version of the same ChatGPT-3.5 model.

The ability to competently use assessments identifying suicidal ideation in client interactions is of utmost importance for mental health professionals. It is estimated that approximately 1 in 4 such professionals will encounter client suicide [42]. The accurate discernment of suicidal ideation thus marks a fundamental stride toward its prevention and efficient management. Although no tool can guarantee absolute certainty, having access to reliable instruments for clinical assessment can prove indispensable for practitioners. Standardized assessments ought to be used in tandem with a clinical interview, mnemonic devices, and the evaluation of risk factors in order to conduct a comprehensive appraisal of patient risk [43]. Although it is evident that ChatGPT-4 showed a tendency to overestimate suicidal ideation compared to the mental health professionals, it did so by only a small margin. Currently, we do not know whether this gap resulted from the AI's overestimation or from mental health professionals' underestimation.

The findings of this study highlight that there was a higher degree of psychache, as assessed by ChatGPT-4, than that rendered by a group of mental health professionals in all observed conditions. By contrast, the assessment of psychache by the May 24 version of ChatGPT-3.5 aligned closely with the mental health professionals' assessment. The ChatGPT-3.5 (May 24 version) findings are congruent with the outcomes from its March version. A potential explanation for these findings might be that ChatGPT-4 tends to overestimate certain metrics. Furthermore, these findings suggest that different ChatGPT versions should still be used in a balanced manner, not exclusively, but in combination with the professional expertise of health care practitioners. This suggestion seems to align with the positions of professionals regarding the use of ChatGPT. In a survey assessing health care workers' interactions with ChatGPT, a significant majority of respondents (75.1%) expressed comfort with the idea of integrating ChatGPT into their health care practice, including in the aiding of medical decision-making (39.5%) [44].

Contrary to the other findings, this study indicates that the level of resilience as evaluated by ChatGPT-4 was lower than the level as assessed by the mental health professionals in all conditions. Similarly, ChatGPT-3.5 (March and May) provided a lower assessment of the level of resilience than did the mental health professionals. Despite the common use of the resilience concept, it should be noted that it has different definitions and a variety of measurement methods [45]. Resilience is a multifaceted concept shaped by various elements that span the individual, environmental, organizational, and cultural spheres [46]. This inherent complexity and multidimensionality renders resilience a challenging construct to operationalize and measure [46]. Experienced professionals, when confronted with case descriptions, are likely to consider an amalgamation of these factors, exhibiting a nuanced understanding that may be challenging to encapsulate fully in technological iterations. Thus, although the different versions of the technology offer significant value, their assessments must be contextualized within this wider understanding of resilience, highlighting the importance of a balanced approach that includes professional insight.

This research highlights the complexity of evaluating an individual's risk for suicide. The foregoing evidence demonstrates the potential advantages of using ChatGPT-4 to bolster clinical decision-making in the realm of suicide risk assessments among professionals [18,19,21]. Furthermore, ChatGPT-4 could play a crucial role in enhancing training and clinical procedures among mental health and medical professionals [26,27,47]. The easy access to ChatGPT and the possibility of reducing feelings of stigma may in the future drive the use of mental health assessment services by the general public.

However, the incorporation of ChatGPT into suicide risk detection mechanisms also presents a series of complexities. The reliability of ChatGPT predictions is intrinsically linked to the quality and demographic inclusivity of the training data [31]. Data biases or inadequate demographic representation could lead to erroneous predictions or exacerbate existing health disparities. Moreover, ChatGPT algorithms often function as opaque entities, obscuring the reasoning behind their predictive mechanisms. This lack of clarity can impede the development of trust and acceptance among users [48]. The deployment of ChatGPT for the identification of depression raises several ethical issues [48]. Ensuring data privacy and security is paramount, particularly given the sensitive nature of mental health information [49]. Research on the ethical facets of mental health within the broader populace suggests that participants express reservations about the widespread acceptance of AI and the implications of its capabilities for human welfare. Furthermore, there is apprehension regarding the potential for medical inaccuracies [50]. Moreover, there is expressed concern from patients regarding the potential discriminatory implications stemming from the use of AI [51]. The potential misuse that could exacerbate health disparities is an imperative issue. There is also the necessity of respecting patient autonomy, as AI can

spread misleading medical information or endorse unverified treatments, which may compromise a general practitioner's understanding of medical conditions [52]. While AI offers the unique capability to craft authentic patient scenarios—thus enhancing medical training—it is vital to remain cognizant of the risks associated with the misuse or misrepresentation of these tools [52]. Users must be thoroughly informed about the use and protection of their data. Importantly, ChatGPT should not replace human clinical judgment in the diagnostic process, but rather supplement it, thereby aiding professionals in making more-informed clinical decisions.

Limitations

This study, although informative, is not without limitations. To begin with, the assessment of the risk of suicide is based on a limited number of vignettes that centered on a woman in relatively stable condition of an age bracket not typically associated with high suicide risk and with no history of previous suicide attempts. Such a narrow representation does not adequately encapsulate the broad spectrum of suicide risk. As a result, we suggest that future research should expand to include a more diverse demographic. Vignettes should be created featuring varied demographics, such as male participants, individuals with psychiatric conditions, adolescents, and the elderly, to ensure a more holistic understanding of suicide risk. Secondly, our comparison of ChatGPT data was limited to a sample drawn from mental health professionals in Israel. To achieve a more comprehensive evaluation and to understand cross-cultural nuances in suicide risk assessment, we advocate for the exploration of therapists' assessments from various global contexts. Additionally, we used norms from a study about professionals as published by Levi-Belz and Gamliel [38]. We were unable to compare other statistical characteristics between the groups. We recommend further examination between an existing sample and ChatGPT. A further limitation of our study was that it relied heavily on the ITS as a theoretical basis. Although this theory provided a useful framework, it could potentially have constrained the breadth of our understanding. Moreover, our research delved into the complex domain of suicide risk assessment via AI; however, to gain more extensive insights, additional studies are required. Such studies should examine additional risk factors, integrate more expansive language models, assess data at different time intervals, and compare results with a broader array of clinical samples. Finally, given the rapid advances in the AI field, there is an inherent challenge in generalizing the results for stable, long-term abilities. Therefore, long-term studies are a necessity to keep pace with the evolving landscape and ensure a lasting understanding of suicide risk assessment.

Conclusion

In summary, this study yields critical knowledge on the potential of AI chatbots, particularly ChatGPT-4, to carry out suicide risk evaluations, while also highlighting the intricacies and subtleties entailed. Even though ChatGPT-4 overassessed suicidal thoughts, it showed an assessment accuracy on par with mental health experts in regard to suicidal behaviors. This research underlines the significance of AI tools such as ChatGPT serving as a supplement rather than as a substitute for professionals. The research also illustrates the rapid development of AI in the field of applied psychology and the need for research at multiple points in time and in regard to multiple versions in order to achieve reliable results.

Data Availability

The data that support the findings of this study are available from the authors upon reasonable request.

Authors' Contributions

IL and ZE contributed to conceptualization and manuscript review and editing. ZE contributed to methodology, formal analysis, and original draft preparation. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Vignette presented to ChatGPT-3.5 and ChatGPT-4.
[DOCX File , 14 KB - mental_v10i1e51232_app1.docx]

Multimedia Appendix 2
Examples of questions and ChatGPT answers.
[DOCX File , 305 KB - mental_v10i1e51232_app2.docx]

References

1. Alberts IL, Mercolli L, Pyka T, Prenosil G, Shi K, Rominger A, et al. Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be? Eur J Nucl Med Mol Imaging 2023 May;50(6):1549-1552 [FREE Full text] [doi: 10.1007/s00259-023-06172-w] [Medline: 36892666]

2. Egli A. ChatGPT, GPT-4, and other large language models - the next revolution for clinical microbiology? Clin Infect Dis 2023 Jul 03:ciad407. [doi: 10.1093/cid/ciad407] [Medline: 37399030]
3. Emenike ME, Emenike BU. Was this title generated by ChatGPT? Considerations for artificial intelligence text-generation software programs for chemists and chemistry educators. J Chem Educ 2023 Mar 20;100(4):1413-1418 [FREE Full text] [doi: 10.1021/acs.jchemed.3c00063]
4. Khalil M, Er E. Will ChatGPT get you caught? Rethinking plagiarism detection. ArXiv :1 Preprint posted online February 08, 2023. [FREE Full text] [doi: 10.35542/osf.io/fnh48]
5. Elyoseph Z, Hadar-Shoval D, Asraf K, Lvovsky M. ChatGPT outperforms humans in emotional awareness evaluations. Front Psychol 2023;14:1199058 [FREE Full text] [doi: 10.3389/fpsyg.2023.1199058] [Medline: 37303897]
6. Elyoseph Z, Levkovich I. Beyond human expertise: the promise and limitations of ChatGPT in suicide risk assessment. Front Psychiatry 2023;14:1213141 [FREE Full text] [doi: 10.3389/fpsyt.2023.1213141] [Medline: 37593450]
7. Guo B, Zhang X, Wang Z. How close is ChatGPT to human experts? comparison corpus, evaluation, and detection. arXiv Preprint posted online 18 Jan 2023. [FREE Full text] [doi: 10.48550/arXiv.2301.07597]
8. Lew B, Lester D, Mustapha FI, Yip P, Chen Y, Panirselvam RR, et al. Decriminalizing suicide attempt in the 21st century: an examination of suicide rates in countries that penalize suicide, a critical review. BMC Psychiatry 2022 Jun 23;22(1):424 [FREE Full text] [doi: 10.1186/s12888-022-04060-5] [Medline: 35739483]
9. Zalsman G. [Suicide: Epidemiology, etiology, treatment and prevention]. Harefuah 2019 Jul;158(7):468-472. [Medline: 31339248]
10. Berman AL, Silverman MM. Near term suicide risk assessment: A commentary on the clinical relevance of protective factors. Arch Suicide Res 2020;24(sup2):S370-S380. [doi: 10.1080/13811118.2019.1612804] [Medline: 31079577]
11. Runeson B, Odeberg J, Pettersson A, Edbom T, Jildevik Adamsson I, Waern M. Instruments for the assessment of suicide risk: A systematic review evaluating the certainty of the evidence. PLoS One 2017;12(7):e0180292 [FREE Full text] [doi: 10.1371/journal.pone.0180292] [Medline: 28723978]
12. Ryan EP, Oquendo MA. Suicide risk assessment and prevention: Challenges and opportunities. Focus (Am Psychiatr Publ) 2020 Apr;18(2):88-99 [FREE Full text] [doi: 10.1176/appi.focus.20200011] [Medline: 33162846]
13. Steele IH, Thrower N, Noroian P, Saleh FM. Understanding Suicide Across the Lifespan: A United States Perspective of Suicide Risk Factors, Assessment and Management. J Forensic Sci 2018 Jan;63(1):162-171. [doi: 10.1111/1556-4029.13519] [Medline: 28639299]
14. Homan S, Gabi M, Klee N, Bachmann S, Moser A, Duri' M, et al. Linguistic features of suicidal thoughts and behaviors: A systematic review. Clin Psychol Rev 2022 Jul;95:102161 [FREE Full text] [doi: 10.1016/j.cpr.2022.102161] [Medline: 35636131]
15. Fernandes AC, Dutta R, Velupillai S, Sanyal J, Stewart R, Chandran D. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. Sci Rep 2018 May 09;8(1):7426 [FREE Full text] [doi: 10.1038/s41598-018-25773-2] [Medline: 29743531]
16. Aladağ AE, Muderrisoglu S, Akbas NB, Zahmacioglu O, Bingol HO. Detecting suicidal ideation on forums: Proof-of-concept study. J Med Internet Res 2018 Jun 21;20(6):e215 [FREE Full text] [doi: 10.2196/jmir.9840] [Medline: 29929945]
17. Tadesse M, Lin H, Xu B, Yang L. Detection of suicide ideation in social media forums using deep learning. Algorithms 2019 Dec 24;13(1):7 [FREE Full text] [doi: 10.3390/a13010007]
18. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: Opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018 Mar;3(3):223-230. [doi: 10.1016/j.bpsc.2017.11.007] [Medline: 29486863]
19. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. London, UK: Hachette; 2019.
20. Vajawat B, Varshney P, Banerjee D. Digital gaming interventions in psychiatry: Evidence, applications and challenges. Psychiatry Res 2021 Jan;295:113585. [doi: 10.1016/j.psychres.2020.113585] [Medline: 33303223]
21. Pham KT, Nabizadeh A, Selek S. Artificial intelligence and chatbots in psychiatry. Psychiatr Q 2022 Mar;93(1):249-253 [FREE Full text] [doi: 10.1007/s11126-022-09973-8] [Medline: 35212940]
22. He N, Yan Y, Wu Z, Cheng Y, Liu F, Li X, et al. Chat GPT-4 significantly surpasses GPT-3.5 in drug information queries. J Telemed Telecare 2023 Jun 22:1357633X231181922. [doi: 10.1177/1357633X231181922] [Medline: 37350055]
23. Nori H, King N, McKinney S. ,. Capabilities of GPT-4 on medical challenge problems 2023 [FREE Full text] [doi: 10.48550/arXiv.2303.13375]
24. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023 Mar 30;388(13):1233-1239. [doi: 10.1056/NEJMsr2214184] [Medline: 36988602]
25. Sanderson K. GPT-4 is here: what scientists think. Nature 2023 Mar;615(7954):773. [doi: 10.1038/d41586-023-00816-5] [Medline: 36928404]
26. Teebagy S, Colwell L, Wood E, Yaghy A, Faustina M. Improved performance of ChatGPT-4 on the OKAP exam: A comparative study with ChatGPT-3. MedrXiv Preprint posted April 03, 2023. [FREE Full text] [doi: 10.1101/2023.04.03.23287957]
27. Takagi S, Watari T, Erabi A, Sakaguchi K. Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: comparison study. JMIR Med Educ 2023 Jun 29;9:e48002 [FREE Full text] [doi: 10.2196/48002] [Medline: 37384388]

28. Wasserman D, Rihmer Z, Rujescu D, Sarchiapone M, Sokolowski M, Titelman D, European Psychiatric Association. The European Psychiatric Association (EPA) guidance on suicide treatment and prevention. Eur Psychiatry 2012 Feb;27(2):129-141. [doi: 10.1016/j.eurpsy.2011.06.003] [Medline: 22137775]
29. Levkovich I, Vigdor I. How school counsellors cope with suicide attempts among adolescents — A qualitative study in Israel. J Psychol Couns Sch 2020 Jun 18;31(1):63-75. [doi: 10.1017/jgc.2020.14]
30. Burnette C, Ramchand R, Ayer L. Gatekeeper training for suicide prevention: A theoretical model and review of the empirical literature. Rand Health Q 2015 Jul 15;5(1):16 [FREE Full text] [Medline: 28083369]
31. Stickl Haugen J, Sutter CC, Tinstman Jones JL, Campbell LO. Teachers as youth suicide prevention gatekeepers: An examination of suicide prevention training and exposure to students at risk of suicide. Child Youth Care Forum 2023;52(3):583-601 [FREE Full text] [doi: 10.1007/s10566-022-09699-5] [Medline: 35789706]
32. Vera-Varela C, Manrique Mirón PC, Barrigón ML, Álvarez-García R, Portillo P, Chamorro J, MEmind Study Group, et al. Low level of agreement between self-report and clinical assessment of passive suicidal ideation. Arch Suicide Res 2022;26(4):1895-1910. [doi: 10.1080/13811118.2021.1945984] [Medline: 34223799]
33. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychol Bull 2017 Feb;143(2):187-232. [doi: 10.1037/bul0000084] [Medline: 27841450]
34. Joiner TE, Van Orden KA, Witte TK, Selby EA, Ribeiro JD, Lewis R, et al. Main predictions of the interpersonal-psychological theory of suicidal behavior: empirical tests in two samples of young adults. J Abnorm Psychol 2009 Aug;118(3):634-646 [FREE Full text] [doi: 10.1037/a0016500] [Medline: 19685959]
35. Van Orden K, Witte T, Gordon K, Bender T, Joiner JT. Suicidal desire and the capability for suicide: tests of the interpersonal-psychological theory of suicidal behavior among adults. J Consult Clin Psychol 2008 Feb;76(1):72-83. [doi: 10.1037/0022-006X.76.1.72] [Medline: 18229985]
36. Levi-Belz Y, Gvion Y, Apter A. The serious suicide attempts approach for understanding suicide: Review of the psychological evidence. Omega (Westport) 2022 Dec;86(2):591-608. [doi: 10.1177/0030222820981235] [Medline: 33327864]
37. Bresin K, Sand E, Gordon K. Non-suicidal self-injury from the observer's perspective: a vignette study. Arch Suicide Res 2013;17(3):185-195. [doi: 10.1080/13811118.2013.805636] [Medline: 23889569]
38. Levi-Belz Y, Gamliel E. The effect of perceived burdensomeness and thwarted belongingness on therapists' assessment of patients' suicide risk. Psychother Res 2016 Jul;26(4):436-445. [doi: 10.1080/10503307.2015.1013161] [Medline: 25751580]
39. Shahsavar Y, Choudhury A. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. JMIR Hum Factors 2023 May 17;10:e47564 [FREE Full text] [doi: 10.2196/47564] [Medline: 37195756]
40. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 04;47(1):33 [FREE Full text] [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
41. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
42. Granello D, Granello P. Suicide: An Essential Guide for Helping Professionals and Educators. Boston, MA: Pearson/Allyn & Bacon; 2007.
43. Erford BT, Jackson J, Bardhoshi G, Duncan K, Atalay Z. Selecting suicide ideation assessment instruments: A meta-analytic review. Meas Eval Couns Dev 2017 Dec 20;51(1):42-59. [doi: 10.1080/07481756.2017.1358062]
44. Temsah M, Aljamaan F, Malki KH, Alhasan K, Altamimi I, Aljarbou R, et al. ChatGPT and the future of digital health: a study on healthcare workers' perceptions and expectations. Healthcare (Basel) 2023 Jun 21;11(13):1812 [FREE Full text] [doi: 10.3390/healthcare11131812] [Medline: 37444647]
45. Jones L. Resilience isn't the same for all: Comparing subjective and objective approaches to resilience measurement. WIREs Clim Change 2018 Sep 09;10(1):1-19 [FREE Full text] [doi: 10.1002/wcc.552]
46. Bonanno GA. The resilience paradox. Eur J Psychotraumatol 2021;12(1):1942642 [FREE Full text] [doi: 10.1080/20008198.2021.1942642] [Medline: 34262670]
47. Walker HL, Ghani S, Kuemmerli C, Nebiker CA, Müller BP, Raptis DA, et al. Reliability of medical information provided by ChatGPT: assessment against clinical guidelines and patient information quality instrument. J Med Internet Res 2023 Jun 30;25:e47479 [FREE Full text] [doi: 10.2196/47479] [Medline: 37389908]
48. Dahmen J, Kayaalp ME, Ollivier M, Pareek A, Hirschmann MT, Karlsson J, et al. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. Knee Surg Sports Traumatol Arthrosc 2023 Apr;31(4):1187-1189. [doi: 10.1007/s00167-023-07355-6] [Medline: 36809511]
49. Roumeliotis KI, Tselikas ND. ChatGPT and Open-AI models: a preliminary review. Future Internet 2023 May 26;15(6):192 [FREE Full text] [doi: 10.3390/fi15060192]
50. Couture V, Roy M, Dez E, Laperle S, Bélisle-Pipon JC. Ethical implications of artificial intelligence in population health and the public's role in its governance: Perspectives from a citizen and expert panel. J Med Internet Res 2023 Apr 27;25:e44357 [FREE Full text] [doi: 10.2196/44357] [Medline: 37104026]

51. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med 2021 Sep 21;4(1):140 [FREE Full text] [doi: 10.1038/s41746-021-00509-1] [Medline: 34548621]
52. Karabacak M, Ozkara BB, Margetis K, Wintermark M, Bisdas S. The advent of generative language models in medical education. JMIR Med Educ 2023 Jun 06;9:e48163 [FREE Full text] [doi: 10.2196/48163] [Medline: 37279048]

Abbreviations

AI: artificial intelligence
ITS: interpersonal theory of suicide
LLM: large language model
NLP: natural language processing

Edited by J Torous; submitted 25.07.23; peer-reviewed by M Alon - Tirosh, S Staubli; comments to author 14.08.23; revised version received 22.08.23; accepted 24.08.23; published 20.09.23.

Please cite as:
Levkovich I, Elyoseph Z
Suicide Risk Assessments Through the Eyes of ChatGPT-3.5 Versus ChatGPT-4: Vignette Study
JMIR Ment Health 2023;10:e51232
URL: https://mental.jmir.org/2023/1/e51232
doi:10.2196/51232
PMID:37728984

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org